A Cityscape Visualization of Video Perspectives

Mark Derthick, Mike Christel, Alex Hauptmann, Dorbin Ng, Scott Stevens, Howard Wactlar

Carnegie Mellon University

CMU's Informedia project has collected and automatically processed a multi-terabyte video corpus containing 8 years of CNN broadcasts and other video sources [5]. Previous work has demonstrated multi-modal querying by text, image, time, and location, and the ability to summarize a single document or a set of documents matching a query. We now plan to organize the corpus or a subset along multiple dimensions, or *perspectives*, adding relevant background material, significantly expanding and accelerating the viewer's comprehension and integration of knowledge. A perspective can provide factual background information, a history of an issue, the view of a biased source, a technical or medical perspective, or any of dozens of others. This abstract proposes a cityscape metaphor for organizing visual context in terms of perspectives.

Problems with current document visualizations Video libraries have traditionally provided visual overviews by showing a sequence of thumbnail images of individual shots in a document. Search engines like Google provide lists of similar documents. Both of these approaches are at too low a level. Others attempt to cluster results hierarchically as a way to provide an overview [2]. Key phrases are extracted as cluster labels. This has two drawbacks. First, it requires cognitive effort to read and interpret the phrases, even in the best case where the extracted phrases adequately describe the clusters. Second, the hierarchy may not reflect the user's interests. For instance, if the user has been reading about oil exploration in Alaska, a top level split labeled Texaco, Exxon, Mobil might be the way he/she mentally organizes the information space, but there are many other perspectives for which this organization does not help.

Themeview [6] is perhaps the best known text document corpus visualization. Like most text information retrieval systems, it treats a document as a vector of normalized word counts, with one element for each word in the corpus vocabulary. This is usually at least several thousand. In order to visualize this high-dimensional space, it must be projected down onto 2D or 3D. This also forces a single organization on the data. Themeview axes are not meaningful because they are an information-loss minimizing projection with thousands of parameters. In contrast, perspectives visualize documents in a few dozen *meaningful* dimensions.

Cityscapes Virtual reality landscapes have often been suggested as ways to organize cyberspace, because people have evolved for orientation and navigation of such spaces. Realistic cityscapes are sufficiently varied and detailed to convey many conceptual dimensions and, importantly, to be memorable.

Some landscape visualizations, such as Data Mountain [3], do not impose any semantics on layout. Rather, the user is free to conceive of neighborhoods and to arrange web page snapshots accordingly. City of News [4] uses a cityscape visualization for visited web pages, with neighborhoods corresponding to districts found in real cities, such as financial or shopping, and sections found in newspapers, such as sports and books. In contrast, in our cityscape example below, buildings correspond to abstract perspectives whose content is dynamically synthesized, rather than to individual web pages.

In previous cityscape work, the objects in the cityscapes have not been treated spatially themselves. Our approach is to combine the idea of parallel coordinates [1] with cityscapes. We can think of a building as one vertical parallel axis. Each building represents a perspective. Thus the system can automatically assign contextual items to a vertical position on the appropriate building. The user is still free to rearrange the buildings into neighborhoods, for instance so that those of greatest interest are larger and/or in front. Concept-space navigation and zooming are accomplished by spatial navigation, possibly by "flying" or possibly by clicking on a building of interest and using "autopilot" to navigate for a good view of it.
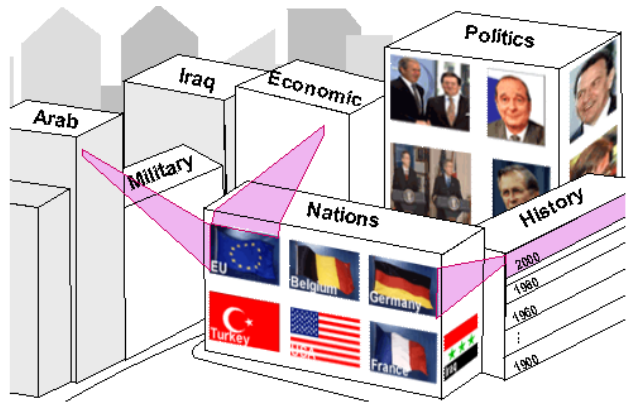
Schematic and Realistic Cityscapes communicating perspectives on a focus video.

A cityscape view is also useful for dynamic summaries. For instance, using a slider to control the range of broadcast dates, the ebb and flow of perspectives and named entities in news casts over time would show up as different parts of the city sparkling and darkening. A night time view will be most effective for this. Corresponding to the line segments of parallel axes, we will use "spotlights" that illuminate relationships among documents and entities. While in parallel coordinates there are d(n-1) segments for d data and n dimensions, we expect to use far fewer, linking only the most important item pairs. For instance, a query or focus document would light up its context in white, with the strongest semantic associations being the brightest. There could be blue spotlights from these to the named entities they mention. In addition, items lit during a previous related query may only gradually fade to yellow and then brown, which provides an overall sense of "direction" to an exploration sequence. A dull red glow could indicate that one of the user's "buddies" has been there recently.

References

1. Inselberg, A. and B. Dimsdale. Parallel Coordinates: A Tool for Visualizing Multi- Dimensional Geometry, in *IEEE Conference on Visualization*. 1990. p. 361.

2. Palmer, C.R., H. Wactlar, R.E. Valdes-Perez, J. Pesenti, D. Ng, A. Hauptmann, and M. Christel. Demonstration of Hierarchical Document Clustering of Digital Library Retrieval Results, in *Joint Conference on Digital Libraries (JDCL '01)*. 2001. Roanoke VA.

3. Robertson, G., M. Czerwinski, K. Larson, D.C. Robbins, D. Thiel, and M.v. Dantzich. Data Mountain: Using Spatial Memory for Document Management, in *User Interface Software and Technology (UIST)*. 1998. San Francisco, CA: ACM Press. p. 153-162.

4. Sparacino, F., R. DeVaul, C. Wren, G. MacNeil, G. Daveport, and A. Pentland. City of News, in *SIGGRAPH 99, Visual Proceedings, Emerging Technologies*. 1999. Los Angeles CA. http://www.siggraph.org/s99/conference/etech/projects.html

5. Wactlar, H., T. Kanade, M. Smith, and S. Stevens, Intelligent Access to Digital Video: Informedia Project. *IEEE Computer*, 1996. **29**(5): p. 46-52.

6. Wise, J.A., J. Thomas, J., K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents, in *IEEE Information Visualization*. 1995. p. 51-58.