Final report

Personalized Content-Based Categorization of Web Videos

- **Researchers names and details**

Lihi Zelnik-Manor Dept. of Electrical Engineering Technion - Israel Institute of Technology Haifa, 32000, Israel e-mail: lihi@ee.technion.ac.il Phone: +972-4-8295736 Fax: +972-4-8295757	Shie Mannor Dept. of Electrical Engineering Technion - Israel Institute of Technology Haifa, 32000, Israel e-mail: shie@ee.technion.ac.il Phone: +972-4-8293284 Fax: +972-4-8295757

Summary

With approximately 10^5 video clips posted on-line daily, video content is the fastest growing media on the web. Browsing web video collections via the textual tags attached to the videos by their owners as in YouTube or Google Video is far from an ideal experience. The tags are typically incomplete or not provided at all. Moreover, the browsing experience is often frustrating as different users have different concepts of the meaning of the same tags. We hence have been developing methods for content-based video search and personalized categorization and classification.
We have proposed to address the challenge of browsing web video collections by advancing in three fronts: (1) design methods for efficiently capturing and representing the contextual information within web video clips; (2) devise methods for adaptive and personalized data categorization, which can handle missing data and individual scores (two individuals could disagree on the "correct" categorization); (3) develop web-tools for collecting user input, e.g., scoring of success of search results or rating the similarity of pairs of clips. The collected user input will be integrated into the categorization system as well as used for evaluating results. We intend to make such tools as well as the data collected available on-line to the public.

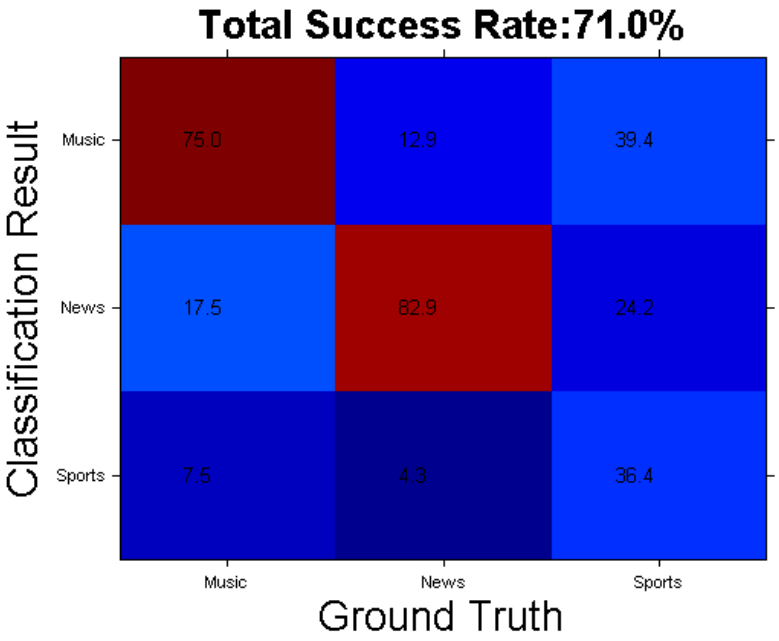
Progress report for the funded period:

- **Infrastructure:**
  - We developed a software tool which enables harvesting multitude of video clips automatically and easily by providing it with a list of keywords. The tool automatically saves the data-base sorted in XML format.
  - A 36 node high-end computing cluster has been purchased and is about to be operable. The server will enable data collection as well as fast processing of vast amounts of data.

- **Research**
  - We have studied previous papers related to video content representation and learned their drawbacks and limitations.
  - We have identified several related works and implemented them:

    1. S. Zanetti , L. Zelnik-Manor and P. Perona. A Walk through the Web's Video clips.
    2. L. Zelnik-Manor and M. Irani. Statistical Analysis of Dynamic Actions.
    3. P. Dollar, V. Rabaud, G. Cottrell and S. Belongie. Behavior Recognition via Sparse spatio-temporal Features.
    4. I. Laptev and T. Lindeberg. On space-time Interest Points.
    5. D.G. Lowe. Object Recognition from Local Scale Invariant Features.
    6. L. Gorelick, M. Blank, E. Shechtman, M. Irani, R. Basri. Actions as Space-Time shapes.
    7. X. Hou and L. Zhang. Saliency Detection: A spectral residual approach.
    8. C. Guo, Q. Ma and L. Zhang. Spatio-Temporal Saliency Detection using Phase Spectrum of Quaternion Fourier Transform.
    9. A. Oliva and A. Torralba. Building the gist of a scene: the role of global image features in recognition.
    10. S. M., Tahaghoghi, H. E., Williams, J. A., Thom, and T. Volkmer, "Video cut detection using frame windows".

  - Initial progress has been made on video content representation. The method we have developed is based on quantization of the frequency-response in the Fourier domain. A video is first filtered to detect only the salient parts. Then a Fourier transform is applied to the salient portions. We quantize the frequency domain and construct descriptors based on the space-time regions related to each bin of frequencies. These descriptors are concatenated into a single descriptor per video. The final descriptor is used for classifying the video content.
    In a preliminary experiment we have classified 300 videos of music, news and sports. The confusion matrix below shows our classification rates with an average of 71% correct classification.
  - We identified dimensionality reduction as a core problem and studied the problem of using principal component analysis when the dimensionality of the data is very high (equal or larger than the number of data samples). This resulted in a new algorithm dubbed High dimensional Robust Principal Component Analysis (HRPCA). A preliminary analysis of the algorithm appeared in the 23rd Annual Conference on Learning Theory (COLT). We believe that dimensionality reduction in the high dimensional regime is the key for successful encoding of data.



o

The project was initially planned to last 3 years. Since the budget was provided was less than one year parts of the project have not been completed. We have worked towards building the theory, algorithms and infrastructure that are needed to complete the project, but have not been able to establish as much as we originally hoped for.