ERCIM "ALAIN BENSOUSSAN" FELLOWSHIP PROGRAMME

Scientific Report

First name / Family name	Umair Ali Khan
Nationality	Pakistani
Name of the *Host Organisation*	Fraunhofer – Institute for Integrated Circuits, Erlangen, Germany.
First Name / family name of the *Scientific Coordinator*	Heiko Sparenberg
Period of the fellowship	01/11/2016 to 30/09/2017

I SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

My research activity at Fraunhofer IIS – Institute for Integrated Circuits (Erlangen, Germany) – has been focussed on exploring the current Machine Learning (ML) techniques, open-source frameworks and tools for movies understanding and extracting key information therein. In the first phase, I highlighted the pros and cons of the existing techniques and performed the preliminary experiments with the currently available open-source libraries, e.g., Tensorflow, Caffe, Theano, etc to analyze their respective performance. This helped to identify the major challenges pertaining to dataset collection, reducing data size, training, building an appropriate hardware and performance evaluation. At the same time, the major focus was given on exploring the potential of a popular type of ML technique, called Deep Learning (DL), for retrieving key information from movies.

I focused on directing the knowledge gained during the first phase to develop a DL framework for classifying a static movie scene. This framework was further extended to automatically extract a compact set of key tags from a movie and subsequently segmenting the movie with respect to the extracted tags.

The designed framework was necessitated by the fact that the sheer volume of movies produced these days poses a huge challenge to their manual processing. Human generated metadata is generally not sufficient to describe the main contents of a movie and/or is not accurate due to the difficulty associated with precise information recall. This requires an intelligent video analysis to automatically extract the salient information from movies. This information can be utilized in a number of tasks including search optimization, scene-driven retrieval, object detection, translating movies to natural language, event detection, action recognition, behavior recognition, recommendation systems (to name a few).

The main contributions of my scientific activity during my ERCIM fellowship include the following.

1. I developed a tag vocabulary and collected a dataset of more than 35000 images for recognizing different scenes in a movie. To the best of my knowledge, no such dataset has ever been developed. It's unique in the sense that the images, representing the individual movie scenes, were collected by a rigorous manual analysis.
2. I trained a convolutional neural network using transfer learning to recognize a given scene of a movie presented as a static image/frame and mapping it to a particular tag in the tag vocabulary.
3. I developed an algorithm that first finds the shots boundaries in a movie and then extracts the key frame that are the representative of individual shots. This significantly reduces the data to be processed and eliminates redundancy.
4. I further developed an algorithm that applies the trained model on individual key frames and summarizes the extracted tags in the form of a compact set which best describes a movie.
5. I further extended the algorithm to segment a movie with respect to the extracted tags. This helps to retrieve the desired scenes/parts of the movie without watching the whole movie.
6. Other contributions include software development of prototypes, documentation, evaluation and field-tests of the developed algorithms

The aim of this research was not only to understand the high level semantics in each frame of the movie, but also to identify a compact set of the movie's representative topics. Retrieving this information further helps in movies classification, context-based search, efficient archiving and content-censorship (e.g, violence, sex and nudity in kids movies).

My developed framework has the following striking features: (i) it works at a higher

semantic level by understanding the overall context in the individual movie frames. The context represents the interaction of the objects in a scene and their overall meaning. The examples of context include romance, violence, fight, action, etc, (ii) it is different from typical event or scene recognition algorithms which attempt to recognize an item belonging to a single event or scene, (iii) it also stands apart from most object recognition algorithm which label everything visible in an image. This will produce thousands of labels for a movie without providing its thematic points, and (iv) this framework does perform, but it is not limited to, genre classification of movies. A movie typically has 2-3 genres which do not reveal other information in the movie (e.g., violence, nudity, sex, etc). My carefully designed vocabulary adequately covers the main theme of a movie and is flexible to scalability.

II PUBLICATION(S) DURING YOUR FELLOWSHIP

1. U. A. Khan, E. Naveed, A. M. Amor, H. Sparenberg, "Movies Tags Extraction Using Deep Learning", 14th IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS), 29 August - 1st September 2017, Lece, Italy.
2. E. Naveed, U. A. Khan, A. M. Amor, H. Sparenberg, “Deep Learning based Beat Event Detection in Action Movie Franchises”, 10th International Conference on Machine Vision (ICMV 2017), 13 - 15 November 2017, Vienna, Austria.
3. U. A. Khan, E. Naveed, A. M. Amor, H. Sparenberg, “Automatic Segmentation of Movies Using Deep Learning” (to be submitted).

III ATTENDED SEMINARS, WORKSHOPS, CONFERENCES

Unfortunately, I could not attend and present papers in the conferences due to the illness of my mother and early resignation.

IV RESEARCH EXCHANGE PROGRAMME (REP)

I visited the Hungarian Academy of Sciences, Institute for Computer Science and Control (MTA SZTAKI) in Budapest, Hungary. from May 15 to May 19, 2017, under the supervision of Prof. Tamas Sziranyi. During the REP, I was able to exchange ideas with him and most of the members of his research group. Specifically, I learnt how to further refine the shot detection algorithm for the key frames extraction. I also gave a seminar entitled “Movies tags extraction using deep learning”, in which I explained my current research challenges, and opened a discussion session with attendees where I also got several ideas.