Architecture for Frame-based metadata

Overview

This result is about the development of a generic layer of persistence for frame-based metadata, i.e. metadata which may have granularity up to the single video frame (or audio sample). In the project a prototype for storage of frame based metadata has been developed to solve the specific case of automated genre recognition. However, the result can be generalised, providing a valuable solution for all such tasks in which efficient storage and quick access to information on a frame-by-frame basis is required.

In depth description

Multimodal information retrieval techniques combine audio, video and textual information to effectively perform content-based analysis and retrieval of multimedia contents., In some tasks, such as shot detection and video segmentation, these information items may be related to each individual frame (or audio sample) of a content item.

The TOSCA_MP project developed a service architecture for extracting and storing multimodal, frame-based features as illustrated in Figure 1. The architecture is multimodal in the sense that it computes a pattern vector to represent the set of features extracted from different media channels included in audiovisual contents. More in detail, these include low-level color, texture and motion features (F), shot-boundary features (S), acoustic features (A) and face-based features (C). The extracted features are saved within the Redis database[5], an in-memory, advanced key-value store system. As an example, Figure 2 shows how storing shot-boundary information is actually performed. Firstly, the output from shot clustering is stored frame-by-frame in the Redis server through the DRF, using the Redis list management interface. An XML containing the information to these data through Redis (i.e. IP address and port number of the Redis server, and name of the list generated by Redis during the insertion process) is then created and stored in the DRF. Finally, at the retrieval stage (see Figure 3), the client application (i.e. any external service that needs to access the shot boundary data) requests this XML to the DRF, and then connects to the Redis database for reading data at either global (i.e. complete content) or local (i.e. segment-based content) level.

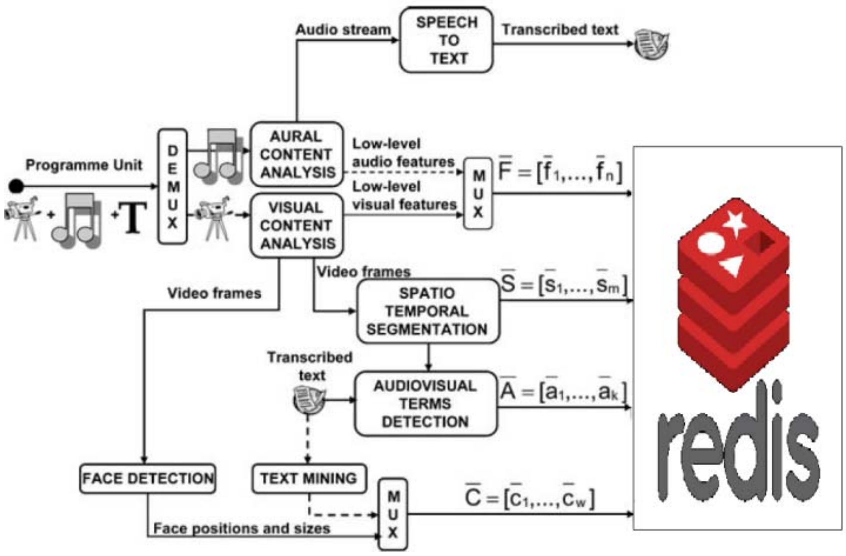

Figure 1. Architecture for multimodal frame-based feature extraction and storing.

[5] http://redis.io/

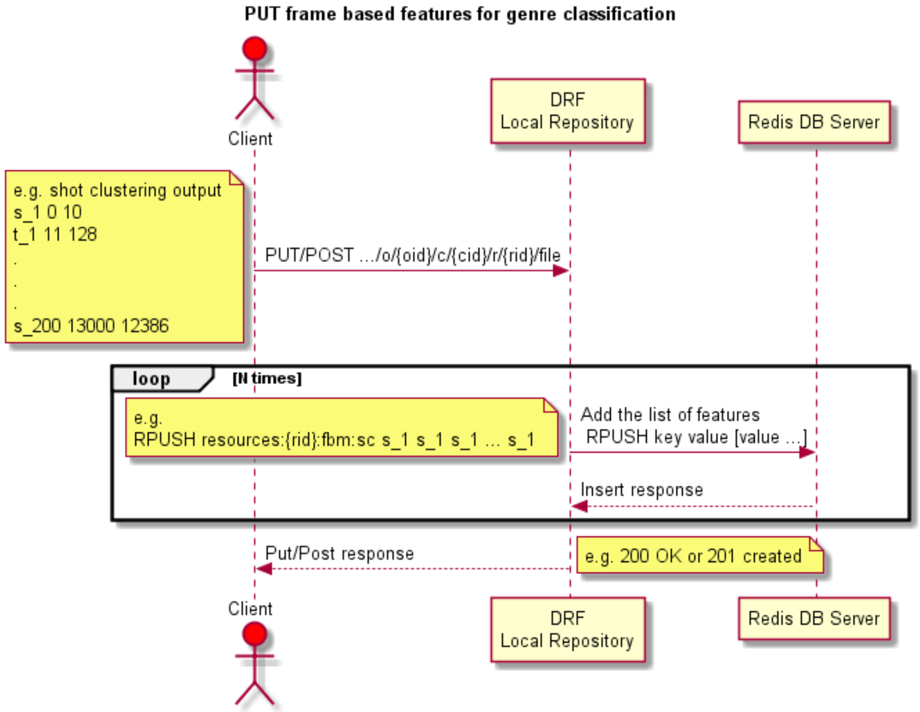

Figure 2. Storing shot-detection information with Redis.

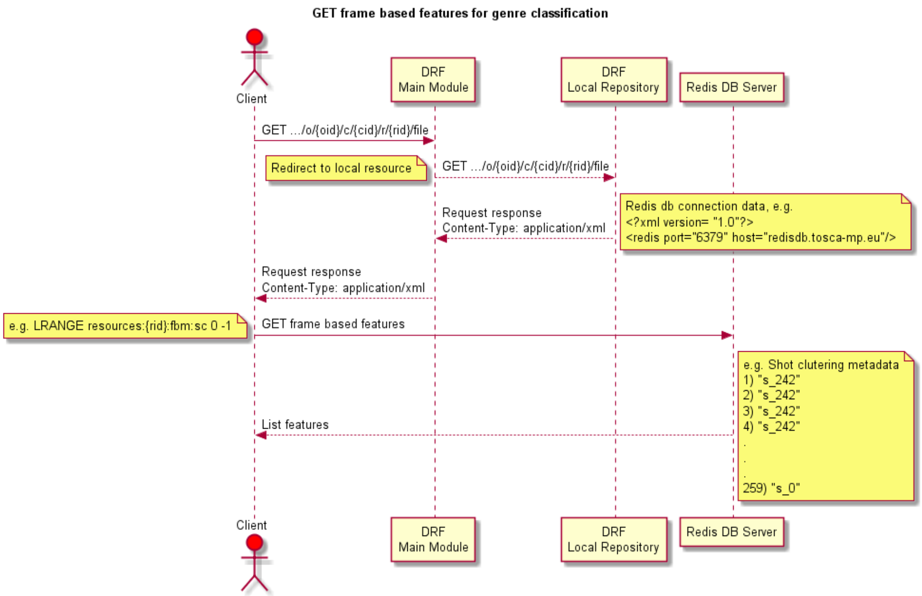

Figure 3. Retrieving shot-detection information with Redis.

Potential fields of Application

The result can be applied in all cases in which frame-based metadata of content are needed, and it is needed to store and access those metadata efficiently and in a scalable manner for different kinds of applications.

Possibilities for exploitation

Exploitation of the result is on a bilateral licensing agreement basis. Licensing information is still to be defined by the owner. Some software integration may be needed to ensure generalisation of the applicability of this result to all possible frame-based metadata cases.

Further Information

Further technical information is available in TOSCA-MP confidential Deliverable D2.2 “Automatic Metadata Extraction and Enrichment”.

Contact Person

Dr. Alberto MESSINA

RAI – Radiotelevisione Italiana - Centre for Research and Technological Innovation

Corso Giambone, 68, I-10135

Torino – ITALY

alberto.messina@rai.it