MEDICAL IMAGE UNDERSTANDING THROUGH THE INTEGRATION OF CROSS-MODAL OBJECT RECOGNITION WITH FORMAL DOMAIN KNOWLEDGE

Manuel Möller, Michael Sintek, Paul Buitelaar
German Research Center for Artificial Intelligence, Trippstadter Str. 122, 67663 Kaiserslautern, Germany

Saikat Mukherjee
Siemens Corporate Research, 755 College Road East, Princeton, USA

Xiang Sean Zhou
Siemens Medical Solutions, 51 Valley Stream Parkway, Malvern PA, USA

Jörg Freund
Siemens Medical Solutions, Hartmannstr. 16, 91052 Erlangen, Germany

Keywords: Semantic Web, Ontologies, NLP, Medical Imaging, Image Retrieval.

Abstract: Rapid advances in medical imaging scanner technology have increased dramatically in the last decade the amount of medical image data generated every day. By contrast, the software technology that would allow the efficient exploitation of the highly informational content of medical images has evolved much slower. Despite the research outcomes in image understanding and semantic modeling, current image databases are still indexed by keywords assigned by humans and not by the image content. The reason for this slow progress is the lack of scalable and generic information representations capable of overcoming the high-dimensional nature of image data. Indeed, most of the current content-based image search applications are focused on the indexing of certain image features that do not generalize well and use inflexible queries. We propose a system combining medical imaging information with semantic background knowledge from formalized ontologies, that provides a basis for building universal knowledge repositories, giving clinicians a fully cross-lingual and cross-modal access to biomedical information of all forms.

1 INTRODUCTION

Rapid advances in imaging technology have dramatically increased the amount of medical image data generated daily by hospitals, pharmaceutical companies, and academic medical research\(^1\). Technologies such as 4D 64-slice CT, whole-body MR, 4D Ultrasound, and PET/CT can provide incredible detail and a wealth of information with respect to the human body anatomy, function and disease associations. This increase in the volume of data has brought about significant advances in techniques for analyzing such data. The precision and sophistication of different image understanding techniques, such as object recognition and image segmentation, have also improved to cope with the increasing complexity of the data.

However, these improvements in analysis have not resulted in more flexible or generic image understanding techniques. Instead, the analysis techniques are very object specific and not generic enough to be applied across different applications. Throughout this paper we will address this fact as “lack of scalability”. Consequently, current image search techniques, whether for Web sources or for medical PACS (Picture Archiving and Communications System), are still dependent on the subjective association of keywords to images for retrieval.

One of the important reasons behind this lack

---

\(^1\)For example, University Hospital of Erlangen, Germany, has a total of about 50 TB of medical images. Currently they have approx. 150,000 medical examinations producing 13 TB per year.
of scalability in image understanding techniques has been the absence of generic information representation structures capable of overcoming the feature-space complexity of image data. Indeed, most current content-based image search applications are focused on indexing syntactic image features that do not generalize well across domains. As a result, current image search technology does not operate at the semantic level and, hence, is not scalable.

We propose to use hierarchical information representation structures, which integrate state-of-the-art object recognition algorithms with generic domain semantics, for a more scalable approach to image understanding. Such a system will be able to provide direct and seamless access to the informational content of image databases.

Our approach is based on the following main techniques:

- Integrate the state-of-the-art in semantics and image understanding to build a sound bridge between the symbolic and the subsymbolic world. This cross-layer research approach defines our road-map to quasi-generic image search.

- Integrate higher level knowledge represented by formal ontologies that will help explain different semantic views on the same medical image: structure, function, and disease. These different semantic views will be coupled to a backbone ontology of the human body.

- Exploit the intrinsic constraints of the medical imaging domain to define a rich set of queries for concepts in the human body ontology. The ontology not only provides a natural abstraction over these queries but also statistical image algorithms could be associated to semantic concepts for answering these queries.

Our focus is on filling the gap between what is current practice in image search (i.e., indexing by keywords) and the needs of modern health provision and research. The overall goal is to empower the medical imaging content-stakeholders (clinicians, pharmaceutical specialists, patients, citizens, and policy makers) by providing flexible and scalable semantic access to medical image databases. Our short term goal is to develop a basic image search engine and prove its functionality in various medical applications.

In 2001 Berners-Lee and others published a visionary article on the Semantic Web (Berners-Lee et al., 2001). The use-case they described was about the use of meta-knowledge by computers. For our goals we propose to build a system on existing Semantic Web technologies like RDF (Brickley and Guha, 2004) and OWL (McGuinness and van Harmelen, 2004) which were developed to lay the foundations of Berners-Lee’s vision. From this point of view it is also a Semantic Web project.

Therefore we propose a system that combines medical imaging information with semantic background knowledge from formalized ontologies and provides a basis for building universal knowledge repositories, giving clinicians cross-modal (independent from different modalities like PET, CT, ultrasound) as well as cross-lingual (independent of particular languages like English and German) access to various forms of biomedical information.

2 GENERAL IDEA

There are numerous advanced object recognition algorithms for the detection of particular objects on medical images: (Hong et al., 2006) at anatomical level, (Tu et al., 2006) at disease level and (Comaniciu et al., 2004) at functional level. Their specificity is also their limitation: Existing object recognition algorithms are not at all generic. Given an arbitrary image it still needs human intelligence to select the right object recognizers to apply to an image. Aiming to gain a pseudo-general object recognition one can try to apply the whole spectrum of available object recognition algorithms. But it turns out that in generic scenarios even with state-of-the-art object recognition tools the accuracy is below 50 percent (Chan et al., 2006; Müller et al., 2006).

In automatic image understanding there is a semantic gap between low-level image features and techniques for complex pattern recognition. Existing work aims to bridge this gap by ad-hoc and application specific knowledge. In contrast our objective is to create a formal fusion of semantic knowledge and image understanding to bridge this gap to support more flexible and scalable queries.

For instance, human anatomical knowledge tells us that it is almost impossible to find a heart valve next to a knee joint. Only in cases of very severe injuries these two objects might be found next to each other. But in most cases the anatomical intuition is correct and, hence, the background knowledge precludes the recognition of certain anatomical parts given the presence of other parts. It is in this use of formalized knowledge that ontologies² come into play within our framework.

In the context of medical imaging it is necessary to define image semantics for parts of human anatomy.

---

²According to Gruber (Gruber, 1995), an ontology is a specification of a (shared) conceptualization.
In this domain the expert’s knowledge is already formalized in comprehensive ontologies like the *Foundational Model of Anatomy* (Rosse and Mejino, 2003) for human anatomy or the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10)\(^3\) of World Health Organization for a classification of human diseases. These ontologies represent a rich medical knowledge in a standardized and machine interpretable format.

In contrast to current work which defines ad-hoc semantics, we take the novel view that within a constrained domain the semantics of a concept is defined by the queries associated with it. We will investigate which types of queries are asked by medical experts to ensure that the necessary concepts are integrated into the knowledge base. We believe that in IR applications this view will allow a number of advances which will be described in the following sections.

We chose the medical domain as our area of application. Unlike common language and many other scientific areas the medical domain has clear definitions for its technical terms. Ambiguities are rare which eases the task of finding a semantic abstraction for a particular text or image. However, our framework is generic and can be applied to other domains with well-defined semantics.

### 3 ASPECTS OF USING ONTOLOGIES

Ontologies (usually) define the semantics for a set of objects in the world using a set of classes, each of which may be identified by a particular symbol (either linguistic, as image, or otherwise). In this way, ontologies cover all three sides of the “semiotic triangle” that includes object, referent, and sign (see Fig. 1). I.e., an object in the world is defined by its referent and represented by a symbol (Ogden and Richards, 1923—based on Peirce, de Saussure and others). Currently, ontology development and the Semantic Web effort in general have been mostly directed at the referent side of the triangle, and much less at the symbol side. To allow for automatic multilingual and multimedia knowledge markup a richer representation is needed of the linguistic and image-based symbols for the object classes that are defined by the ontology (Buitelaar et al., 2006).

Cross-Modal Image Retrieval. Current systems for medical IR depend on the modality of the stored images. But in medical diagnosis very different imaging techniques are used such as PET, CT, ultrasonography, or time series data from 4D CT etc. Each technique produces images with characteristic appearance. For tumor detection, for example, often PET (to identify the tumor) and CT (to have a view on the anatomy) are combined, to formulate a precise diagnosis with a proper localization of the tumor. The proposed system will allow to answer queries based on semantic similarity and not only visual similarity.

Full Modality Independence. Cross-Modality even can be driven another step forward by integrating documents of any format into one single database. We plan to also include text documents like medical reports and diagnoses. On the level of semantic representation they will be treated like the images. Accordingly, the system will be able to answer queries not only with images but also with

---

\(^3\)http://www.who.int/classifications/icd/en/
text documents including similar concepts as in the retrieved images.

**Improved Relevance of Results.** Current search engines retrieve documents which contain the keyword from the query. The documents in the result set are ranked by various techniques using information such as their inter-connectivity, statistical language models, or the like. For huge datasets search by keyword often returns very large result sets. Ranking by relevance is hard.

This holds for low-level image retrieval as well. Here only two similarity measures are applicable: through visual similarity which can be completely independent from the object and context and via a comparison between keywords and image annotations. With current IR systems the user is forced to use pure keyword-based search as a detour while in fact he or she is searching for documents and/or images including certain concepts.

Our notion of keyword-based querying goes beyond current search engines by mapping keywords from the query to ontological concepts. Our system provides the user with a semantic representation. That allows the user to ask for a concept or a set of concepts with particular relationships. This allows for more precise queries than a simple keyword-based retrieval mechanism and likewise better matching between query and result set.

**Inferencing of Hidden Knowledge.** By mapping the keywords from a text-based query to ontological concepts and the use of semantics the system is able to infer implicit results. This allows us to retrieve images which are not annotated explicitly with the query concepts but with concepts related to them through the ontology.

To represent the complex knowledge of the medical domain and allow a maximum of flexibility in the queries we will have to enrich the ontology by rules and allow to use rules in the queries. Another point will be an integration of spatial representation of anatomical relations as well as an efficient implementation of spatial reasoning.

### 4 LEVELS OF SEMANTIC ABSTRACTION

Our notion of semantic imaging is to ground the semantics of a human anatomical concept on a set of queries associated with it. The constrained domain of a human body enables us to have a rich coverage of these queries and, consequently, define image semantics at various levels of the hierarchy of the human anatomy.

Fig. 3 gives an overview of the different abstraction levels in the intended system. For the proposed system we want to take a step beyond the simple dichotomy between a symbolic and subsymbolic representation of images. Instead, from our perspective there is a spectrum ranging from regarding the images as simple bit vectors over color histograms, shapes and objects to a fully semantic description of what is depicted. The most formal and generic level of representation is in form of an ontology *(formal ontological modeling)*. The ontology holds information about the general structure of things. Concrete entities are to be represented as *semantic instances* according to the schema formalized in the ontology.

To emphasize the difference to the dichotomic view, we call the lower end of this spectrum *informal* and the upper end *formal* representation. From our perspective the abstraction has to be modeled as a multi-step process across several sublevels of abstraction. This makes it easier to close the gap between the symbolic and subsymbolic levels from the classic perspective. Depending on the similarity measure that is to be applied for a concrete task different levels of the abstraction process will be accessed.

If a medical expert searches for images that are looking similar to the one he or she recently got from an examination, the system will use low-level features like histograms or the bit vector representation. In another case the expert might search for information about a particular syndrome. In that case the system will use features from higher abstraction levels like the semantic description of images and texts in the database to be able to return results from completely different modalities.

We believe that text documents have to be understood in a similar way. Per se, a text document is just a string of characters. This is similar to regarding an image as a sheer bit vector. Starting from the string of characters, in a first step relatively simple methods can be used to identify terms. In further steps technologies from concept based Cross Language Information Retrieval (CLIR)**5** are applied to map terms in the documents to concepts in the ontology (Volk et al., 2003; Vintar et al., 2003; Carbonell et al., 1997; Eichmann et al., 1998). CLIR currently can be divided into three different methods: approaches based on bilingual dictionary lookup or

---

**Footnotes:**

4We aim at using standard OWL reasoners like Racer, FaCT++ or Pellet.

5The research project MUCHMORE (http://muchmore.dfki.de/) was focused on this aspect.
query either through sample images, or pose structured queries using conceptual descriptors, or use natural language to describe queries. In the following, we explain each of these different methods.

Query by Sample Image. Basically there are two different approaches to image based queries. The first approach retrieves images from the database which are looking similar. Only low-level image features are used to select results for this type of query. The ability to match the image of a current patient to similar images from a database of former medical cases can be of great help in assisting the medical professional in his diagnosis (see Sect 4) we believe that image understanding has to be an abstraction over several levels. To answer queries by sample image we will make use of the more informal features extracted from the images. The support for these queries is based on state-of-the-art similarity-based image retrieval techniques (Deselaers et al., 2005).

Today there are various image modalities in modern medicine. Many diseases like cancer require to look at images from different modalities to formulate a reliable diagnosis (see example in Sect. 3). The second approach therefore takes the image from the query and extracts the semantics of what can be seen on the image. Through mapping the concrete image to concepts in the ontology, an abstract representation is generated. This representation can be used for a matching on the level of image semantics with other images in the database. Applying this method makes it possible to use a CT image of the brain to search for images from all available modalities in the database (see Fig. 4–6).

Query by Conceptual Descriptions. Similar to the use of SQL for querying structured relational databases, special purpose languages are also required for querying semantic metadata. Relying on well-established standards we propose using a language on top of RDF, such as SPARQL, for supporting generic structured semantic queries.

Query by Natural Language. From the point of the medical expert having a natural language interface is very important. Through a textual interface the user directly enters keywords which are mapped to ontology concepts. Current systems like the IRMA-Project (see Sect. 6) only allow to search for keywords which are extracted offline and stored as annotations. Since our system has to compose a semantic representation of each query, the ontological background knowledge can be used in an iterative process of query refine-

ment. Additionally, it will be possible to use complete text documents as queries.

In cases where the system cannot generate a semantic representation—due to missing knowledge about a knew syndrome, therapy, drug or the like—it will fall back to a normal full text search. If the same keyword is used frequently this can be used as evidence that the foundational ontology has to be extended to cover the corresponding concept(s).

6 RELATED WORK

Most current work in content based image retrieval models object recognition as a probabilistic inferencing problem and use various mathematical methods to cope with the problems of image understanding. These techniques use image features which are tied to particular applications and, hence, suffer from a lack of scalability.

Among extant work in fusing semantics with image understanding, (Hunter, 2001) describes a technique for modeling the MPEG-7 standard, which is a set of structured schemas for describing multimedia content, as an ontology in RDFS. There has been some research (Barnard et al., 2003; Lavrenko et al., 2003; Lim, 1999; Carneiro and Vasconcelos, 2005; Town, 2006; Mezaris et al., 2003; Mojsilovic et al., 2004) on semantic imaging relying primarily on associating word distributions to image features. However, these works used hierarchies of words for semantic interpretation and did not attempt to model image features themselves in levels of abstraction. Furthermore, the lack of formal modeling made these techniques susceptible to subjective interpretations of the word hierarchies and, hence, were not truly scalable. Especially in the context of medical imaging, our notion of semantics is tied to information gathered from physics, biology, anatomy, etc. This is in contrast to perception-based subjective semantics in these works.

The goal of the project “Image Retrieval in Medical Applications” (IRMA) (Lehmann et al., 2003) was an automated classification of radiographs based on global features with respect to imaging modality, direction, body region examined and the biological system under investigation. The aim was to identify image features that are relevant for medical diagnosis. These features were derived from a database of 10.000 a-priori classified and registered images. By means of prototypes in each category, identification of objects and their geometrical or temporal relationships are handled in the object and the knowledge layer, respectively. Compared to the system proposed


