Information Fusion 20 (2014) 225–241

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

A Multi-Modal Incompleteness Ontology model (MMIO) to enhance information fusion for image retrieval

Stefan Poslad [b], Kraisak Kesorn [a,*]

[a] Computer Science and Information Technology Department, Science Faculty, Naresuan University, Phitsanulok 65000, Thailand
[b] School of Electrical and Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, United Kingdom

ARTICLE INFO

Article history:
Received 7 October 2013
Received in revised form 3 February 2014
Accepted 20 February 2014
Available online 7 March 2014

Keywords:
Multi-Modal Ontology
Knowledge base
Incomplete Ontology
Visual and textual information fusion

ABSTRACT

A significant effort by researchers has advanced the ability of computers to understand, index and annotate images. This entails automatic domain specific knowledge-base (KB) construction and metadata extraction from visual information and any associated textual information. However, it is challenging to fuse visual and textual information and build a complete domain-specific KB for image annotation due to several factors such as: the ambiguity of natural language to describe image features; the semantic gap when using image features to represent visual content and the incompleteness of the metadata in the KB. Typically the KB is based upon a domain specific Ontology. However, it is not an easy task to extract the data needed from annotations and images, and then to automatically process these and transform them into an integrated Ontology model, because of the ambiguity of terms and because of image processing algorithm errors. As such, it is difficult to construct a complete KB covering a specific domain of knowledge. This paper presents a Multi-Modal Incompleteness Ontology-based (MMIO) system for image retrieval based upon fusing two derived indices. The first index exploits low-level features extracted from images. A novel technique is proposed to represent the semantics of visual content, by restructuring visual word vectors into an Ontology model by computing the distance between the visual word features and concept features, the so called *concept range*. The second index relies on a textual description which is processed to extract and recognise the concepts, properties, or instances that are defined in an Ontology. The two indexes are fused into a single indexing model, which is used to enhance the image retrieval efficiency. Nonetheless, this rich index may not be sufficient to find the desired images. Therefore, a Latent Semantic Indexing (LSI) algorithm is exploited to search for similar words to those used in a query. As a result, it is possible to retrieve images with a query using (similar) words that do not appear in the caption. The efficiency of the KB is validated experimentally with respect to three criteria, correctness, multimodality, and robustness. The results show that the multi-modal metadata in the proposed KB could be utilised efficiently. An additional experiment demonstrates that LSI can handle an incomplete KB effectively. Using LSI, the system can still retrieve relevant images when information in the Ontology is missing, leading to an enhanced retrieval performance.

1. Problem statement

The main benefit of using knowledge representation models for Image Retrieval System (IMR), is that they are able to reduce the semantic gap, *the gap between the user perception and the low-level feature abstraction from the visual content,* providing relations between these low-level and high-level concepts can be identified, enhancing concept-based retrieval. Typically an Ontology model is used for knowledge representation, which represents physical things in this world using a hierarchical model expressed in the form of classes and relationships to support human decision-making, learning, reasoning and explanation.

An Ontology provides a useful way for formalising the semantics of the represented information. In principle, an Ontology can actually be the semantic representation for an information system in a concrete and useful manner [1]. Ontologies are used by an IMR for reducing the semantic gap by storing the knowledge structures for summarising, discovering, classifying, browsing, retrieving and annotating images. Ontology-based frameworks are proposed for IMR in numerous collections. Ontologies for manual image annotation and semantic retrieval for collections of animal pictures have

* Corresponding author. Tel.: +66 55 963267.
E-mail addresses: stefan@eecs.qmul.ac.uk (S. Poslad), kraisakk@nu.ac.th (K. Kesorn).

http://dx.doi.org/10.1016/j.inffus.2014.02.003
1566-2535/

been presented in [2]. An Ontology for considering art images has been presented in [3]. In [4], Ontologies also have been applied successfully for handling museum collections. These frameworks have validated the hypothesis that Ontologies can help to improve information retrieval effectiveness by making it possible to find semantically similar documents.

Gasevic et al. [5] has summarised the benefits of using Ontologies in IR and IMR systems as follows. Firstly, the Ontology structure can be exploited to measure the semantic similarity. For example, the term list (''Michael Phelps'', ''Swimming'', ''Gold medal'') has no syntactic similarity to the term list (''London'', ''2012'') although the two lists are semantically relevant. This is because Michael Phelps was the Swimming (Gold medal) winner at the London 2012 Olympic Games. The similarity can be obtained using the relationship between concepts, e.g., Michael Phelps-<*is_champion_of*>-Free_style_swimming-<*participate_in*>-London 2012. Secondly, semantic annotations that may not be explicitly mentioned in a caption can be identified using knowledge stored in an Ontology. For example, if many entities, locations and athletes related to the London 2012 Olympic Games appear in a text caption, and also the time context is London 2012, the annotation system can infer that the London 2012 Olympic Games itself is relevant to an image. Hence, this could be added to the semantic annotation although the text caption does not explicitly contain the time context ''London 2012''. Ontologies may also be used to enable query expansion [6,7] but this will not be described in detail here. These are some of the potential uses of Ontologies in IMR.

Existing work on IMR tends to be based upon only single-modality information, either textual information or visual features. Consequently, such work suffers from several limitations. For example, an IMR system is not able to describe the high-level semantics of images, based only on any distinctive low-level visual features when text descriptions of images are not supplied, because the extracted visual features themselves cannot be used to represent the content of images effectively. Text and image are two different modalities that can be fused to represent 'things' more completely. However, there are some invariant and implicit connections between them that complicates the information fusion of the two [8]. Often, the textual information surrounding images includes descriptions of images generated by humans. These image captions should not be disregarded, as they can aid image interpretation. Nonetheless, exploiting only text information for visual content interpretation can suffer from the ambiguity of the text descriptions used because they are written using natural language, which may be ambiguous and imprecise. As such, using single-modality information is not adequate to enhance the interpretation power for IMR. Multimodality information should be utilised to facilitate image interpretation, classification and retrieval.

Fusing text and image features has been proposed by several researchers in order to improve the image search results [9–15]. These approaches focus on improving the retrieval performance in order to get more accurate results. However, there are some challenges in integrating visual and textual metadata in a knowledge base for IMR. Firstly, the KB model should be designed to interlink both visual and textual metadata together, in order to facilitate the image classification and retrieval performance. Secondly, automatic KB construction and metadata extraction from text captions are very challenging tasks from which to build a complete KB due to several factors:

(1) Those text captions may be ambiguous because they are written using natural language.
(2) Standard Ontology languages such as the W3C Resource Description Framework (RDF) or Web Ontology Language (OWL) cannot directly represent some semantic aspects, e.g. uncertainty and gradual truth, value) because the latter hard-wires a specific logic, description logic, into the Ontology representation.

These are the reasons why a complete Ontology cannot be built even when the system processes a large training set in order to acquire the metadata to populate the KB model. In this paper, Ontology incompleteness refers to an absence of some semantic metadata and also to relationships between concepts that cannot be represented in an Ontology. The KB may be incomplete, resulting in the failure of finding relevant information of the retrieval mechanism. Image retrieval systems operating solely on information in the KB, sometimes, are less effective than systems that use information directly from text captions. This is because of the inadequate coverage of annotations by a KB [16]. As such, IMR should be able to deal with information incompleteness in the KB.

These limitations drive the research objectives described in the following sections. Solutions to these problems are vital to achieve a good quality knowledge-base for use by an image retrieval system and are, as such, the main focus in this paper.

2. State of the art

The current survey focuses on a discussion of Ontology-based frameworks for IMR that use a shared or a standard encoding, i.e., MPEG-7.

2.1. Ontology-based frameworks for IMR

Numerous techniques have been introduced to resolve the semantic gap problem in the past decade. Early IMR approaches were based on low-level features which fail to capture the underlying conceptual associations in images. Since visual data cannot be used in its original form, it needs to be analysed and transformed into a format which can be used by Knowledge Management (KM) systems. Typically, knowledge is extracted by image processing algorithms and transformed into metadata. This metadata describes the content, context and visual features of an image document, which are manipulated and processed by standard information retrieval methods. Image data contains large numbers of unstructured and dynamic visual features. How to establish a good knowledge representation model to represent visual content is very important for IMR [9–13]. In part through the emergence of the Semantic Web [17], an Ontology model has become common to represent visual content, enabling an IMR system to perform semantic retrieval.

Tansley [18] proposed a method to bridge the semantic gap using Web images and their surrounding text, file name and alterative tags. Using the WordNet Thesaurus [19], the system can solve the NL vagueness problem of text captions. Unfortunately, this framework exploits only textual information and supports only text-based queries. Schreiber et al. [2] presented a method to index and search image collections using Ontologies. The system uses the terminology from WordNet for annotations. The main limitation of this framework is that the knowledge-base is designed as a closed KB, because if a query concept is outside the scope of KB, the system cannot find any relevant images for users. Dasiopoulou et al. [17] presented a framework comprising two main modules, a semantic analysis module and a retrieval module. The domain Ontology provides the conceptualisation and vocabulary for structuring content annotations. The analysis module is used to guide the analysis process and to support the detection of certain concepts defined in a domain Ontology using low-level features. The system exploits the low-level features of an image and matches