



ERCIM "ALAIN BENSOUSSAN"
FELLOWSHIP PROGRAMME



Scientific Report

First name / Family name

Naveed Ejaz

Nationality

Pakistani

Name of the *Host Organisation*

Fraunhofer Institute of Integrated
Circuits

First Name / family name
of the *Scientific Coordinator*

Heiko Sparenberg

Period of the fellowship

15/07/2016 to 14/02/2018

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

During the fellowship, the major focus was develop state of the art machine learning based solutions for the task of movies understanding. In the first phase of fellowship, a thorough survey of the existing Object detection and activity recognition techniques in videos was conducted. At the same time, different open source deep learning frameworks were explored which included Tensorflow, Caffe, and Theano. As a result of the survey and continuous discussions in the group, it was decided to work on Beat Event detection in Hollywood movies for the rest of the fellowship.

The beat event detection is one component of a big goal of automatic understanding of the movie content. The automatic understanding of the movie contents is important for movie recommendation systems, efficient archiving and retrieval, content censorship, and scene driven retrieval systems etc. A movie can be temporally sub-divided into a hierarchy of acts, scenes, shots, and frames. In a movie, a “beat” refers to the punctual changes in the storyline. The frames of a video are easily available and efficient techniques for shot detection exist in the literature. However, grouping shots into scenes is a much harder problem. Similarly, from the state of art in computer vision, detecting the semantic level of beats and acts is not possible. Therefore, the problem of video understanding can be effectively addresses at the level of beat events which are a combination of consecutive shots having the same event.

There was an existing data set called “Action Movie Franchises”, developed at INRIA(France), which included annotation of 20 Hollywood action movies, each of which belonged to five famous franchises: Indiana Jones, Lethal Weapon, Die Hard, Rocky, and Rambo. The eleven beat event categories identified in the data set included: (1) Pursuit, (2) Battle Preparation, (3) Battle, (4) Romance, (5) Despair Good, (6) Joy Bad, (7) Good Argue Good, (8) Good Argue Bad, (9) Bad Argue Bad, (10) Victory Good, (11) Victory Bad.

A deep learning based scheme was developed for beat event detection on “Action Movie Franchises” dataset. An appropriate training dataset was constructed for each of the 11 beat categories after adjustment to some of the categories. The developed deep learning model used visual features in the key frames of the video to categorize each shot into one of the pre-defined categories which ultimately led to detection of beat events. Despite not using audio, script and motion information, the proposed scheme outperforms the state of the art on "Action Movies Franchises" data set. The proposed technique is flexible for adding new beat categories and can be easily tailored to other genres of movies apart from action movies. The proposed method is also open for addition of other modalities including semantic audio analysis and text analysis for improving the performance.

Along with the team in the group, I also partly worked on the problem of extracting tags from movies. The primary purpose of this work was to extract the key information from a movie and summarize it into a few key tags, representing the overall theme of the movie. The aim 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). A deep learning solution was developed for this task. In this work, I was involved in collecting the data set and assisting in conducting the experiments for subjective evaluation.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

1. "Movies Tags Extraction using Deep Learning",14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Italy, 2017.
2. "Deep Learning based Beat Event Detection in Action Movie Franchises", 10th International Conference on Machine Vision (ICMV), Austria, 2017.

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

1. 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Italy, 2017.
2. 10th International Conference on Machine Vision (ICMV), Austria, 2017.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

During my fellowship, I visited Spatio-Temporal Activity Recognition Systems(STAR) team in INRIA, Sophia Antipolis France on my research exchange program of one week. The REP was done between 15.05.2017 and 19.05.2017. The one week in STAR team INRIA was spent under the supervision of Dr. François BRÉMOND. During this short stay at INRIA, following activities were conducted:

1. A seminar was delivered for introducing my work to people in INRIA in general and STAR team in general. After the seminar, the follow up discussions were done with the head and senior members of the team. Some useful observations and suggestion were made by the host team.
2. Technical meetings with the head of team, post doctorates and PhD students were arranged. The meetings were mostly focused at presenting the work done in INRIA and the general discussion on the area of research.
3. Attending the team's quarterly meeting and getting to know their work style.
4. Meeting with the heads of other teams for getting know-how of their work and seeking feedback and suggestions for our work.