Previewing Video Content with Dynamic and Interactable Collage

Sheng-Jie Luo[1*] Chun-Yu Tsai[2] Hsiao-Ching You[1*] Hsun-Pei Wang[1*] Wei-Chao Chen[3] Bing-Yu Chen[1†]

[1]National Taiwan University [2]Columbia University [3]Skywatch Innovation Inc.

Figure 1: *A series of frames to illustrate the interaction process of our system. The manipulated video is highlighted with red boundaries. (The video clips shown in the figure, including Shark Tale, Shrek, How to Train Your Dragon©Dreamworks Animation, and Toy Story, Ratatouille©Walt Disney Pictures, are used for research purpose.)*

1 Introduction

With the growing use of digital cameras and video recorders, video creation and distribution are becoming easier. While more and more videos are easily accessible, the burden of browsing the video contents also grows exponentially. It is challenge to effectively browse a collection of videos at once and pick those users want. Even current video summarization methods either combine salient shots into a short video, or pack a collection of keyframe in a static image. In this extended abstract, we present a dynamic collage system that aims to enable simultaneous browsing of a collection of videos. The important portions of videos are extracted and packed one by one in the same canvas, and are played simultaneously. An efficient layout algorithm is proposed to enable the system to respond to content changes and users' interactions efficiently, and the method encourages the canvas to be compact and informative. Therefore, the system provides an interface to efficiently browse and rediscover those interesting contents from a collection of video files.

2 Dynamic and Interactable Collage

Our goal is to facilitate users to efficiently browse video content. The final collage should satisfy the following criteria: (1) The important regions should appear in the final canvas. (2) The non-essential parts of a video can be eliminated in the canvas. (3) The collage should respond the change of video contents and users' interaction. Our method consists of three steps: *importance portion extraction*, *collage layout*, and *canvas rendering*.

Important portion extraction. This step aims to extract the important regions of videos. We first divide each video into a number of shots using cut detection. Then we compute the temporal-spatial saliency for each shot, which is the combination of the saliencies of all frames. Frame saliency is computed using the features of motion contrast, color contrast, and face information. Finally, top 5% salient pixels are used to form the important volume for computing layout.

*e-mail:{forestking, hcyou, garfieldduck}@cmlab.csie.ntu.edu.tw
†e-mail:robin@ntu.edu.tw

Collage layout. In this step, our objective is to find an optimal configuration without occluding each important volume. Unfortunately, it is an NP-complete problem, and we approximate the optimal solution by a two-stage heuristic. First, we use a greedy algorithm to initialize a layout frame by frame, which is described as follows. The first important region is placed at the center of the canvas. Then, we iteratively place each remaining important region R_i radially around the canvas center while ensuring no overlap between R_i and all other already on the canvas. An optimal direction is selected to minimize the empty space while respecting the aspect ratio of the canvas. Second, this layout is iteratively refined. We formulate the energy function based on penalty measures on empty space, important region occlusion, and aspect ratio deviation: $E = E_{es}^{\alpha} + kE_{occ}$, where E_{es} and E_{occ} refers to *empty space energy* and *occlusion energy*, respectively. They are defined as

$$E_{es} = \frac{Area(R_B - \cup_i R_i)}{Area(R_B)}, E_{occ} = \sum_i Area(R_i) - Area(\cup_i R_i),$$

where R_B is the bounding box formed by all important regions R_i. α refers to aspect ratio deviation penalty, which is defined as $\alpha = 1/(q_c - q_p)^2 + \varepsilon$.

Canvas rendering. This step generates the final collage on the canvas with the multiple-site Voronoi method. Specifically, we sample a number of Voronoi sites along the boundaries of the important regions and proceed the Voronoi algorithm. Then the segmented regions are textured with the their corresponding video portions. We also render the boundary lines for distinguishing different videos.

3 Conclusion and Future Work

The number of videos has grown rapidly in recent years, and more and more are accessible from the Internet. Simultaneous browsing of a large amount of videos has become a valuable issue for academe and industry. When using our dynamic and interactable system to browse a collection of videos, users can efficiently get what they are finding with direct manipulation of the canvas. In the future, we would improve the usability of the system and endow users with more choices of managing their ever-growing media collections.