ABSTRACT
What are the evolving trends in video search? What has changed technologically after many decades of solid research and development in computer vision, image processing, information retrieval and pattern recognition to enable products that search millions of video from the Web? We attempt to highlight some of the changes that are altering the fundamental approach to video information retrieval.

One of the fundamental problems of applying information retrieval techniques to media domains is need of “training/contextual” data. With video search, it is still very challenging to get vast amounts of content feature transcribed data. The power of community description and annotation frameworks is altering the nature of that problem. For instance, the web already contains millions of media files that have descriptive information about the media generated by the web page developer. Another related powerful aspect is user tagging. In this scenario, community effects that encourage descriptive information sharing/tagging of media provide a vast amount of useful information. A third dimension is leveraging context to build improved media search systems. Context encompasses a wide variety of aspects ranging from user context, geo-location, and time.

There are enormous opportunities with this shifting paradigm of leveraging non-content based meta-data to bootstrap media information retrieval systems. This is not a new idea: we have seen such paradigm shifts in the area of speech recognition where the model leveraged context derived not from the audio data, but from textual corpus. In the speech recognition example, using non-media derived context fuelled the success of speech technology enormously. The approach of utilizing millions of text from news sources such as Wall Street Journal and building statistical text based language models which then provide the context to improve the acoustic matching of the audio enabled building large vocabulary systems for recognizing continuous speech with reasonable accuracies. The same is true for media search, where the non-media content can be used to restrict contexts for application of media based technologies. A lot more work still needs to be done in investigating how best to combine the different content based classifiers with the meta-data based classifiers for video search. Along with this approach comes the challenge of dealing with meta-data inaccuracies and annotation spam. Just having more data does not imply a better model. The emergent challenges around this approach are also discussed.

This is a defining time for video search with the advent of web based video search products that are generating high consumer interest/adoptions. The research community can help enormously by designing algorithms and techniques that can combine media based models with the non-media derived meta-data effectively. Other key opportunities influenced by the changing landscape of cheaper cameras, mobile, coupled with more media aware devices make this an exciting time for researchers to be evaluating next-generation strategies for video search.