



IDAKS 2016

Workshop of the Institute of Document Analysis and Knowledge Science

19-21 september 2016
La Rochelle, France

UNIVERSITÉ
La Rochelle

3i Informatique
Image
Interaction
Univ. La Rochelle

la
Charente
Maritime

Communauté
d'Agglomération de
La Rochelle



Table of Contents

Document forensics - a set of recent contributions	3
<i>by Jean-Marc Ogier and Nicolas Sidère, L3i Laboratory, University of La Rochelle – France</i>	
Comics Image Analysis Based on Various Elements	4
<i>by Motoi Iwata, Osaka Prefecture University – Japan</i>	
Historical document analysis and indexing	5
<i>by Mickaël Coustaty, L3i Laboratory, University of La Rochelle – France</i>	
Tapping into The Interplay Between Visual and Textual Scene Content to Improve Robust Reading Systems	6
<i>by Dimosthenis Karatzas, Computer Vision Centre, Barcelona - Spain</i>	
Graph-Based Visual Recognition and Indexing Based On Random Walk Embeddings	8
<i>by Josep Lladós, Computer Vision Center, Universitat Autònoma de Barcelona – Spain</i>	
ICDAR – an Important Community of Data Analysis with Reality	9
<i>by Seiichi Uchida, Kyushu University - Japan</i>	
More Dynamic Documents Imply More Dynamic Document Analysis Systems	10
<i>by Jean-Yves RAMEL, Polytech'Tours - France</i>	
Speeding up Research and Benefit in the Community	11
<i>by Fotini Simistira & Marcus Liwicki, University of Fribourg – Switzerland</i>	
Document-creator	12
<i>by Nicholas Jounet, LaBRI, University of Bordeaux – France</i>	
Augmented Document: A new challenge in document analysis field?	14
<i>by Jean-Christophe Burie, L3i Laboratory, University of La Rochelle – France</i>	
Toward Smart and Affective Documents	15
<i>by Olivier Augereau, Osaka Prefecture University – Japan</i>	
Activity Recognition and Intentional Intervention for a Human Behavior Change	16
<i>by Yutaka Arakawa, Nara Institute of Science and Technology – Japan</i>	
Analysis of Reading for Behavior Change	17
<i>by Koichi Kise, Osaka Prefecture University – Japan</i>	

Document forensics - a set of recent contributions

by Jean-Marc Ogier and Nicolas Sidère, L3i Laboratory, University of La Rochelle – France

Document engineering is the area of knowledge concerned with principles, tools and processes that improve our ability to create, manage, store, compact, access, protect and maintain documents, ...

The fields of document recognition and retrieval have grown rapidly in recent years.

Such development has been fueled by the emergence of new application areas such as forensics for document examination.

Indeed, the explosion of the digital contents production and manipulation tools offers a lot of opportunities for criminal organizations to create, violate steal, fraud digital documents, in the context of intellectual property protection, privacy protection, military application, ...

This talk will address some recent developments in the area of Document forensics and security, and will give an overview of our developments in terms of watermarking, automatic fraud and forgery detection (based on graphic and semantic analysis), and semantic hashing technique.

Comics Image Analysis Based on Various Elements

by Motoi Iwata, Osaka Prefecture University – Japan

Comics is a widespread cultural expression all over the world. Comics is not only an important industry but also the complex type of document images combined with text and other visual information in order to narrate a story. In industrial aspects, the market of comics is about 4 billion EUR in 2015 only in Japan, where the market of printed comics and digital comics are 3 and 1 billion EUR, respectively. In academic aspects, comics images are worth investigating because a comics page includes various elements, for example, panels, speech balloons, captions, leading characters, and so on. Moreover, the styles of drawing are also various depending on authors. Here, I would like to investigate manga image analysis based on not only drawing but also texts, effects, shape of speech balloon, and so on, where manga is Japanese style comics mainly drawn in grayscale. In addition to them, I will use the information obtained from readers, for example, eyegaze or the information by physiological sensors, where it is a collaborative work with Olivier.

Historical document analysis and indexing

by Mickaël Coustaty, L3i Laboratory, University of La Rochelle – France

Graphical part of historical documents are quite difficult images to process, and many of them are falling into decay. All these images are important to retrieve historical information but also to identify printers or copies.

A lot of works have been done on these peculiar images, with many dedicated features based on structural or statistical representation. Finding a proper representation of objects is a key issue in pattern recognition and document analysis. Common ways of object representation generally rely on the use of statistical pattern recognition techniques (like features vectors) to summarize the radiometric content of an image on one hand, or on the use of structural pattern recognition approaches to summarize

the topological organization of image's content (trees, graphs, ...) on the other hand. Most of the recognition systems are limited to work with a statistical representation, mainly due to the need of computing distances between documents (feature vectors) or because of the necessity of finding a representative of a cluster of documents. When a numerical feature vector is used to represent the document, the structural information is generally discarded while the structural representation allows keeping it. Structural representations are generally more powerful in terms of their representational abilities. But none of them were completely efficient and this presentation will try to propose some new trends in order to mix these approaches and to get the best of these two worlds.

Tapping into The Interplay Between Visual and Textual Scene Content to Improve Robust Reading Systems

by Dimosthenis Karatzas, Computer Vision Centre, Barcelona - Spain

Textual information appears in a wide range of contexts, apart from traditional documents. Over the last five years, the computer vision community has turned its attention on reading systems that operate on images acquired in unconstrained conditions, such as scene images, video sequences, born-digital images, wearable camera, lifelog feeds, social media images, etc. As an indicator of interest of the community, at the time of writing, the Robust Reading Competition has more than 1,600 registered users and has received more than 4,300 submissions.

Text content is not always present in scene images, but when it is, it tends to be important for understanding the scene. Recent statistics from the large-scale Coco-Text dataset, primarily focused on urban scene imagery, indicate that about 50% of the images contain some textual information. When present, text offers high-level semantic information, not easily obtainable from analyzing the rest of the scene. As such, it requires an explicit and accurate recognition.

At the time of writing this summary, robust reading research has reached an important milestone, as the availability of large-scale scene-text datasets, combined with latest developments in the Deep Learning paradigm, enable for a holistic contextual reasoning between scene text and the rest of the scene contents.

In this presentation, I will discuss our ideas of developing a unified model where visual, textual and user information provide mutual context for the holistic interpretation of scene images.

In particular, I will present recent work on how to take advantage of the visual information to optimise the interpretation of textual content in the scene.

Many scene text understanding methods approach the end-to-end recognition problem from a word-spotting perspective and take huge benefit from using small per-image lexicons. Such customized lexicons are normally assumed as given and their source is rarely discussed.

In this presentation I will discuss a method that generates contextualized lexicons for scene images using only visual information. For this, we exploit the

correlation between visual and textual information in a dataset consisting of images and textual content associated with them. Using the topic modeling framework to discover a set of latent topics in such a dataset allows us to re-rank a fixed dictionary in a way that prioritizes the words that are more likely to appear in a given image. Moreover, we train a CNN that is able to reproduce those word rankings but using only the image raw pixels as input.

Graph-Based Visual Recognition and Indexing Based On Random Walk Embeddings

by Josep Lladós, Computer Vision Center, Universitat Autònoma de Barcelona – Spain

Graph-based representations are experiencing a growing usage in visual recognition and retrieval due to their representational power as complementary techniques to classical appearance-based representations. However, (sub)graph isomorphism computation for recognition or retrieval, especially in large graph databases implies a high computational complexity. In this talk we will present our recent work at the Computer Vision Center on the definition of contextual similarities between graph nodes to speed up the inexact (sub)graph matching. A very interesting property of the adjacency matrix A of any graph G is that the $(i; j)$ -th entry of n -th power of A denotes the number of walks of length n from node v_i to node v_j . Based on this, we can enrich node labels of graphs in terms of attributes representing incident walks at each node. These attributes represent the structural context of a node that allows to compute efficient matches. We will first present a graph hashing formalism applied to visual retrieval where a binary embedding is defined as hashing keys for graph nodes. Secondly, we will present a work in which contextual similarities between nodes are incorporated in the tensor product graph computation. We will finally illustrate the proposed methods in applications related to word and symbol spotting in document analysis, and other well-known graph databases representing real scene images.

ICDAR – an Important Community of Data Analysis with Reality

by Seiichi Uchida, Kyushu University - Japan

“Data science” is now one of the international buzz words and “data scientists” are treated as the sexiest and highest-paid job... Well, are we (DAR researchers) OK just doing researches toward better OCR performance? Isn't it a waste of our skill? I personally believe that DAR researchers are the most suitable for data science because they are practically dealing with realistic data, i.e., character patterns. Character patterns are somewhat similar to “model organism” in biology research and thus have many interesting and unique properties suitable for analysis. For example, character patterns are linear (i.e., drawn with strokes), nonlinear (i.e., deformable), categorical (i.e., symbolic), artificial (i.e., designable), artistic, linguistic (i.e., readable and legible), less semantic-gapped, multi-modal, fine-grained, contextual, high-dimensional, psychological, perceptual, temporal, combinatorial, universal, and massive. Now is the time of transferring our great realistic skills on character pattern analysis to other data science researches!

More Dynamic Documents Imply More Dynamic Document Analysis Systems

by Jean-Yves RAMEL, Polytech'Tours - France

As mentioned in previous IDAKS workshops, the Document Image Analysis research field is at a cross-roads of its evolution mainly because of the last evolutions of what we call Documents. With the last new electronic devices, documents are now more and more dynamic and interactive (multimodality, streams, video, tactile interface) bringing new applications and new types of problems (construction of new reading systems).

First, I would like to present the last projects and research, we are working on in Tours showing that future Document Analysis System should not deal only with static images but with more and more dynamic documents or streams of documents.

Then, I would like to present and discuss with you the impacts of these evolutions on the design of future Document Analysis System that also have to become more Dynamic. It will allow me to introduce some initial works we are exploring like incremental and interactive analysis systems, anytime and/or budgeted classification, active online learning that we try to use to elaborate new Dynamic Document Analysis Systems.

Speeding up Research and Benefit in the Community

by Fotini Simistira & Marcus Liwicki, University of Fribourg – Switzerland

Progress in Document Image Analysis Research is comparably slow due to the slow process of publications and presentations at conferences and workshops. The provision of algorithms as research outcomes is (if any) typically done by putting the source code into github repositories or sharing it elsewhere. While this is helpful for engineering experts, many researchers struggle during the installation and compilation of the code. In recent projects on historical document analysis we developed WebServices for faster exchange and usage of methods as research outcomes. These DIVAServices <https://diuf.unifr.ch/main/hisdoc/divaservices> will be one focus of my presentation. Future ideas on mutual community benefit as well as related research will be presented as well.

Document-creator

by Nicholas Journet, LaBRI, University of Bordeaux – France

At one point or another in his/her scientific activities, almost every researcher in the field of Document Image Analysis and Recognition (DIAR) had to face the problem of obtaining a ground-truthed (labeled) data-set. Indeed, in general, whatever the DIAR system is designed for, ground-truthed document images are needed to evaluate the performances of the system or for running training algorithms embedded in the system.

When facing this problem, many researchers try to use training/benchmarking datasets available on the internet. Some of these datasets are free, some are not. Some are still updated, but many are no longer. Some are easy to download, but for many, obtaining them quickly is not straightforward, in particular due to the restrictions linked to intellectual property. For this purpose, creating manually the ground-truth is widespread in the DIAR community.

This manual ground-truth creation remains a tedious and time-consuming task. This inherently limits the amount of ground-truthed data available. In some cases, such relatively small amount of data cannot be representative of the whole content of the dataset. Another drawback of manually (or semi-automatically) annotating real documents is that many of them are protected by copyrights, which raises many difficulties to make the experiments reproducible.

As an alternative to the creation of manually (or semi-automatically) ground-truthed documents, from the beginning of the 90's with the seminal work of Baird, some solutions were proposed to generate synthetic images that look realistic. The generation of synthetic document images enables to create (quickly and with lower human cost) huge masses of ground-truthed documents that can be used for both training and evaluating DIAR algorithms. Despite some reference works, in 2016, the number of papers dealing specifically with sophisticated synthetic document image generation processes is still confidential. This can seem paradoxal given the rising need of researchers for ground-truthed data. We believe that, if we propose a complete and generic solution for them to generate easily and quickly huge masses of semi-synthetic ground-truthed documents sharing similar visual features as real documents, we can greatly help DIAR researchers for training, fitting and evaluating the performances of their approaches.

During this talk I will present Document-creator (<http://doc-creator.labri.fr/>), a

software that is able to create synthetic document images and its associated ground-truth.

Augmented Document: A new challenge in document analysis field?

by Jean-Christophe Burie, L3i Laboratory, University of La Rochelle – France

For many years, research in the field of document analysis try hard to analyze different categories of documents (historical, cadastral, administrative, ...) in order to extract and understand their content. Most of the time, the methods have been developed in order to improve the measure of recall and precision

Nowadays, the technological advances have facilitated the emergence of new systems using augmented reality or virtual reality. Combined to such a devices, the document analysis offers new interesting fields of study in order to develop interactive applications. The new technologies can bring an added value to the paper document by turn it into augmented document linking the physical support with digital information.

However, to achieve this goal, the approaches in the field of document analysis must be adapted to fulfil new constraints. For example, responsiveness of the analysis methods is essential to provide to the user the digital information. So document analysis must be robust enough to give relevant information and must be done in real-time to allow fluid interaction.

Moreover, designing an augmented document system needs to consider other fields of study such as: human computer interaction, gesture recognition, ...

This talk will address some recent developments in the area of augmented document. It will give an overview of new interactive interfaces and the new challenges in the field of document analysis.

Toward Smart and Affective Documents

by Olivier Augereau, Osaka Prefecture University – Japan

We introduced recently a new way for analyzing and understanding the documents by the way we read them. By estimating the reader's comprehension, we can estimate which layers, fonts, etc. are more suitable to make the document more understandable. But this is a user dependent process. So a straightforward idea is to create 'smart documents' that can adapt themselves to the reader. For example, we are now able to estimate the TOEIC score of a reader after reading few pages. If the reader skill is weak, the document could change for a simpler version of the text.

The second part of the discussion I would like to have is about the analysis of document emotions. By using some physiological sensors (skin conductivity, heartbeat, skin temperature, breathing patterns...) it is possible to know which part of a document is stressful, funny, boring... This information will be useful to make a summary of the document, to give a feedback to the author to improve it, to make a new generation of recommendation systems... The document could also adapt itself to the reader's mood.

Activity Recognition and Intentional Intervention for a Human Behavior Change

by Yutaka Arakawa, Nara Institute of Science and Technology – Japan

This research is the future research topic based on activity recognition that we did until now. Sensors are spreading into various scene of our life. Based on the sensing results, various services and useful information are provided to us. Typical service is a traffic jam information. In the academic field, a participatory sensing becomes one of attractive researches. Both utilize the human as a sensor of the city. We, human, have already been involved into the fusion of AI system and society.

In order to realize a sustainable society, I believe that a human behavior change is one of the promising technique. So, the intentional intervention aims to prompt a human behavior change by giving a useful information to a target user.

In this presentation, I will introduce our big project about it. Our lab bought a social system itself. It is a one-way car sharing system composed of 3 EV cars, 2 parking places (each 3 slots), a mobile application for unlocking/locking a car, and management software on the cloud. The problem in this society is a bias of cars and demand matching. I try to solve a bias by using intentional intervention from the system to users.

Although how to control the human is a quite challenging research issue, I try to find some way to prompt behavior change through an activity recognition research. I hope to discuss about it in IDAKS meeting.

Analysis of Reading for Behavior Change

by Koichi Kise, Osaka Prefecture University – Japan

As a possible new research direction, I would like to discuss with you a topic I call “analysis of reading”. As I presented in the last IDAKS, we have been working on the research called “reading-life log”, which include analysis of human reading behavior. It is originally for knowing more about readers in terms of the quantity and quality of reading. This year I would like to take a step forward towards a direction of how to use the results of analysis. A final goal is to guide readers towards a direction a bit better in some sense, by changing his/her behavior.

An example field is “learning and education”. Needless to say, they are inseparable from reading. Thus by knowing reading, we can find what should be improved. An important question is how to convince readers to follow the recommendation. Another example is “wellness”, especially dementia and depression. It is known that their seriousness can be reflected to the change of the reading behavior. Again, an important point is to how to guide them towards recovery or how to make the progress of disease slow down.

I am also interested in other fields of application or different way of the use of analysis results for behavior change. I am happy if I could stimulate you to take part in the discussion around analysis of reading and behavior change.