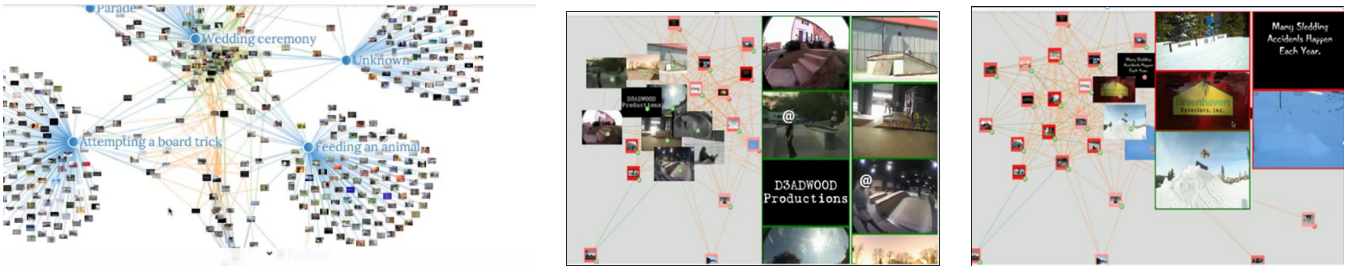
Large-Scale Human Geography from Crowdsourced Multimedia

(EMERGING TECHNOLOGIES)

Matt Turek (Asst. Dir. of Computer Vision), Sangmin Oh, Anthony Hoogs, Amitha Perera
Kitware, Inc.
matt.turek@kitware.com, 518-881-4942

New multimedia content is being shared through the Internet (e.g., YouTube, Facebook) at an unprecedented pace. On YouTube alone, video data is being uploaded at the rate of 30 million hours a year. Such massive crowdsourced multimedia collections provide a unique opportunity to learn and explore human geography—the *terrain* of human events. Crowdsourced multimedia is particularly advantageous because it provides an aggregate participant's view of events, living conditions, and economic vitality without collection cost or risk.

While crowdsourced multimedia provides unique opportunities, its vast scale requires automatic analysis and visual exploration tools to help organize and identify its content. The ability to quickly and interactively explore a large multimedia collection is a key capability, particularly for analysts faced with quickly evolving events such as natural disasters and geo-political events.



(a) (b) (c)

Figure 1: (a) **Automatic video grouping using semantic concepts**. (b, c) **Interactive multimedia organization**. Nearby videos have similar contents. The similarity measure is configurable as a combination of physical distance, appearance, audio, etc. Two snapshots from different graph neighborhoods show (b) dynamic activities in an urban area and (c) activities in snowy, mountainous areas.

Here we present two solutions enabling more efficient analysis of such large-scale, crowd-sourced multimedia collections through the use of advanced multimedia visualization. Videos are first processed to automatically extract descriptors to characterize video and audio content. A large array of semantic concepts, such as scene type (e.g. urban, seaport, factory); visual objects (e.g. person, car, building); and audio events (e.g. explosion, shout, speech) are computed. Lower-level descriptors, such as visual shape, motion, and audio patterns are also computed to model content outside previously known semantic categories.

Our two interactive visualization approaches can then be used to organize and rapidly explore large multimedia collections efficiently. Our first approach is automatic video organization based on semantic concepts. A snapshot from our system is shown in Fig. 1(a). In this mode, videos are grouped by automatically detected semantic concepts. The users can select the semantic concepts of interest and are able to rapidly filter out videos of less interest. This mode provides a straightforward way to search videos from multiple semantic angles. However, it is infeasible to know a-priori all semantic concepts that may be of interest in the future. Consequently, we also allow users to explore the multimedia collection with user-defined similarity measures.

In our second visualization approach, we present the collection of videos embedded on a graph, as shown in Fig. 1(b,c). Each node corresponds to an image or video, and an edge between two nodes corresponds to the similarity between them. The figure illustrates how two different neighborhoods will show different types of content. One of the major innovations in this work allows users to interactively define the similarity as a combination of multiple dimensions such as geographical proximity, similarity of visual shape or audio patterns, and the distribution of detected semantic concepts. For example, if users are looking for more audio-driven content, the weight of the audio information can be increased, and the overall organization of videos will update in real-time based on the new similarity measure.