UNIVERSITÉ DE RENNES 1

Direction de la Recherche et de l'Innovation

Appel à projets - Année 2022

Contrats doctoraux «Ordinaires»

ECOLE DOCTORALE : MATH-STIC **POLE DE RECHERCHE : Mathématiques-Numérique**

Unité de recherche d'accueil

Unité de Recherche : IRISA

Equipe : Mimetic (en attente de la création de VIRTUS)

Encadrement

Directeur de thèse : Marc CHRISTIE **Unité de recherche** : UMR IRISA 6074

Employeur/Établissement d'origine : Université de Rennes 1

HDR : ☐ oui ☐ non ☑ en cours (soutenance Avril 2022) **Nb de thèses en cours** : co-supervision sur 3 thèses

Mail : marc.christie@irisa.fr **Tél.** : 06 50 01 29 22

Co-directeur de thèse : Vicky Kalogeiton **Unité de recherche** : LiX, Ecole Polytechnique

Employeur/Établissement d'origine : Ecole Polytechnique

HDR : ☐ oui ☑ non **Nb de thèses en cours** : co-supervision sur 2 thèses

Mail : vicky.kalogeiton@polytechnique.edu **Tél.** :

Sujet

Titre : Better understanding movies: towards high-level features for content and structure analysis and synthesis

Mots clés : Video analysis, film style, deep learning, visual transformers, computer animation

Cotutelle internationale (~~oui~~/non) : **Si oui, préciser l'établissement et le pays de rattachement** :

Descriptif (15-30 lignes) :

This thesis proposal lies at the intersection of computer vision field and computer graphics fields, and relies on deep learning technologies. The project aims at developing new methods for understanding the contents and high-level filmic structures behind video scenes. It will consist in first identifying the filmmaker's decisions (what to portray, and how to portray it), building a *director signature* that would include aspects of film genre (western, action, romance,...), film style (framing/editing/image depth/camera static/dynamic behavior) and narrative construction (temporal arrangement of shots and scenes). Second it will consist in exploiting the filmmaker's decisions (or director signature) to compose and create novel sequences (typically using 3D contents).

In essence, this thesis proposes to extract and formalize key cinematographic features at a degree that would enable understanding the content, style and structure of a movie and to exploit it in interactive 3D applications. In more details, our goal is to develop methods for understanding the *decisions behind the scenes from the filmmaker's perspective*, and we aim at understanding and transferring these director choices to new contexts. It will consist of two tasks.

First, we will focus on *movie-style and structure understanding*, for instance, by examining whether directors are using the same style patterns for their movies. We will consider several cinematographic features that we recently explored, such as frame composition, speed of speech, dialogue duration, interchange of dialogues, colors, lighting, shot length, soundtrack analysis. Following our previous contributions, we will develop pipelines to extract these high-level features and learn how to combine them to reveal the particularities of each director, and build *film signatures* that can enable high-level identifications and comparisons.

Second, we will focus on *creating novel cinematographic sequences of specific styles*, as an attempt to capture the director's staging as a particular intention at a specific instant. Specifically, we are interested in understanding how to rely on the film signature to re-design a sequence with different contents, different structures or different style. For this, we will use automatic editing techniques and cast the problem as a next-shot prediction task (from current scene analysis, current shot, and high-level characteristics, we will determine the optimal next shot). This would be useful for interactive videos (the audience could control the camera on the fly) and could be a step towards better content-based recommendation systems.

For these tasks, we will exploit recent advances in multimodal transformers by leveraging cinematographic conventions. Indeed, transformers have shown their capacity to generalize structures on from multimodal contents (see recent contributions in classification, summarizing and generative tasks). In addition, we have also demonstrated how strongly cinematographic conventions are tied to elements of structure, content and style. The proposed thesis will therefore build on these recent contributions to improve the computational understanding and exploitation of cinematographic contents.

➢ **Ce document est à renvoyer par la direction de l'unité de Recherche aux animateurs du Pole Recherche et à l'école Doctorale pour le 7 février 2022 au plus tard**