I-010

Video2Manga: A Comic-like Video Summary

傅 晨† 山崎 俊彦† 相澤 清晴†‡
Chen Fu Toshihiko Yamasaki Kiyoharu Aizawa

1. Abstract

Video summarization and visualization plays an important role in recent years because we need efficient techniques for faster retrieval and easier understanding of video content. In this paper, we propose a “Video2Manga” system that can summarize and represent a video sequence as “Manga” style. Video key-frames first are extracted from video frames by using clustering-based method. In order to understand the video content easier and faster, we then perform Non-Photorealistic Rendering (NPR) for each key-frame to create an actual “Manga” style. The result shows that our system not only can keep the most significant information of video but also can make it easy to understand.

2． Introduction

Video summarization and visualization have been hot research topics in the fields of image processing, computer vision, and computer graphics. Researchers proposed various kinds of solutions such as Pictorial Summary [1], Video Manga [2], Video booklet [3], Video Snapshot [4], Video Collage [5], etc.

“Manga” is a graphic art form that can represent a story efficiently. Therefore, the intuition of our system is using “Manga” style to represent the video content. In this paper, we propose a “Video2Manga” system that extracts the key-frames from video frame sequence automatically and represents it as “Manga” style based on non-photorealistic rendering. It not only can keep the most significant contents of video but also can represent it easy to understand, even more artistic.

“Video2Manga” system mainly includes two steps: video key-frame extraction and Non-Photorealistic Rendering (NPR). Video is a sequence of frames, and key-frame extraction is a process to select the most representative frames from video frames. We first divide video sequence into several shots by similarity measure, and then separate frames of each shot into clusters by unsupervised clustering. Finally, we extract the centroid from clusters as key-frames. On the other hand, NPR is a field of computer graphics that can create a variety of artistic styles such as drawing, painting, and so on. The main objectives of NPR are directing our attention on the most meaningful areas in an image, and help us easier and faster to understand. Therefore, we involve NPR techniques in our system to design a more efficient video summary. The NPR framework we proposed mainly combines mean-shift segmentation, canny edge detection, and bilateral filtering, for edge enhancement and texture simplification.

† 東京大学大学院 情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

‡ 東京大学大学院 情報学環
Interfaculty Initiative in Information Studies, The University of Tokyo

3. Related Works

In recent years, there have been many studies on video summarization and visualization. For example, Yeung et al. [1] proposed a pictorial video summary by arranging video poster in timeline to tell the story. Each poster is an image summarizing the dramatic incident taking place in the story unit. Uchihashi et al. [2] proposed another form of video summary called “Video Manga”, which likes the creation of comic books. Hua et al. [3] developed a video booklet system, which chooses a set of thumbnails from original video, and prints them out on a predefined set of templates in a variety of forms. Ma et al. [4] described a video snapshot system, which integrated a set of content analysis techniques to provide compact static video summarization. Mei et al. [5] presented an automatic procedure for constructing a compact synthesized collage from a video sequence, which called “Video Collage”. They extract salient regions of interest (ROI) from each key-frame, and arranging ROIs on a given canvas seamlessly.

4. System Overview

The “Video2Manga” system includes two main components: video key-frame extraction and image non-photorealistic rendering. We first extract the key-frames from original video automatically based on clustering algorithm, then perform non-photorealistic rendering on each key-frame to create a real “Manga” style.

4.1 Key-frame Extraction

Key-frames are still images extracted from original video which best represent the video content. In the “Video2Manga” system, key-frame extraction can be viewed as the first important step of how to map the video content into a small number of representative images.

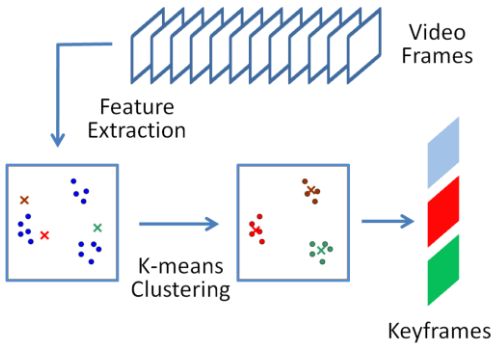

Figure 1. Overview of Key-frame Extraction

Our approach is based on unsupervised clustering method [6]. At first, we use three-dimensional color histogram in CIEL*a*b* color space and divide video sequence into several shots by computing similarity between frames. Each shot then is divided into clusters by using K-means clustering. In terms of clustering algorithm, each frame can be seen a feature point in feature space, thus we can choose the centroid of clusters and combine it as key-frames. Figure 1 shows the overview of our method based on clustering algorithm.

4.2 Non-Photorealistic Rendering

Non-photorealistic Rendering (NPR) is a field of computer graphics, which creates a variety of artistic styles such as drawing, painting, and so on. NPR directs our attention to the most meaningful areas in images, and makes it easier and faster to understand [7]. In our "Video2Manga" system, NPR is used to create the "Manga" style for each key-frame being extracted, by simplifying low contrast areas while enhancing high contrast areas.

Two main objectives of non-photorealistic rendering are summarized as edge enhancement and texture simplification. We perform canny edge detector to find the boundaries of objects and drawing the contour of detected objects to enhance these high contrast regions, while using mean-shift algorithm to simplify these low contrast regions to enhance edge detection result. Meanwhile, we repeatedly use bilateral filter, an edge-preserving smooth filter, on the luminance channel of images in CIEL*a*b* color space, which can smooth the textures while preserving the edge information. Image binarization also is used to enhance some information such as shadow, dark regions, etc. Figure 2 shows the processing flowchart we proposed.

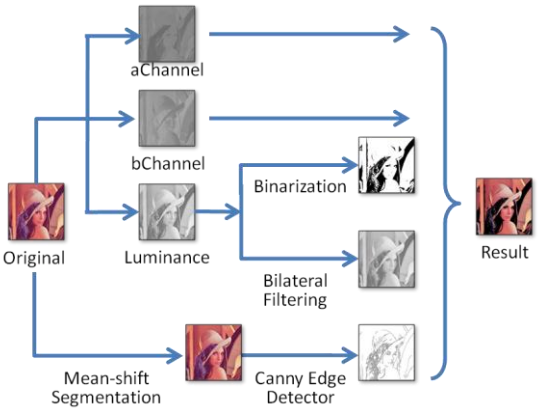

Figure 2. Overview of Non-Photorealistic Rendering

4.3 Result

We implement our framework using Visual C++ 2008 and OpenCV 2.0. Total length of test video is about 35 seconds and the frame rate is 30 frames per second (FPS). The computational time takes a few minutes. Figure 3 illustrate the preliminary result.

5. Conclusion

In this paper we describe a "Video2Manga" system for video summarization and visualization. In order to convert video content to "Manga" style, we mainly focus on these two steps: video key-frame extraction and non-photorealistic rendering. The result shows that our system not only can keep the most significant information of video but also can make it easy to understand.

There are still many tasks remained in the "Video2Manga" system such as layout arrangement, and speech balloon. For example, the "speech balloon" is one of the most expressive characteristics and an indispensable part of "Manga". We are working on how to generate the "speech balloon". There are several plans: extracting from subtitles, or by user interaction. In all, we expect our "Video2Manga" system can be applied to many possible fields such as video preview at portable devices, user-centered creation, etc.

Figure 3. A result by "Video2Manga" system

6. References

[1] M.M. Yeung, and B.L. Yeo, "Video Visualization for Compact Presentation and Fast Browsing of Pictorial Content," IEEE TCSVT, 7(5): 771-785, 1997

[2] S. Uchihashi, J. Foote, A. Girgensohn, J. Boreczky, "Video Manga: Generating Semantically Meaningful Video Summaries". In Proceedings of ACM Multimedia: 383-392, 1999

[3] X.S. Hua, S. Li, and H.J. Zhang, "Video booklet". In Proceedings of IEEE ICME: 189-192, 2005

[4] Y.F. Ma, and H.J. Zhang,"Video Snapshot: A Bird View of Video Sequence". In Proceedings of MMM: 94–101, 2005

[5] T. Mei, B. Yang, S.Q. Yang, and X.S. Hua,"Video Collage: Presenting a Video Sequence Using a Single Image", The Visual Computer, 25(1): 39-51, 2009

[6] Y.T. Zhuang, Y. Rui, Thomas S. Huang, and S. Mehrotra, "Adaptive Key Frame Extraction Using Unsupervised Clustering", In Proceedings of IEEE ICIP: 866-870,1998

[7] H. Winnem¨oller, S. C. Olsen, and B. Gooch. "Real-Time Video Abstraction". ACM Trans. Graph., 25(3): 1221–1226, 2006