**LABORATOIRE
BORDELAIS
DE RECHERCHE
EN INFORMATIQUE**

LaBRI

le 2 septembre 2019

**Report by Jenny Benois-Pineau
On the Habilitation Thesis by Jakub Lokoc
« Methods for content-based Interactive Retrieval**

Interactive content –based retrieval of visual information is a technology which is required in a wide range of applications in modern society. It is very demanding efficient methods mixing approaches from information retrieval and visual information analysis, description and classification.
The proposed Habilitus thesis contains a synthesis of research of the candidate in this field.
The manuscript is organised in introductory part and two parts split into 12 chapters presenting contributions of the author.

In the *Introduction* the author presents the context and the objectives of his work. He gives the panorama of his research: i) use of traditional hand-crafted features, ii) interactive video retrieval.
In Chapter 2, he presents the motivation of his work in the fielded of research in multimedia retrieval. The ever growing volumes of information in digital form appear, a large part of it being multimedia data. He very clearly outlines the requirements for multimedia retrieval systems: effectiveness, efficiency and the ergonomics of interfaces. The quality of the present thesis is witnessed by the attention the author pays to the formalisation of the research questions addressed, shortly describing relevant formalisms and introducing quality metrics which are common to quantify the performance of MM retrieval systems.
A short summary of vector spaces is given as the majority of MM search and retrieval methods are based on nearest searches in the description space which is a metric vector space. He also summarizes two complementary parts of MM retrieval approaches: metadata search and content based search. Finally indexing data structures are introduced which is necessary for fast MM search in large databases.

The *Part 1* of the manuscript contains the description of the author's contributions in Models Based on feature signatures and in indexing structures and approaches to optimize the query time.

In *Chapter 3* he gives motivations for this research conducted in the framework of two projects: malware detection and 3D shape search. He also summarises main contributions. The feautre-based approach is justified in both cases. The signature of objects is presented as an « adaptive histogram ». In literature of the period when this research has been conducted, such kind of histograms is called Bag-of-Words (BOW) or in case of visual information indexing and retrieval, Bag-of–Visual Words, (BOVW). The author does not stress this. What is the reason ?
He defines the feature signatures as weighted histograms, this is also known as « Soft assignment » - it would be nice to find this mention too.
The main contribution of the author consists in using and adaptive distance he calls Signature Quadratic Form Distance (SQFD). He explains its advantages, its efficiency and also the adaptability for parallel computing and finally presents efficient retrieval approaches.
Chapter 4 is devoted to the contribution of the author which is the approximation of the SQFD using Scalable Feature Signatures. This is a conference paper at MMM'2014 which was then extended in Multimedia Tools and Applications ACM Journal. Scalability of feature signatures is an important research question which was also intensively studied specifically when the images were represented on the scalable basis, e.g. Daubechies wavelets in JPEG2000 compressed form. Having worked on the subject the reviewer

can appreciate this contribution, where the scalability is understood in the view of order relation defined on feature signatures.

In *Chapter 5* the author is interested in indexing on Metric spaces using Cut regions. This chapter is a paper published in journal « Information Systems ». The cut regions represent compact metric regions and the author shows that they are more efficient for indexing than ball regions. Without being a specialist in the matter, the reviewer thinks that this contribution is original and well-formalised.

Chapter 6 represents the next contribution of the author such as D-Cache. This memory structure contains hashed distances from previous queries. It allows for acceleration of search and filtering of non-relevant objects. The chapter is a paper in IEEE Transactions on Knowledge Engineering. The distances computed in previous queries are stored in the cache and can serve as approximation for current distance. The good property is that the proposed approach is generic and adapted for all metric queries.

In *Chapter 7* the author is interested in Ptolemaic properties of his proposed adaptive distance SQFD. This means that this distance satisfies metric properties together with Ptolemaic inequality. The latter stands that for any quadrilateral, the pairwise products of opposing sides sum to more than the product of the diagonals. The author uses the Ptolemaic inequality for distance-based indexing to construct a pivot-based lower bound applied to introduced metric SQFD. The comparative study of triangle lower bound estimation and Ptolemaic estimation is quite interesting. The paper in Information System journal is very well written and experimental results on comparison of Ptolemaic indexing with other kinds of indexing are impressive.

Part 2 of the manuscript is devoted to the second subject of research of the author, namely Interactive video retrieval.

In *Chapter 8* the author presents a summary of his contributions in interactive video retrieval, the research which was confirmed by large-scale experiments in the framework of TRECVID international challenge. The reviewer has been a leader of competing team in TRECVID challenge in different tasks, such as « visual summary » and « instance search » and is aware what an effort is required to produce competitive systems for TRECVID tasks. We can but congratulate the candidate for such a validation of his research. The author has participated in the task such as « Known-item search » and « Ad-hoc search ».

Here he uses more modern feature extractors such as CNNs and formulates the problem as similarity search. He convincingly shows that the automatic simulated search is not yet able to bring satisfactory solution and interactive search with query re-formulation is required. A valuable contribution represents the developed interactive search system VIRET.

In this chapter he also presents popular methods for interactive video retrieval. He starts with video pre-processing referencing three main problems: shot boundary detection, representative key-frame selection and scene segmentation. Quite a large literature has been devoted to the subject, it would be nice if the author could reference research of French community which has been intensively participating in TRECVID competition on these subjects in the early 2000ties. The author also comprehensively presents the initialisation of interactive search with an initial query and refinement by the users who browse the first response pages. It has to be noted, that in his approach the author used not only visual, but multi-modal queries, which improves results.

A mandatory part of interactive search systems is an user-friendly visualization interface. A part of this chapter is devoted to the analysis of state-of-the-art solutions. Having worked on graph-based visualization of video with key-frames, the reviewer can appreciate the contribution of the author in 3D visualisation interfaces.

Relevance feed-back in interactive search system is a standard mechanism to refine the query. The author has developed an interactive video retrieval system for mental visual browsing inspired by Bayesian framework and used it in Video Browser Showdown 2016 campaign.

To evaluate interactive video retrieval systems, they have developed a simulation of image retrieval task based on a sketch. A good analysis of protocols of qualitative and quantitative evaluations « post-task » is presented. The experience with real users remains in perspective. The evaluation of video retrieval was performed in Video Browser Showdown campaign, where the team of the author got a second place.

Chapter 9 is the author's paper in IEEE transaction on multimedia where an international team analyses the trends in Interactive Video Retrieval.

In *Chapter 10*, which is a paper in ACM Transaction on Multimedia computing the author and other participants of international team report on the experience of 2018.

Finally in *Chapter 11*, the developed Interactive video retrieval tool VIRET is presented in the TRECVID task of known-item search.

Chapter 12 concludes the thesis and analyses proposed contributions.

In general, the manuscript present a rich and deep research work of the candidate with several contributions, all proposed methods are well formalised and experimentally evaluated in world-known campaigns, such as TRECVID and Video SowDown Search. The manuscript is well organised and perfectly written.

Taking all this in consideration the reviewer concludes that M. Jakub Lokoc perfectly merits the grade of Habilitation.

Jenny Benois-Pineau
Professor,
Knight of Academic Palms