Finding single media documents that are relevant to a query involves several stages: data/knowledge extraction (computing appropriate descriptors of the content, possibly showing application-specific invariances), followed by indexing (structuring the extracted data to improve the efficiency of retrieval) and retrieval (the query of the user is analyzed and possibly reformulated so as to query the database in order to decide which results are relevant and then to present these results to the user). Each of these stages becomes a challenge when very large volumes of data have to be processed. One has to explicitly consider "large-scale" processing. While in the data/knowledge extraction stage one only has to deal with the large volume of data, in the subsequent stages one also has to take into account the size and complexity of the descriptors, as well as increasing external requirements (e.g. a very large number of classes for supervised learning). By its very nature, a multimedia document (web page, video, etc.) is characterized by heterogeneous and complementary multimodal data. This can be textual data, digital content from a video or audio channel (speech, music), data concerning document structure (spatial layout, temporal relations), etc. The multimodal aspect brings in both opportunities and new challenges. The opportunities come from the possibility of jointly using several media that complement each other. The new challenges concern both the increase in complexity of all the stages described above and the difficulty of defining methods for the effective cooperation of several media. Goals and existing work Among all existing difficulties that have been identified in efficiently managing multimedia data, this thesis will focus on the following points: - Efficient combination of multimedia information: although the extraction within each single medium is still a rather open problem \cite{Ballas2011jmre,Popescu2011,shabou.12.cvpr}, we will focus here on the effective combination of information extracted from several media. On this subject specifically, the CEA LIST has already proposed efficient late fusion algorithms \cite{znaidia.12.icmr}. We also developed a promising early fusion scheme, resulting in an effective system with relatively compact signatures \cite{znaidia.12.icpr}. Knowledge extraction from multimedia documents is a multi-step process. To deal with large scale databases, one must determine the relative importance of each of these steps in order to simplify the whole chain or optimize the most expensive ones. The goal is to find an appropriate compromise between effectiveness and efficiency, for each media considered (image, text,...) and for their combination. - Efficient large-scale search: multiple issues are identified to search and classify images and videos in large scale databases \cite{crucianu12visualscalability}. One may distinguish between multimedia retrieval and classification \cite{perronnin12goodpractice}. Multimedia retrieval consists in finding all documents from a database that respond to a given query (text \cite{information retrieval}). Recognizing an object (or a class of objects) is rather solved by supervised or unsupervised (text \cite{clustering}) classification. Some methods can reduce the "naive" complexity of algorithms (for a base of $N$ elements, the complexity is of order $O$ (N) for retrieval and usually $O$ (N^2) for classification), including the use of hash functions \cite{hashfunctions} and research in a compressed domain signature \cite{hashfunctions}, or effective learning algorithm \cite{perronnin12goodpractice}. Particular attention will be paid to the evaluation of the developed methods. This is not only of first importance regarding the scientific community (since required to publish in the best conferences and journals in the field), but also of interest for the industry (benchmarking). Moreover, public agencies have a growing interest in these approaches. A privileged framework for conducting such assessments consists of using data from international campaigns \cite{imagecampaigns,webcampaigns}. Although actual participation to such campaigns is not strictly required for the thesis, it will be necessary to evaluate the work on such data. Alternatively, the scientific community commonly uses publicly available databases according to established protocols. Applications This work may be of interest for many projects and applications, including cultural heritage, video browsing, online selling, private data protection, and much more. Although the CEA LIST is involved in many European and national projects for which this thesis is of interest, the candidate will not have to directly participate to software development for these projects. The unique requirement of the thesis is to conduct research on the proposed topic and produce relevant scientific publications.
Both the CEA LIST and the CEDRIC have multiple international collaborations. For the CEDRIC we can mention the National Institute of Informatics (NII, Tokyo, Japan), the New Jersey Institute of Technology (NJIT, New Jersey, USA), or the German Aerospace Center (DLR, Germany).