University of Modena and Reggio Emilia

D.I.I. - DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

VidiVideo

Interactive semantic video search with a large thesaurus of machine-learned audio-visual concepts

Tech Rep 5.0 - 16/12/2008

Video Surveillance User Community

Roberto Vezzani, Rita Cucchiara

Dipartimento di Ingegneria dell'Informazione

University of Modena and Reggio Emilia

via Vignolese 905 – 41100 Modena, Italia

Tel +39-059-2056111 Fax +39-059-2056129

email: {roberto.vezzani, rita.cucchiara}@unimore.it

1. Introduction

From the Description of Work document:

> *"To guarantee the concreteness of the project and in order to reach the requirements and opportunities of the market it is important to create a relationship between the research activity and software development in the project and the methodological definition suiting the user requirements through the establishment of a User Group for all the three field of application."*

Three user groups were identified:

- Scientific and cultural heritage
- Broadcast and Entertianment industry
- Video surveillance

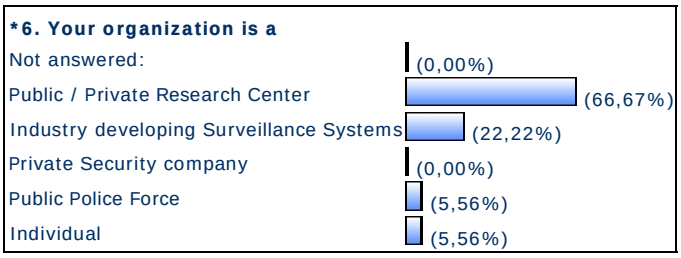
The first and the second fields have been previously faced by the FRD partner, who spread a questionnaire and collected responses from about 20 qualified users of each field. Following their suggestions and guide lines we proposed and distributed a similar questionnaire to the Video Surveillance community. In this document commented results of this activity are reported.

2. The Questionnaire

As in the others fields, a web survey has been created and articulated into 23 questions. The original questionnaire from FRD has been formulated following the guidelines of IEEE Standard 830-1984 Guide to Software Requirements Specifications and Standard P1233-1992 Guide for Developing System Requirements Specification. Appendix A and B contain the questionnaire and the invitation letter we sent to the community.

3. Video surveillance user community

The community of video surveillance users is not homogeneous and collects different types of actors, ranging from scientific researchers to companies producing surveillance systems and final users such as private citizens or police force. As in the other two fields, we collected qualified users tring to cover all these different user types.



4. The Questionnaire results

A summary of the answers is here reported. Aggregated values in percentage format are shown. If the question allows more than one answer, the reported values indicate the number of users which selected the corresponding answer; thus dor those questions the sum of the percentages can exceed 100%.

Overview of existing systems and practices

***7. Your organisation own a VideoSurveillance archive**

True (33,33%)
False (66,67%)

***8. Which kind of software platform your organisation is currently using for surveillance**

Not answered: (0,00%)
no software platform used (22,22%)
In-house developed software platform (61,11%)
Other:
- AlphaTech (1 times)
- Linux ReHat Enterprise (1 times)
- matlab (1 times)

***9. Are you currently digitising and storing the video material in your archive?**

True (66,67%)
False (33,33%)

10. Which formats are you using? (Please choose all that apply)

Not answered: (0,00%)
MPEG-1 (16,67%)
MPEG-2 (27,78%)
MPEG-4 (55,56%)
Quicktime (11,11%)
RealMedia (0,00%)
Windows Media (11,11%)
Flash (11,11%)
Other (33,33%)

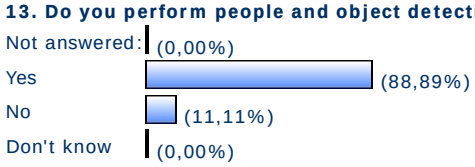
11. Which is the bitrate used to encode your digital video?

Not answered: (0,00%)
under 500 Kbps (27,78%)
500 Kb/s to 1,5 Mb/s (50,00%)
1,5 Mb/s to 2,5 Mb/s (27,78%)
Above 2,5 Mb/s (16,67%)

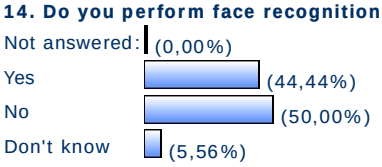
12. Do you perform event detection on your video?

Not answered: (0,00%)
Yes (83,33%)
No (16,67%)
Don't know (0,00%)

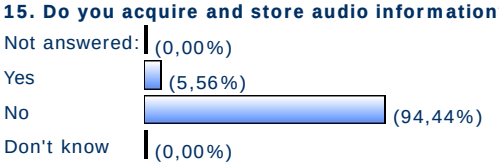
13. Do you perform people and object detection on your digital video?



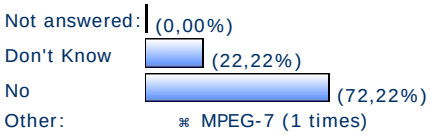
14. Do you perform face recognition on your digital video?



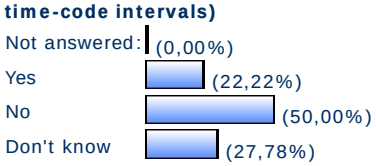
15. Do you acquire and store audio information?



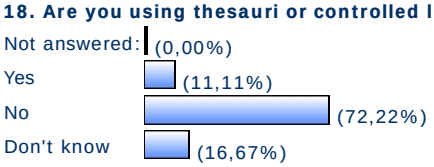
16. Are you using a standard metadata schema for annotating material? (Dublin Core, Dublin Core for video, MPEG-7, etc.)



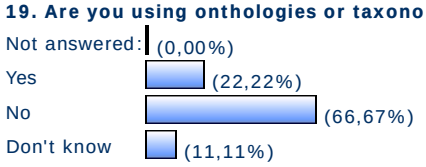
17. Are you tagging your videos using time-aligned metadata? (metadata synchronised with time-code intervals)



18. Are you using thesauri or controlled lexicons to annotate videos?

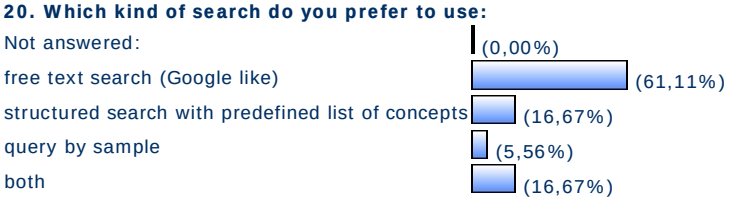


19. Are you using onthologies or taxonomies to annotate videos?

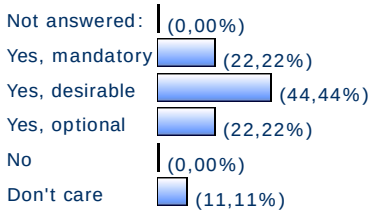


Search options and preferences

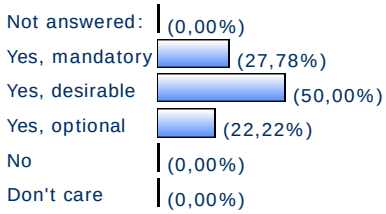
20. Which kind of search do you prefer to use:



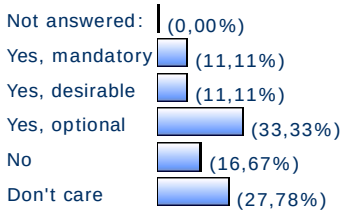
21. Would you like to have the possibility to search using another video as input? (Query by Visual Example)



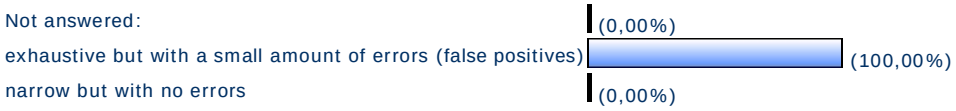
22. Would you like to have the possibility to search using concepts and their relations? (e.g. search for "scenes containing walking people", or "scenes containing traffic", or "scenes containing Ferrari GTO" that can automatically be expanded to -> "containing sport cars" and then to -> "containing vehicles")



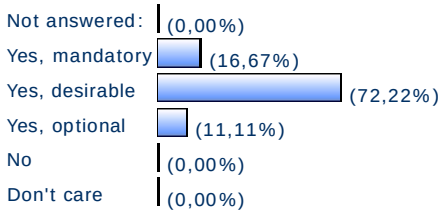
23. Would you like to have the possibility to search for sounds or audio events? (e.g. search for "ringing bells", or "traffic noise")



24. Would you prefer to have a query result

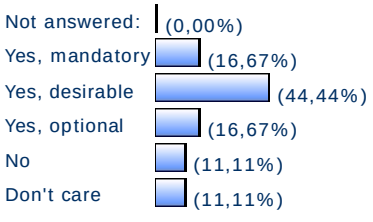


25. Would you like to have your query results presented at a frame level? (showing directly the frames matching your query)

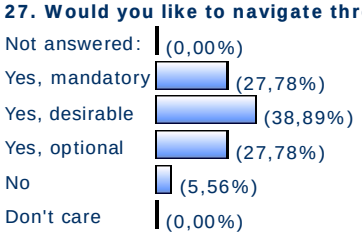


User interface

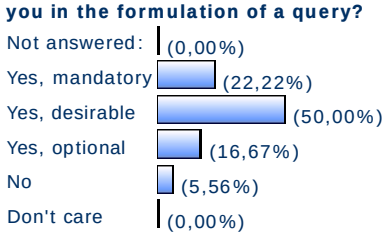
26. Would you like to use a web based search interface



27. Would you like to navigate through the retrieved videos at a frame level?



28. Would you like to have a visual interface representing concepts and their relations to assist you in the formulation of a query?



5. Results analysis and comments

Current status

- Most of the users which hold a digital archive have developed in-house their own database application (61%)
- Most of them hold a digitised video collection but there is still a wide use of analog supports (33%)
- Mpeg is the widest used video format, but about 33% of the users adopt also some custom formats
- The majority of them perform event and object detection on their videos
- The majority of them do not make use of the audio
- Only few users have a taxonomy, an ontology or a controlled set of terms to annotated videos

Most requested features:

- A web based search interface is a must
- Both free text (Google like) and structured search are requested
- Query results presented at a frame level (showing directly the frames matching the query) are highly requested
- Search for AV documents using concepts and their relations is requested
- A visual interface representing concepts and their relations to assist in the formulation of a query is highly demanded

6. Conclusions

This activity has highlighted the emerging need of new technology for videosurveillance archives. In particular, users pointed out the lack of shared ontologies and lexicon terms for video annotation. The VidiVideo project can improve the current technology level adopted and meet user requirements, allowing new query and search interfaces for videosurveillance archives.

Appendix A – the questionnaire

VIDI-Video - Users requirements questionnaire

1. Please, fill in the following contact info. We will keep you updated about the development of the VIDI-Video project.

Name:
Company:
City/Town:
Country:
Email Address:

Overview of existing systems and practices

2. Your organisation own a VideoSurveillance archive?

[ ] Yes *(if yes, the survey continues to question 3.)*
[ ] No *(if no, the survey continues to question 12.)*

3. Which kind of software platform your organisation is currently using for surveillance?

[ ] Commercial software platform. Please specify: _____
[ ] In-house developed software platform.
[ ] no software platform used.

4. Are you currently digitising and storing the video material in your archive?

[ ] Yes
[ ] No

If yes,

4a. Which formats are you using? (Please choose all that apply)

MPEG-1
MPEG-2
MPEG-4
Quicktime
RealMedia
Windows Media
Flash
Other

4b. Which is the bitrate used to encode your digital video?

under 500 Kbps
500 Kb/s to 1,5 Mb/s
1,5 Mb/s to 2,5 Mb/s
Above 2,5 Mb/s

5. Do you perform event detection on your video?

[ ] Yes
[ ] No
[ ] Don't know

6. Do you perform people and object on your digital video?

[ ] Yes
[ ] No
[ ] Don't know

7. Do you acquire and store audio information?

[ ] Yes
[ ] No
[ ] Don't know

8. Are you using a standard metadata schema for annotating material? (Dublin Core, Dublin Core for video, MPEG-7, etc.)

[ ] Yes Please, specify: _____
[ ] No
[ ] Don't know

9. Are you tagging your videos using time-aligned metadata? (metadata synchronised with time-code intervals)

[ ] Yes
[ ] No
[ ] Don't know

10. Are you using thesauri or controlled lexicons to annotate videos?

[ ] Yes
[ ] No
[ ] Don't know

11. Are you using onthologies or taxonomies to annotate videos?

[ ] Yes
[ ] No
[ ] Don't know

Search options and preferences

12. Which kind of search do you prefer to use:

[ ] free text search (Google like)
[ ] structured search
[ ] query by sample
[ ] both

13. Would you like to have the possibility to search using another video as input? (Query by Visual Example)

[ ] Yes, mandatory [ ] Yes, desirable [ ] Yes, optional [ ] No [ ] Don't care

14. Would you like to have the possibility to search using concepts and their relations? (e.g. search for "scenes containing walking people", or "scenes containing traffic", or "scenes containing Ferrari GTO" that can automatically be expanded to -> "containing sport cars" and then to -> "containing vehicles")

[ ] Yes, mandatory [ ] Yes, desirable [ ] Yes, optional [ ] No [ ] Don't care

15. Would you like to have the possibility to search for sounds or audio events? (e.g. search for "ringing bells", or "traffic noise")

[ ] Yes, mandatory [ ] Yes, desirable [ ] Yes, optional [ ] No [ ] Don't care

16. Would you prefer to have a query result

[ ] exhaustive but with a small amount of errors (false positives)

[ ] narrow but with no errors

17. Would you like to have your query results presented at a frame level? (showing directly the frames matching your query)

[ ] Yes, mandatory [ ] Yes, desirable [ ] Yes, optional [ ] No [ ] Don't care

User interface

18. Would you like to use a web based search interface?

[ ] Yes, mandatory [ ] Yes, desirable [ ] Yes, optional [ ] No [ ] Don't care

19. Would you like to see in the search results the keyframes of the selected videos?

[ ] Yes, mandatory [ ] Yes, desirable [ ] Yes, optional [ ] No [ ] Don't care

20. Would you like to navigate through the retrieved videos at a frame level?

[ ] Yes, mandatory [ ] Yes, desirable [ ] Yes, optional [ ] No [ ] Don't care

21. Would you like to have a visual interface representing concepts and their relations to assist you in the formulation of a query?

[ ] Yes, mandatory [ ] Yes, desirable [ ] Yes, optional [ ] No [ ] Don't care

Appendix B – The invitation letter

Dear user,
thank you very much for your interest in the ViSOR project.

We are very glad to have you in our community and we hope ViSOR can be a valid support to your research activities.

We want to remember you that ViSOR is supported by the VidiVideo project. VIDI-Video project aims at enhance access to video, by developing a semantic search engine.
The project will boost the performance of video search by developing a 1000 element thesaurus for automatically detecting instances of semantic concepts in the audio-visual content.
The main novelties of VIDI-Video are the size and quality of the thesaurus. The VIDI-Video project is a FP6 project funded by the EU. (http://www.vidivideo.info)
In order to improve our work and focus the research to real user requirements, we are asking you to fill in a short questionnaire about surveillance systems. The questionnaire is available at the following address: http://imagelab.ing.unimore.it/survey

Please carefully answer to the reported questions and feel free to distribute and forward the link to your collaborators and industrial partners.

Thanks in advance for your collaboration and your time.

The ViSOR Staff
Roberto Vezzani