A Videography Analysis Framework for Video Retrieval and Summarization

Kang Li*[1]
kangli@buffalo.edu
Sangmin Oh*[2]
sangmin.oh@kitware.com
A. G. Amitha Perera[2]
amitha.perera@kitware.com
Yun Fu[3]
raymondyunfu@gmail.com

[1] Department of CSE
State University of New York
Buffalo, NY, USA

[2] Kitware, Inc.
Clifton Park, NY, USA

[3] Department of ECE and College of CIS
Northeastern University
Boston, MA, USA

Overview: In this work, we focus on developing features and approaches to represent and analyze videography styles in unconstrained videos. By unconstrained videos, we mean typical consumer videos with significant content complexity and diverse editing artifacts, mostly with long duration. We present an approach for *unsupervised videography analysis* for unconstrained videos. Intuitively, each videography can be understood as a camera director's direction on a movie script, *e.g.*, "capture the running actress by panning the camera, to have her face appear at 20 percent size of the video". The idea is that different classes of video content will have different styles—the videography style of a wedding video should be different from a sports video—and so, the videography style should provide a valuable signal for automated content analysis.

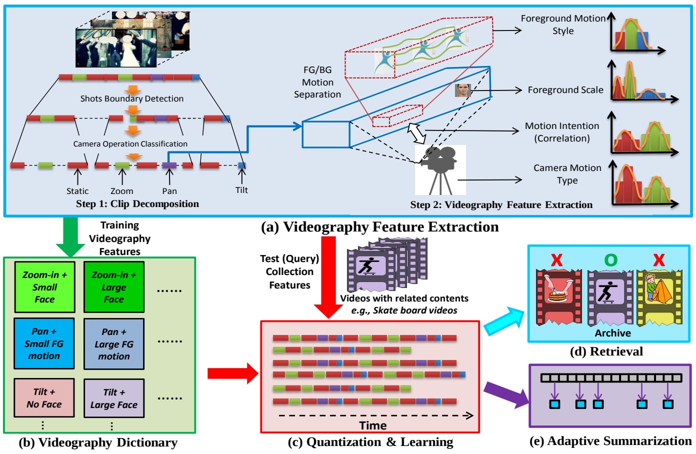
Shots Boundary Detection
Camera Operation Classification
Static
Zoom
Pan
Tilt
Step 1: Clip Decomposition
FG/BG Motion Separation
Foreground Motion Style
Foreground Scale
Motion Intention (Correlation)
Camera Motion Type
Step 2: Videography Feature Extraction
(a) Videography Feature Extraction
Training Videography Features
Zoom-in + Small Face
Zoom-in + Large Face
Pan + Small FG motion
Pan + Large FG motion
Tilt + No Face
Tilt + Large Face
......
(b) Videography Dictionary
Test (Query) Collection Features
Videos with related contents e.g., Skate board videos
Time
(c) Quantization & Learning
X
O
X
Archive
(d) Retrieval
(e) Adaptive Summarization

Figure 1: Framework for videography analysis and applications.

Videography Analysis: The overall framework of our approach is illustrated in Fig. 1(a). First, a two-level motion analysis is conducted to decompose long clips into sequences of segments with coherent motion types (S/P/T/Z). Second, multiple features related to motion and scale patterns are measured from every segment, which are used to characterize videography. Throughout this work, we utilize densely computed KLT tracks over the entire clips as main basis for the derived features.

We assume that there are diverse videography styles in unconstrained videos, which are discovered as a *videography dictionary* via unsupervised clustering on proposed features. Then, a video clip can be represented as a series of segments with varying videography words. For the underlying videography features, we extend conventional features such as camera motion and foreground (FG) object motion (e.g., [1]) by incorporating two novel features: *motion correlation* and *scale* information.

Once videography features are obtained from segments, they are used to build *videography dictionary* (VD) shown in Fig. 1(b). The computed VD will be used to quantize video clips into sequences of videography words (VWs), as shown in Fig. 1(c). Our analysis shows that there are regularized patterns in the videography used in the unconstrained Internet videos, and correlations between the exhibited videography styles and video contents. Such observation on discriminative correlations suggests that videography analysis can actually be used for challenging tasks such as content-based retrieval and content summarization.

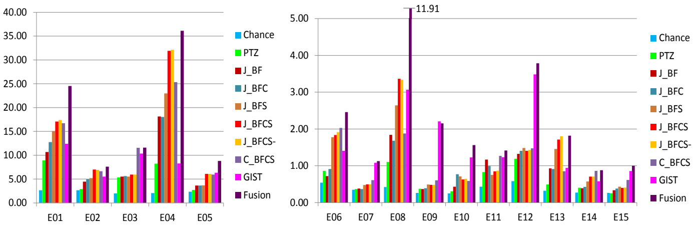

Figure 2: Average Precision (%) of video retrieval results on MED corpus, for 15 events: ***(E01) Board trick***, *(E02) Feeding animal, (E03) Fishing,* ***(E04) Wedding***, *(E05) Working wood project,* ***(E06) Birthday party***, *(E07) Change vehicle tire,* ***(E08) Flash mob***, *(E09) Getting vehicle unstuck, (E10) Groom animal, (E11) Make sandwich,* ***(E12) Parade***, ***(E13) Parkour***, *(E14), Repair appliance, and (E15) Sewing project.*

Video Retrieval: For retrieval, we computed bag-of-word representations based on the videography word sequences and employed them as the basis for content-based video retrieval tasks. We have conducted experiments on a large TRECVID '11 MED dataset where we tried diverse variations of the proposed approach as well as using more conventional features such as GIST. Our results indicate that the proposed videography features effectively improve the retrieval performance and are complimentary to traditional appearance features such as GIST, improving performance further when both features are used jointly. Figure 2 shows the list of video event classes and the extent of conducted retrieval experiments as well as summarized performance profiles. Event classes that show the most benefits by videography-based analysis are marked in bold.

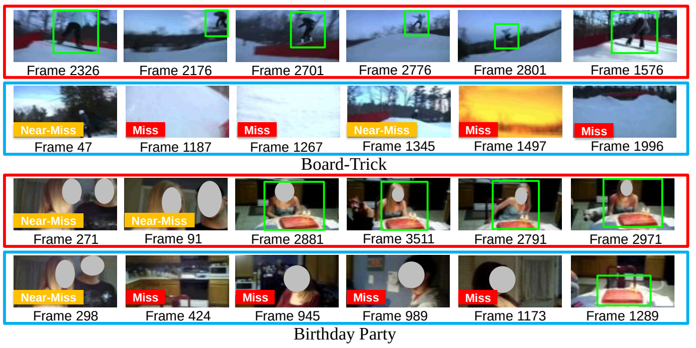

Figure 3: Videography-aware adaptive summarization. Three summarization results by this work (red rows) and baseline (blue rows). Detected FG regions (green) and human judgements on relevance of key frames (good:none, near-miss: yellow, miss: red) are marked on each image.

Video Summarization: We also show that the proposed videography analysis can be used to provide videography-aware adaptive summarization method. For example, Fig. 3 shows example summarization results for different events where the segments with distinctive videography styles for particular events are highlighted in the summaries, e.g., board tricks during snowboarding and candle blowing during a birthday party. Summarization produced by our proposed approach is shown in red and results by baseline approaches of using color histogram changes are shown in blue.

[1] Xingquan Zhu, Ahmed K. Elmagarmid, Xiangyang Xue, Lide Wu, and Ann Christine Catlin. InsightVideo: Towards hierarchical video content organization for efficient browsing, summarization and retrieval. *IEEE Transactions on Multimedia*, 7(4):648–666, 2005.

This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20069. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI/NBC, or the U.S. Government.

* Indicates equal contributions.