# Introduction

Retrieval of multimedia information has become an important topic since the development of the Internet has greatly expand the types of information that we transmit over the web. Users send numerous images, audios and videos on social media to share about their life, while researchers do a lot of work to find a way to collect, index and analyze these data, and eventually extract semantic meaning and information from the data collected. Methods for multimedia information retrieval can be generally categorized into three types:

a) methods for multimedia information classification.

In the context of multimedia information classification problems, we take the example of classification of unbalanced data, which is an important research problem because many real-world datasets have skewed class distributions in which the majority of data instances (examples) belong to one class, and far fewer instances belong to other classes. Although in many applications a small number of instances actually represent concepts of interest (e.g. niche hobbies, niche places to visit, etc.), classifiers induced from unbalanced datasets are more likely to be biased towards the majority class and have poor classification accuracy for the minority class. Despite a large body of research work, imbalanced data classification remains one of the most challenging problems in data mining and machine learning, especially for multimedia data.

b) methods for filtering multimedia information.

Filtering multimedia content is an extensive process that involves extracting and modelling semantic and structural information and metadata about the content. The problem with multimedia content is that the information presented in any document is by definition multimodal. The properties of different types of media vary greatly in the way in which the format of the content is stored and perceived. There is no direct way to relate the semantic content of a video stream to the semantic content of an audio stream, unless this is done manually. A content model of the spatial and temporal characteristics of an object can be used to define the actions in which the object participates. This content model can then be filtered according to the user profile in order to refine the filtering of the content for effective ranking and relevance of the files.

c) methods for feature extraction and content description for multimedia information.

Feature extraction is motivated by the huge size of multimedia objects and their redundancy and possible noise. In general, two possible goals can be achieved by feature extraction.

1. Generalization of the media content. Methods of induction include in the audio domain, e.g., melting-frequency cepstrum coefficients, zero-crossing rate, short-time energy. In the visual domain, color histograms, such as MPEG-7 scalable color descriptors, can be used for summarization.
2. Patterns are detected by auto-correlation and/or cross-correlation. Patterns are recurring media blocks that can be detected by comparing media blocks on media dimensions (temporal, spatial, etc.) or by comparing media blocks to templates (e.g., face templates, phrases).

In our project, we have chosen four papers in related area. We plan to carry out intensive reading on the papers, figuring out the mechanism introduced respectively. We hope to obtain a general view of multimedia and social media data retrieval, as well as an inside view of the ideas from the selected papers.