Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 38 (2014) 100 – 107

Procedia Computer Science

10th Italian Research Conference on Digital Libraries, IRCDL 2014

# A distributed system for multimedia monitoring, publishing and retrieval

Giuseppe Becchi, Marco Bertini*, Alberto Del Bimbo, Andrea Ferracani, Daniele Pezzatini

*Università di Firenze – MICC, Firenze, Italy*

**Abstract**

In this paper we present a distributed and interactive multi-user system which provides a flexible approach to collect, manage, annotate and publish collections of images, videos and textual documents. The system is based on a Service Oriented Architecture that allows to combine and orchestrate a large set of web services for automatic and manual annotation, retrieval, browsing, ingestion and authoring of different multimedia sources. These tools can been used to create several publicly available vertical applications, addressing different use cases. Positive results of usability test evaluations have shown that the system can be effectively used to create video retrieval systems.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Peer-review under responsibility of the Scientific Committee of IRCDL 2014

*Keywords:* semantic multimedia annotation; SOA; multimedia retrieval

## 1. Introduction

The explosion of digital data in recent times, in its varied forms and formats (H.264 and Flash videos, MP3 and Wav audio, HTML and PDF documents, images), requires the creation of effective tools to organize, manage and link digital resources, in order to maximize accessibility and reduce cost issues for everyone concerned, from content managers to online content consumers. On a larger scale, isolated multimedia repositories developed by content owners and technology providers can be connected, unleashing opportunities for innovative user services and creating new business models, in the vein of on-demand, online or mobile TV ventures. The system presented in this paper stems from above conditions and potentialities to connect publicly available multimedia information streams

* Corresponding author. Tel.: +39 055 2751395; fax: +39 055 2751390.
*E-mail address:* marco.bertini@unifi.it

1877-0509 © 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).
Peer-review under responsibility of the Scientific Committee of IRCDL 2014
doi:10.1016/j.procs.2014.10.017

under a unifying framework, which additionally allows publishers of audio-visual content to monetize their products and services. The backbone of the system, developed within two EU funded projects (IM3I and euTV), is a scalable audio-visual analysis and indexing system that allows detection and tracking of vast amounts of multimedia content based on Topics of Interest (TOI), that correspond to a user's profile and set of employed search terms. The front-end is a portal that can be customized to be suitable to different scenarios and scopes, displaying syndicated content, allowing users to perform searches, refine queries, and produce faceted presentation of results.

In this paper we present an interactive multi-user multimedia retrieval framework composed by: *i)* a SOA-based system that provides services for semantic and syntactic annotation, search and browsing for different media such as videos, images, text and audio that belong to different domains (possibly modeled with different ontologies) with query expansion and ontology reasoning; *ii)* web-based interfaces for interactive query composition, archive browsing, annotation and visualization. The usability of the system has been assessed in field trials performed by professional video archivists and multimedia professionals, proving that the system is effective, efficient and satisfactory for users. The paper is organized as follows: Sect. 2 briefly describes the state-of-the-art related to the main topics of this work, in Sect. 3 the architecture and the main functions of the framework are discussed; Sect. 4 reports on the usability assessment techniques used and on the outcomes of the field trials.

## 2. Related work

A thorough review of concept-based video retrieval has been presented by Snoek et al.[31]. The approach currently achieving the best performance in semantic concept annotation in competitions like TRECVid[29] or PASCAL VOC[18] is based on the bag-of-words (BoW) approach[30]. The bag-of-words approach has been initially proposed for text document categorization, but can be applied to visual content analysis[15,22,28], treating an image or keyframe as the visual analog of a document that is represented by a bag of quantized descriptors (e.g. SIFT), referred to as visual-words. A comprehensive study on this approach has been presented by Zhang et al.[40], considering the classification of object categories; a review of action detection and recognition methods has been recently provided by Ballan et al.[4]. Since the BoW approach disregards the spatial layout of the features, some researchers have proposed several methods to incorporate the spatial information to improve classification results[5]: Lazebnik et al[24] have proposed to perform pyramid matching in the two-dimensional image space, while Bronstein et al.[11] have proposed the construction of vocabularies of spatial relations between features. Regarding CBIR we refer the reader to the complete review of Datta et al.[17].

A web based video search system, with video streaming delivery, has been presented by Halversen et al.[21], to search videos obtained from PowerPoint presentations, using the associated metadata. A crowd-sourcing system for retrieval of rock'n'roll multimedia data has been presented by Snoek et al.[32], which relies on online users to improve, extend, and share, automatically detected results in video fragments. Another approach to web-based collaborative creation of ground truth video data has been presented by Yuen et al.[38], extending the successful LabelMe project from images to videos. A mobile video browsing and retrieval application, based on HTML and JavaScript, has been shown by Bursuc et al.[12]. An interactive video browsing tool for supporting content management and selection in postproduction has been presented by Bailer et al.[3], comparing the usability of a full-featured desktop application with a limited web-based interface of the same system.

## 3. The system

The system platform has a powerful infrastructure, that embraces different programming and scripting languages (C++, Java, PHP, JavaScript, LUA), required for the many tools and services, and libraries and open source software products (FFMpeg, Red5, Mule etc.) to process audio, visual and textual multimedia documents.

The projects scale has required a pragmatic and robust approach to integrate all components into a scalable platform. This goal has been reached by extensively using a services oriented approach for system integration, with an agnostic approach to languages and protocols.

This framework is based on a service oriented architecture (SOA), that allows different applications and clients to exchange data with each other on a shared and distributed infrastructure, permitting to build new digital library

applications on the base of existing services[10]. The framework has three layers (Fig. 1): analysis, SOA architecture and interfaces. The analysis layer provides services for syntactic and semantic annotation of multimedia data, using series of user-definable processing pipelines that can be executed on distributed servers. The SOA Architecture layer routes communications between interfaces and analysis services and provides the main repository functions. Finally, the interface layer provides applications and systems for manual annotation, tagging, browsing and retrieval. The platform has a central services-based architecture layer consisting of a file-store, a repository and a web services hub (Apache Mule). This layer acts as the central hub to which the media processor framework and the end-user interfaces could rely on for storing and retrieving data in a flexible manner (Fig. 1).

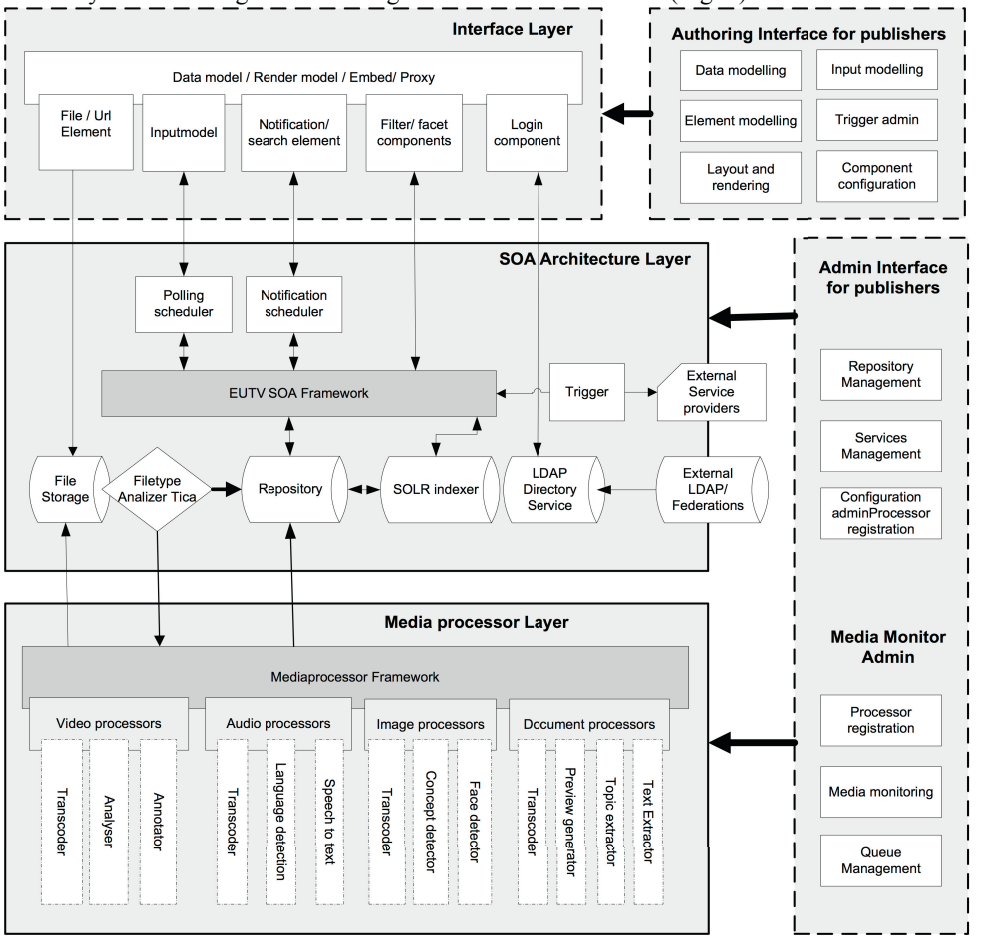


Fig. 1. The system architecture.

### *3.1. Media annotation and indexing*

Text indexing and annotation is performed by services that provide: *i)* language classification, based on n-grams and Naïve Bayes classifiers that despite the simplicity have shown to work effectively also on short fragments[39]; *ii)* topic detection based on LDA; *iii)* named entity extraction based on gazetteers and a rule-based system, to handle entities that have not been added yet to lists[16]. Topic detection and named entity identification can be used also with the outcomes of speech transcription services.

Speech and audio indexing and analysis tools provide services for: *i)* audio segmentation[25] that separates audio streams into six separate components: four for classification (speech/non-speech, gender, background and speaker

identification), one for speaker clustering and one for acoustic change detection. These components are mostly model-based, making extensive used of feed-forward fully connected Multi-Layer Perceptrons trained with back-propagation; *ii)* audio language identification, based on Abad[1], that identifies the 12 most spoken languages across the European Union, using SVMs for the phonotactic system and an I-vector based acoustic sub-system; *iii)* an audio event module that recognizes 54 sound concepts, using a combination of MFCC, ZCR and MPEG features to feed SVM classifiers; *iv)* audio transcriptions, an engine that uses an hybrid approach combining the temporal modelling capabilities of Hidden Markov Models with the pattern discriminative classification capabilities of Multi-Layer Perceptrons[26], and works with English, Spanish, Portuguese and German.

Visual annotation and indexing deal with images and videos at syntactic and semantic levels. Similarity-based retrieval deals with images and video keyframes, using a combination of MPEG global features (in particular have been used Scalable Color, Color Layout and Edge Histogram descriptors to capture different visual aspects with compact and low computational cost descriptors[7]) and SIFT descriptors, indexed using approximate similarity searching based on inverted files[2] for scalability. Semantic annotation is obtained using a BoW-based approach, following the success of this approach for scene and object recognition[19,28,37], with a model selection step to select the best combination of interest point detectors/descriptors (e.g. SIFT, SURF and MSER) for each concept classifier, and using the Pyramid Match Kernel kernel[20], that is robust to clutter and outliers, and is efficient thanks to its linear time complexity in matching. Classifiers can be trained with a specific service, that exploits social media as training source[34].

Each type of media has a specific component that can handle visualization, media search and manual annotation. A web-based authoring system allows to combine all the services to design specific applications for each use scenario, from the automatic ingestion of media to their processing, search and presentation.

### *3.2. The search engine*

The Orione search engine (preliminarly presented by Bertini et al.[6]) permits different query modalities (free text, natural language, graphical composition of concepts using Boolean and temporal relations and query by visual example) and visualizations, resulting in an advanced tool for retrieval and exploration of video archives for both technical and non-technical users. It uses an ontology that has been created semi-automatically from a flat lexicon and a basic lightweight ontology structure, using WordNet to create concept relations (is a, is part of and has part). The ontology is modeled following the Dynamic Pictorially Enriched Ontology model[8], that includes both concepts and visual concept prototypes. These prototypes represent the different visual modalities in which a concept can manifest; they can be selected by the users to perform query by example, using MPEG-7 descriptors (e.g. Color Layout and Edge Histogram) or other domain specific visual descriptors. Concepts, concepts relations, video annotations and visual concept prototypes are defined using the standard Web Ontology Language (OWL) so that the ontology can be easily reused and shared. The queries created in each interface are translated by the search engine into SPARQL, the W3C standard ontology query language. The system extends the queries adding synonyms and concept specializations through ontology reasoning and the use of WordNet. As an example consider the query "find shots with vehicles": the concept specializations expansion through inference over the ontology structure permits to retrieve the shots annotated with "vehicle" and also those annotated with the concept's specializations (e.g. "trucks", "cars", etc.). In particular, WordNet query expansion, using synonyms, is enabled when using free-text queries, since it is not desirable to force the user to formulate a query selecting only the terms from a predefined lexicon. As an example consider the case in which a user types the word "automobile": also videos that have been annotated using the word "car" will be returned. The search engine provides also services for browsing the ontology structure and to filter query results using metadata, like programme names or geolocalization information.

### *3.3. The web-based user interfaces*

The web-based search and browse system, based on the Rich Internet Application paradigm (RIA), does not require any software installation and it is composed by several integrated and specialized interfaces. It is complemented by a web-based system for manual annotation of videos, developed with the aim of creating, collaboratively, manual annotations and metadata, e.g. to provide geolocalization of concepts' annotations. The tool

can be used to create ground truth annotations that can be exploited for correcting and integrating automatic annotations or for training and evaluating automatic video annotation systems.

The Sirio semantic search engine is composed by two different interfaces (Fig. 2): a simple search interface with only a free-text field for Google-like searches and an advanced search interface with a GUI to build composite queries that may include Boolean and temporal operators, metadata (like programme broadcast information and geo tags) and visual examples. The advanced interface allows also to inspect and use a local view of the ontology graph for building queries. This feature is useful to better understand how one concept is related to the others, thus suggesting possible changes in the composition of the query.

Sirio has two views for showing the list of results: the first presents a list of four videos per page; each video, served by a video streaming server, is shown within a small video player and it is paused in the exact instant of the occurrence of a concept. The other view is an expanded list of results that can show, in a grid, up to thousands of keyframes of the occurrences of the concepts' instances. In both cases users can then play the video sequence and, if interested, zoom in each result displaying it in a larger player that shows more details on the video metadata and allows better video inspection. The extended video player allows also searching for visually similar video shots, using the CBIR search interface Daphnis.


Fig. 2. The Sirio web-based user interfaces: *left)* simple user interface; *right)* advanced user interface with ontology graph and geolocalization filtering.

Furthermore another interface (Andromeda), also integrated with Sirio, allows to browse video archives navigating through the relations between the concepts of the ontology and providing direct access to the instances of these concepts; this functionality is useful when a user does not have a clear idea regarding the query that he wants to make. The Andromeda interface is based on some graphical elements typical of web 2.0 interfaces, such as the tag cloud. The user starts selecting concepts from a tag cloud, than navigates the ontology that describes the video domain, shown as a graph with different types of relations, and inspects the video clips that contain the instances of the annotated concepts. Users can select a concept from the ontology graph to build a query in the advanced search interface at any moment.

## 4. System evaluation

According to ISO, usability is defined as the extent that a user can utilize a product effectively, efficiently and satisfactorily in achieving a specific goal. However, researchers in the field of usability have defined a variety of views on what usability is and how to assess it. Some[14] follow ISO standards on quality models (ISO 9126), others follow user-centered design (ISO 9241) or user-centered approaches[36]. Moreover, it has been observed that usability guidelines are often not comprehensive, e.g. large portions of recommendations are specific of each guidelines[9], and thus studies on usability of digital video library systems have used a variety of methods and guidelines[13]. The methodology used in the field trials of the system follows the practices defined in the ISO 9241 standard and of the guidelines of the U.S. Department of Health and Human Sciences[35], and gathered: *i)* observational notes taken

during test sessions by monitors, *ii)* verbal feedback noted by test monitors and *iii)* a survey completed after the tests by all the users. The usability tests have been designed following task-oriented and user-centered approaches[33,35]; in particular the test aimed at evaluating the main characteristics of usability (as defined in ISO 9126 and ISO 9241) that are: *i)* understandability, i.e. if the user comprehends how to use the system easily; *ii)* learnability, i.e. if the user can easily learn to use the system; *iii)* operability, i.e. if the user can use the system without much effort; *iv)* attractiveness, i.e. if the interface looks good; *v)* effectiveness, i.e. the ability of users to complete tasks using the system; *vi)* satisfaction, i.e. users subjective reactions to using the system. Before the trials, users received a short interactive multimedia tutorial of the systems, as in Christel et al.[13].

The questionnaires filled by the users contained three types of questions[23]: factual types, e.g. the number of years they have been working in a video archive or the types of tools commonly used for video search; attitude-type, to focus the respondent's attention to inside themselves and his response to the system; option-type, to ask the respondent what they think about features of the systems, e.g. what is the preferred search mode. The answers to the attitude-type questions followed a five point Likert scale questionnaire, addressing frequency and quality attitudes.

### *3.4. Web usability test*

The web-based interfaces have been tested to evaluate the usability of the system in a set of field trials. A group of 18 professionals coming from broadcasting, media industry, cultural heritage institutions and video archives in Italy, The Netherlands, Hungary and Germany have tested the system on-line (running on the MICC servers), performing a set of pre-defined tasks, and interacting with the system. These activities deal with the video components and were:

**Task 1**: Concept search. In this task users had to perform two search activities using the advanced and simple search interfaces. Each search activity was composed by sub-activities, e.g. using the ontology graph to inspect concepts' relations, filter results using metadata, etc. Each sub-activity was evaluated but, for the sake of brevity, we report only the overall results.

**Task 2**: Concept browsing. In this task users had to browse the ontology, searching for concepts and concepts' instances (i.e. video clips) related to "person".

**Task 3**: CBIR search. In this task users had to perform a semantic-based search for "face" or "person" using the advanced search interface and then search for shots containing an "anchorman" or "Angela Merkel" based on visual similarity.


Fig. 3. Web-based interfaces usability tests: usability of the advanced search and simple search interfaces in Task 1.

Test monitors have recorded observational notes and verbal feedbacks of the users; these notes have been analyzed to understand the more critical parts of the system and, during the development, they have been used to redesign the functionalities of the system. In particular, a search interface based on natural language queries has been dropped from the system because users found it too difficult to be used and not providing enough additional functionalities w.r.t. the advanced and simple search interfaces.

Fig. 3 reports evaluations for the advanced and simple search modalities used in Task 1. Fig. 4 reports results for Task 2 and 3. These tests were carried on only by 6 participants, which work as archivists in a national video archive. However, this number of participants is enough for such usability tests[27,35].


Fig. 4. Web-based interfaces usability tests: usability of the Andromeda and Daphnis interfaces in Task 2 and 3.

Fig. 5 summarizes two results of the tests. The overall experience is very positive and the system proved to be easy to use, despite the objective difficulty of interacting with a complex system for which the testers received only a very limited training. Users appreciated the combination of different interfaces. The type of search interface that proved to be more suitable for the majority of the users is the Sirio advanced interface because of its many functionalities that are suitable for professional video archivists.

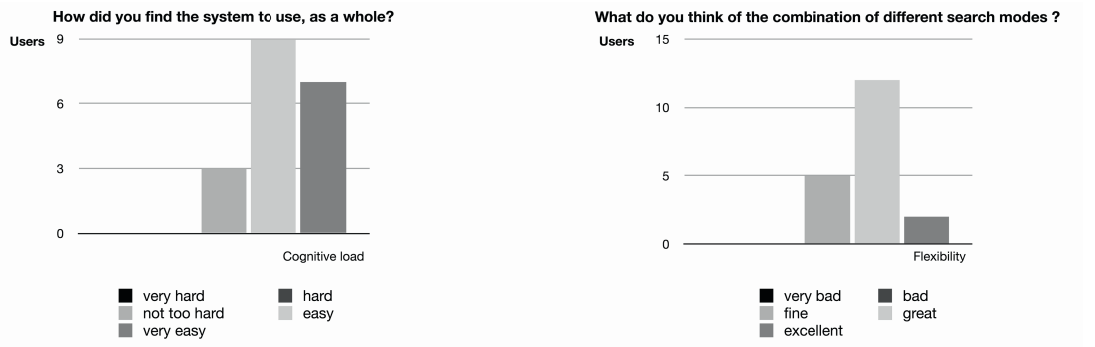


Fig. 5. Overview of web-based interfaces usability tests: overall usability of the system, usability of the combination of search modalities.

## Acknowledgements

This work was partially supported by the EU IST IM3I project (http://www.im3i.eu/ - contract FP7-222267) and EU IST euTV project (http://www.eutvweb.eu - contract FP7-262428).

## References

1. Abad, A.: The L2F language recognition system for Albayzin 2012 evaluation. In: *Proc. of Iberspeech* (2012)
2. Amato, G., Savino, P.: Approximate similarity search in metric spaces using inverted files. In: *Proc. of InfoScale* (2008)
3. Bailer, W., Weiss, W., Kienast, G., Thallinger, G., Haas, W.: A video browsing tool for content management in postproduction. *International Journal of Digital Multimedia Broadcasting* (2010)
4. Ballan, L., Bertini, M., Del Bimbo, A., Seidenari, L., Serra, G.: Event detection and recognition for semantic annotation of video. *Multimedia Tools and Applications* **51**(1), 279–302 (2011)
5. Behmo, R., Paragios, N., Prinet, V.: Graph commute times for image representation. In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition* (CVPR). pp. 1 –8 (2008)
6. Bertini, M., D'Amico, G., Ferracani, A., Meoni, M., Serra, G.: Sirio, Orione and Pan: an integrated web system for ontology-based video search and annotation. In: *Proc. of ACM International Conference on Multimedia (ACM MM)*. pp. 1625–1628 (2010)

7. Bertini, M., Del Bimbo, A., Nunziati, W.: Video clip matching using MPEG-7 descriptors and edit distance. In: *Proc. of International Conference on Image and Video Retrieval (CIVR)*. pp. 133–142. Tempe, AZ, USA (July 2006)
8. Bertini, M., Del Bimbo, A., Serra, G., Torniai, C., Cucchiara, R., Grana, C., Vezzani, R.: Dynamic pictorially enriched ontologies for digital video libraries. *IEEE MultiMedia* **16**(2), 42–51 (2009)
9. Bevan, N., Spinhof, L.: Are guidelines and standards for web usability comprehensive? In: Jacko, J. (ed.) *Human-Computer Interaction. Interaction Design and Usability, Lecture Notes in Computer Science*, vol. 4550, pp. 407–419. (2007)
10. Brettlecker, G., Milano, D., Ranaldi, P., Schek, H.J., Schuldt, H., Springmann, M.: ISIS and OSIRIS: a process-based digital library application on top of a distributed process support middleware. In: *Proc. of 1st International Conference on Digital Libraries: Research and Development (DELOS)*. pp. 46–55 (2007)
11. Bronstein, A.M., Bronstein, M.M.: Spatially-sensitive affine-invariant image descriptors. In: *Proc. of European Conference on Computer Vision (ECCV)*. pp. 197–208 (2010)
12. Bursuc, A., Zaharia, T., Prêteux, F.: Mobile video browsing and retrieval with the Ovidius platform. In: *Proc. of ACM International Conference on Multimedia (ACM MM)*. pp. 1659– 1662 (2010)
13. Christel, M., Moraveji, N.: Finding the right shots: assessing usability and performance of a digital video library interface. In: *Proc. of ACM International Conference on Multimedia (ACM MM)*. pp. 732–739 (2004)
14. Chua, B., Dyson, L.: Applying the ISO9126 model to the evaluation of an e-learning system. In: *Proc. of the 21st ASCILITE Conference*. pp. 184–190 (2004)
15. Chum, O., Philbin, J., Sivic, J., Isard, M., Zisserman, A.: Total recall: Automatic query expansion with a generative feature model for object retrieval. In: *Proc. of International Conference on Computer Vision (ICCV)*. pp. 1 –8 (2007)
16. Cunningham, H., Maynard, D., Bontcheva, K., Tablan,V. :GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In: *Proc. of ACL* (2002)
17. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys* **40**, 5:1–5:60 (2008)
18. Everingham, M., VanGool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The Pascal Visual Object Classes (VOC) challenge. *International Journal of Computer Vision* **88**(2), 303–338 (Jun 2010)
19. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2003)
20. Grauman, K., Darrell, T.: The pyramid match kernel: Efficient learning with sets of features. *Journal of Machine Learning Research* (2007)
21. Halvorsen, P., Johansen, D., Olstad, B., Kupka, T., Tennøe, S.: vESP: enriching enterprise document search results with aligned video summarization. In: *Proc. of ACM International Conference on Multimedia (ACM MM)*. pp. 1603–1604 (2010)
22. Jiang, Y.G., Ngo, C.W., Yang, J.: Towards optimal bag-of-features for object categorization and semantic video retrieval. In: *Proc. of ACM International Conference on Image and Video Retrieval (CIVR)*. pp. 494–501 (2007)
23. Kirakowski, J.: Questionnaires in usability engineering: A list of frequently asked questions (3rd ed.). Available at: http://www.ucc.ie/hfrg/resources/qfaq1.html (2000)
24. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. vol. 2, pp. 2169 – 2178 (2006)
25. Meinedo, H.: Audio pre-processing and speech recognition for broadcast news. *Ph.D. thesis*, IST, Lisbon (2008)
26. Neto, J., Meinedo, H., Viveiros, M., Cassaca, R., Martins, C., Caseiro, D.: Broadcast news subtitling system in Portuguese. In: *Proc. ICASSP* (2008)
27. Nielsen, J.: Why you only need to test with 5 users. Available at: http://www.useit.com/alertbox/20000319.html (2000)
28. Sivic, J., Zisserman, A.: Video Google: A text retrieval approach to object matching in videos. In: *Proc. of International Conference on Computer Vision (ICCV)* (2003)
29. Smeaton, A., Over, P., Kraaij, W.: High-level feature detection from video in TRECVid: a 5-year retrospective of achievements. *Multimedia Content Analysis, Theory and Applications* pp. 151–174 (2009)
30. Snoek, C.G.M., van de Sande, K.E.A., de Rooij, O., Huurnink, B., Gavves, E., Odijk, D., de Rijke, M., Gevers, T., Worring, M., Koelma, D.C., Smeulders, A.W.M.: The MediaMill TRECVID 2010 semantic video search engine. In: *Proc. of TRECVID Workshop*. (2010)
31. Snoek, C.G.M., Worring, M.: Concept-based video retrieval. *Foundations and Trends in Information Retrieval* **2**(4), 215–322 (2009)
32. Snoek, C.G.M., Freiburg, B., Oomen, J., Ordelman, R.: Crowdsourcing rock'n'roll multimedia retrieval. In: *Proc. of ACM International Conference on Multimedia (ACM MM)*. pp. 1535– 1538 (2010)
33. Taksa, I., Spink, A., Goldberg, R.: A task-oriented approach to search engine usability studies. *Journal of Software (JSW)* **3**(1), 63–73 (2008)
34. Ulges, A., Schulze, C., Koch, M., Breuel, T.: The challenge of tagging online video. *Computer Vision and Image Understanding* (2009)
35. U.S. Department of Health and Human Sciences: Research-based web design & usability guidelines. Available at: www.usability.gov/guidelines/ (2006)
36. van Velsen, L., König, F., Paramythis, A.: Assessing the effectiveness and usability of personalized internet search through a longitudinal evaluation. In: *Proc. of 6th Workshop on User-Centred Design and Evaluation of Adaptive Systems (UCDEAS)* (2009)
37. Yang, J., Jiang, Y.G., Hauptmann, A.G., Ngo, C.W.: Evaluating bag-of-visual-words representations in scene classification. In: *Proc. of Int'l Workshop on Multimedia Information Retrieval (MIR)* (2007)
38. Yuen, J., Russell, B., Liu, C., Torralba, A.: Labelme video: Building a video database with human annotations. In: *Proc. of Int'l Conference on Computer Vision (ICCV)*. pp. 1451 – 1458 (29 2009-oct 2 2009)
39. Zhang, H.: The optimality of Naïve Bayes. In: *Proc. of FLAIRS'2004* (2004)
40. Zhang, J., Marszałek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision (IJCV)* **73**, 213–238 (2007)