

## ViSen

### Abstract

#### Tagging visual data with semantic descriptions

Today a typical Web document will contain a mix of visual and textual content. Most traditional tools for search and retrieval can successfully handle textual content, but are not prepared to handle heterogeneous documents. The new type of content demands the development of new efficient tools for search and retrieval.

The visual sense project aims at mining automatically the semantic content of visual data to enable ?machine reading? of images. In recent years, we have witnessed significant advances in the automatic recognition of visual concepts (VCR). These advances allowed for the creation of systems that can automatically generate keyword-based image annotations. The goal of this project is to move a step forward and predict semantic image representations that can be used to generate more informative sentence-based image annotations. Thus, facilitating search and browsing of large multi-modal collections. More specifically, the project targets three case studies, namely image annotation, re-ranking for image search, and automatic image illustration of articles. It will address the following key open research challenges:

1. To develop methods that can predict a semantic representation of visual content. This representation will go beyond the detection of objects and scenes and will also recognize a wide range of object relations.
2. To extend state-of-the-art natural language techniques to the tasks of mining large collections of multi-modal documents and generating image captions using both semantic representations of visual content and object/scene type models derived from semantic representations of the multi-modal documents.
3. To develop learning algorithms that can exploit available multi-modal data to discover mappings between visual and textual content. These algorithms should be able to leverage ?weakly? annotated data and be robust to large amounts of noise.

For this purpose, the current project will build on expertise from multiple disciplines, including computer vision, machine learning and natural language processing (NLP), and gathers four research groups from University of Surrey (Surrey, UK), Institut de Robòtica i Informàtica Industrial (IRI, Spain) , Ecole Centrale de Lyon (ECL, France), and University of Sheffield (Sheffield, UK) having each well established and complementary expertise in their respective areas of research.

(2011)

From Data to New Knowledge (D2K)

## Partnership & Contact

ViSen starts in October 2012, lasts 36 months and involves the partnerships below. The financial support of CHIST-ERA is about 1300000 €.

Partnership	
Institut de Robòtica i Informàtica Industrial	Spain
University of Surrey Department of Electronic Engineering	United Kingdom
University of Sheffield Department of Computer Science	United Kingdom
Ecole Centrale de Lyon	France

Contact: Dr. **Krystian MIKOLAJCZYK** (coordinator), [K.Mikolajczyk@surrey.ac.uk](mailto:K.Mikolajczyk@surrey.ac.uk) <sup>[1]</sup>



## Attachment

## Size

 <a href="#">CHIST-ERA Call 2011 - D2K Topic - ViSen 2016.pdf</a> <sup>[2]</sup>	878.48 KB
 <a href="#">CHIST-ERA Call 2011 - D2K Topic - ViSen 2014.pdf</a> <sup>[3]</sup>	1.27 MB

© CHIST-ERA

◦ [Administration](#)

**Source URL:** <http://www.chistera.eu/projects/visen>

## Links:

[1] <mailto:K.Mikolajczyk@surrey.ac.uk>

[2] <http://www.chistera.eu/sites/chistera.eu/files/CHIST-ERA%20Call%202011%20-%20D2K%20Topic%20-%20ViSen%202016.pdf>

[3] <http://www.chistera.eu/sites/chistera.eu/files/CHIST-ERA%20Call%202011%20-%20D2K%20Topic%20-%20ViSen%202014.pdf>