An integrated approach

Professor Rolf Ingold highlights how his collaborative research into document, image and voice analysis is pushing technological boundaries and producing new insights into a range of different manuscripts.

To start, could you introduce your research team and outline what skills and experience are represented?

The Document, Image and Voice Analysis (DIVA) research group at the University of Fribourg, Switzerland, is located in the Department of Informatics and is currently composed of a group leader (myself), two senior researchers (experts in document analysis and user interaction, respectively), one postdoctoral researcher and eight PhD students. The group also collaborates intensively with researchers from the University of Applied Sciences and Arts Western Switzerland (HES-SO) in Fribourg, who complement our expertise in signal processing and ICT.

The DIVA group deals in general with multimodal signal analysis obtained from various sensors such as scanners, cameras, digital pens, microphones and, more recently, physiological sensors. Our expertise can be applied to different research fields such as image analysis, document analysis and understanding, speech processing, writer identification, signature verification, behavioural biometry, emotion analysis, forensics and various types of user interaction.

In 2011, you and your collaborators on the project VisualAudio won the James A. Lindner Prize – awarded jointly by the International Association of Sound and Audiovisual Archives (IASA), Association of Moving Image Archivists (AMIA) and Southeast Asia-Pacific Audiovisual Archive Association (SEAPAVAA) – which is intended to recognise and reward outstanding contributions to research in the technology of preservation of moving images and recorded sound. What made this project unique?

The VisualAudio project was conducted in collaboration with the Swiss National Sound Archives in Lugano and HES-SO Fribourg. Overall, the aim was to recover the sound from old phonographic records that had deteriorated too much to allow it to be played by mechanical means. The idea was to replace the traditional needle with an ultrahigh-resolution scanner (resolution of 1 μm per pixel) and retrieve the groove shape by image analysis performed on a fine-grain photograph taken with a dedicated camera, before transforming it back into an acoustic signal.

Could you briefly outline your research on computer-assisted palaeography and explain how HisDoc 2.0 builds on the first project?

In the original HisDoc project (2009-13) funded by the Swiss National Science Foundation, we focused on the development of a complete processing chain composed of layout analysis, handwritten text recognition, and finally indexation and retrieval of historical documents. With HisDoc 2.0 we will now move towards computer-assisted palaeography, aiming to develop new techniques to automatically extract valuable information about scripts and scribes, and incorporate existing information derived from catalogues into script analysis.

Why have you chosen to concentrate on medieval manuscripts in the HisDoc projects? Will your research translate for use with other historical documents?

Medieval manuscripts pose many challenges to researchers. Some of these challenges concern the original characteristics of the manuscripts, such as complex non-rigid layouts that may include several text entities of different kinds, touching text lines, interlinear gloss, annotations, ornaments, illuminations and several scripts within one page. Additional challenges come about through degradation as a result of the ageing process of the paper (parchment) and ink. These can include bleed-through, stains, mould and other forms of background noise.

In our research group, it is therefore our aim to tackle these very complex problems. Moreover, in our studies we also work with synthetically generated realistic distortions in order to stretch the boundaries of what has been possible so far. Although we are currently concentrating on medieval manuscripts, we are intending to develop methods that are easily adaptable to other kinds of documents and scripts.

Regarding the research on electronic documents, why is the recovery of logical structures still so challenging?

Unlike physical structures – which are considered to be universal and applicable for nearly all kinds of modern documents – logical structures are very complicated to handle. They cannot be defined in a generic way and must always be adapted to different document classes. For instance, legal texts, business letters, administrative forms, scientific journals and advertisements all have their own logics. Additionally, the logical structure is even specific to the considered applications and has to be adapted accordingly. This adaptation is difficult and usually requires a high level of expertise in both computer science and the application field.

Looking ahead, by what means will your research facilitate the examination and annotation of manuscripts in future?

While the focus of our project HisDoc 2.0 is on fundamental research, in one of our other projects we are targeting the development of an holistic digital workspace for the humanities scholars of the future, entitled DivaDesk. This will allow such scholars to make full use of the new possibilities that are arising as a result of technological developments and the digitisation of documents.
DIGITAL HUMANITIES IS a field of research that exists at the juncture between computing and the humanities. Academics in this area utilise computational methods to answer research questions, undertake extensive document analysis and examine large cultural datasets. Excitingly, this discipline was proclaimed “the first ‘next big thing’ in a long time” at the 2009 Modern Language Association (MLA) convention in Philadelphia, USA.

One prominent research group contributing to digital humanities is the Document, Image and Voice Analysis (DIVA) team from the Department of Informatics at the University of Fribourg in Switzerland. Led by Professor Rolf Ingold, the team conducts extensive research into multimedia engineering and multimodal data analysis. Together, the researchers are collaborating to develop innovative software solutions that aim to provide academics with useful tools for producing, collating and analysing knowledge that exists in digital contexts.

IMAGE PROCESSING

Rapid technological developments over the past few decades have brought about a seismic shift in the work practices of researchers in the humanities. No longer is it always necessary to visit repositories to view physical historical documents; instead, researchers are now able to work with digital facsimiles, allowing them to compare different documents and annotate digital images. Yet while this provides unprecedented access to a vast amount of information, handling these data necessitates specialised and user-friendly tools that make text passages searchable and forge links between corresponding documents. Although tools for these specific tasks do exist, the DIVA group’s vision is to create a single, holistic workspace integrating useful predictions, searchable texts and adequate visualisations. Indeed, it is in this context that DivaDia/DivaDesk – two visionary workspaces, respectively used for labelling and analysing historical document images – are being developed.

Essentially, the DivaDia workspace enables the fast generation of significant amounts of Ground Truth data, which refers to a set of sample images annotated by experts. These annotations are important because they incorporate the location of text regions, their transcriptions and key information about the script and scribes.

DivaDia consists of three main modules: Item Description, Content Representation and Research Data. In Item Description, the primary concept is incremental learning, whereby the computer system adapts and applies its existing knowledge to new data. In Content Representation, there are sub-modules for transcription, critical editions of several different textual versions, and the representation of non-textual elements such as images or diagrams. Finally, Research Data enables the generation and access of data through the integration of state-of-the-art natural language processing and information retrieval methods.

Looking ahead, the aim is for DivaDesk to become a user-friendly tool implemented across a variety of platforms – Windows, Mac and Linux. Additionally, the goal is to ensure that different texts are searchable, providing researchers with semi-automatic state-of-the-art methods to make predictions: “The envisaged workspace will provide further functionalities, supporting high-performance search, the comparison of editions and texts, and seamless connections to diverse research data,” discloses Ingold. “DivaDesk will be useful for research in the humanities, as well as for pushing the limits in state-of-the-art document image analysis.”

EXPLORING CULTURAL HERITAGE

Another DIVA research initiative – HisDoc – is in its second iteration, after successful completion of the original project in 2013. The initial HisDoc project aimed to develop methods and algorithms for textual heritage analysis and indexing. Through HisDoc 2.0, Ingold and his team are now advancing computer-assisted...
INNOVATION

A HOLISTIC APPROACH INTEGRATING TEXT LOCALISATION, SCRIBE IDENTIFICATION, AND SEMANTIC KNOWLEDGE FOR HISTORICAL DOCUMENT ANALYSIS

OBJECTIVES

To contribute to cultural heritage with a focus on document analysis and understanding applied to the context of historical documents.

KEY COLLABORATORS

Professor Dr Marcus Liwicki, University of Fribourg, Switzerland, and University of Kaiserslautern, Germany • Dr Andreas Fischer, École Polytechnique de Montréal, Canada • Angelika Garz, Michael Baechler, Chen Kai, Mathias Seuret; Hao Wei, University of Fribourg, Switzerland

FUNDING

Swiss National Science Foundation (SNSF)

CONTACT

Professor Rolf Ingold
Head, Document, Image and Voice Analysis (DIVA) research group
Department of Informatics
University of Fribourg
Boulevard de Pérolles 90
1700 Fribourg
Switzerland
T +41 26 300 84 66
E rolf.ingold@unifr.ch
http://diuf.unifr.ch/main/diva

PROFESSOR ROLF INGOLD is Head of the DIVA research group, former President of the Department of Informatics (2006-08) and Dean of the Faculty of Science (2010-13). He received his PhD in Computer Science from the Swiss Federal Institute of Technology Lausanne (EPFL), and joined the University of Fribourg as Associate Professor in 1989, promoted to Full Professor in 1997. He was highly involved in the Swiss National Centre of Competence in Research on Interactive Multimodal Information Management (NCCR-IM2). During the last decade, his activities have concentrated on image analysis and recognition, speech processing, and multimodal signal analysis, with applications in biometry, cultural heritage and user interaction. Ingold’s most recent achievements cover historical document analysis for cultural heritage preservation, in which the University of Fribourg has become an incontestable leader. In 2011, he was awarded the James A. Lindner Prize.

There are three major innovative components to the DIVA team’s HisDoc 2.0 project. First, it will comprise a single, holistic approach with integrated methods for text localisation, script discrimination and scribe identification. Second, it will incorporate additional domain knowledge into document image analysis (DIA) algorithms, thereby improving existing analysis methods. Finally, it will enable the examination of documents with complex layouts and several different scribes – a task that has not been possible to date: “Existing approaches presume laboratory conditions, such as high-quality binarisation or pre-segmented text regions, and focus on sub-tasks, treating interrelated tasks independently,” explains Ingold. “In real-world applications, reliable script analysis depends on the exact localisation of text regions on a page, which again depends on discriminating scripts. With HisDoc 2.0 we will develop approaches that open new doors for the improvement of DIA methods and beyond.”

DOCUMENT ANALYSIS

In addition to developing document analysis methods for historical documents, a further focus of the DIVA team’s research is centred on methods for processing and analysing modern documents that exist in a digital form. In an attempt to

An innovative team

The DIVA research group is conducting a wide range of projects centred on the following research themes:

- Document analysis, recognition and reengineering
- Speech analysis and processing
- Biometry
- Gesture and emotion recognition
- Human-machine interaction

The DIVA group’s vision is to create a single, holistic workspace that facilitates useful predictions, searchable texts and adequate visualisations

MAKING PROGRESS

Looking to the future, DIVA researchers are eager to continue refining and developing their cutting-edge computational tools and methods, thereby enabling academics to benefit from rapid technological developments and the digitisation of documents. Specifically, the long-term objective of Ingold’s research is the widespread use of DIVA’s state-of-the-art tools and holistic digital workspace. The hope is that these innovations will provide researchers with the apparatus to answer a range of cultural, social, historical and philosophical questions, consequently advancing knowledge and understanding in the realm of digital humanities.