Summarization of Egocentric Moving Videos for Generating Walking Route Guidance

Masaya Okamoto Yoshiyuki Kawano Keiji Yanai
Department of Informatics, The University of Electro-Communications
{okamoto-m,kawano-y,yanai}@mm.inf.uec.ac.jp

Abstract

In this paper, we propose a new application of egocentric videos which is automatically generating a walking route guidance video by summarizing an egocentric video recorded while walking [1]*. To do that, we analyze it by applying pedestrian ego-motion classification and crosswalk detection, and estimate an importance score of each section of the given video. Based on the estimated importance scores, we dynamically control video playing speed instead of generating a summarized video file in advance. In the experiments, we confired the effectiveness of the propose method.*

1. Introduction

In this paper, we propose a system to summarize an egocentric moving video. The goal of summarization of an egocentric moving video in this work is to produce a compact summary video, assuming that an input video is an egocentric video recorded continuously from a departure place to a destination by a wearable camera that a walking person is wearing. Therefore, the generated summary video can be used as a guidance video which explains the walking route from a departure place to a destination. In general, making a route guidance video manually is a time-consuming job. With this system, we can generate compact guidance videos on walking path routes very easily by walking along the routes with a wearable camera.

To produce a summarized video, we make use of egomotion (motion of the person wearing a wearable camera) and detection of a pedestrian crosswalk as cues. Egomotion is estimated based on optical flow of the scene, while crosswalks are detected by Geometric Context [1] and a standard object recognition method based on bag-of-features representation and non-linear SVM classifier. Based on these cues, the system estimates importance of each section in the given video.

[1]The original paper was preseted at the Pacific-rim Symposium on Image and Vision Technology (PSIVT) 2013 as [2]

Actually, in the proposed system, summarized video files are not generated. Instead, a playing scenario is generated which instructs relative playing speed on each video section according to the importance scores of video sections. Basically, we play important scenes at a normal speed, while we play less important scenes at a high speed or skip them. We regard turning on a corner or a crossing as an important scene, and crossing a crosswalk as very important to remember walking routes. Acoording to the generated playing scenario, the system dynamically controls video playing speeds. Figure 1 shows a screenshop of the walking route video viewer which is implemented as an Web application integrated with an playing-speed-controllable HTML5 video player. "x 2.80" in the Figure represents the current playing speed, which is a relatively slow speed, because the maximum speed is set as 6.0 times in this case.

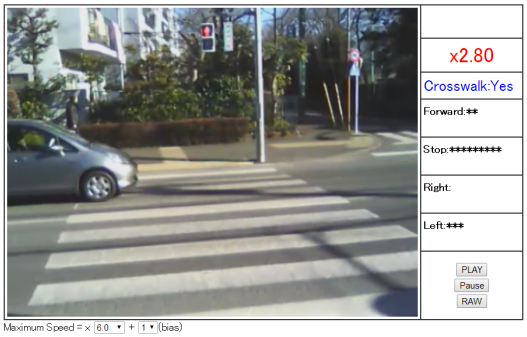

Figure 1. The walking route video viewer where the playing speed are dynamically controlled based on ego-motion and crosswalk detection. This scene is estimated as an importance scene and the playing speed is set with a relatively low speed, because crosswalk is detected.

2. The Proposed System

2.1. Overview

In the proposed system, we estimate importance of each video section of the given egocentric video based on egomotion and existence of crosswalks, and generate playing scenario off-line. Based on the scenario, we summarize the

given video dynamically by controlling playing speed of the HTML5 video player on-line.

2.2. Ego-motion Classification

Since we assume that the given video is taken while walking, we estimate likelihood of four types of ego-motion conditions on each four-second-long section of the given video: (1) moving forward, (2) stopping, (3) turning right, and (4) turning left. To do that, firstly we extract twelve frame images per second from the given video section, and secondly estimate optical flows for each interval of frame images using the Lucas-Kanade method, and thirdly we convert them into histogram-based representation. Forthly, we apply one-ve-rest SVMs to the feature vector of each frame, and finally we average output values over the given section. Note that we use the pseudo-probability values obtained by applying a sigmoid function to the output values of SVMs.

2.3. Crosswalk Detection

For walking route guidance in cities, crosswalk is an important and remarkable cue to explain walking routes. To prevent the scenes containing crosswalks from being skipped or played in a high speed, we detect pedestrian crosswalks in the given egocentric video as well as ego-motion conditions.

For crosswalk detection, we extract three frames per second, and detect the frames including crosswalks after road region estimation for each extracted frame.

We estimate road regions for each extracted frame. To do that, we use Geometric Context [1]. We regard ground regions as road regions which may contain crosswalks.

To detect crosswalks in the groud regions finally, we use standard bag-of-feature representation with densely-sampled SIFT and SVM with non-linear chi-square-RBF-kernel.

2.4. Estimation of importance of video sections

Based on the results of ego-motion classification and crosswalk detection, we estimate importance score of each video section of the given egocentric video. We divide the given video into video sections every four seconds. The importance score varies between 0.0 and 1.0, which decides playing speed of the corresponding video section. Refer to [2] for the detail.

3. Experiments

For this experiment, we prepared nine ego-centric videos recorded while walking, the average duration of which are about nine minutes.

First, we evaluate the performance of ego-motion classification. The results are shown in Table 1. The classification rate over four-kinds of ego-motion was 83.8%.

Table 1. Ego-motion classification result

Motion	# section	recall	precision	F-number
Forward	244	0.943	0.697	0.801
Stop	72	0.694	0.893	0.781
Go right	84	0.738	0.969	0.838
Go left	88	0.795	0.972	0.875

Next, we evaluate the performance of crosswalk detection. The experimental results are shown in Table 2.

Table 2. Crosswalk detection accuracies

Method	Classification Accuracy
w/ ground estimation	0.635
w/o ground estimation	0.605

Finally, we show the results of user study employing ten subjects. To evaluate the proposed method on walking egocentric video summarization, we compared the proposed summarization method with two baseline methods. The first baseline is just playing videos in fast-forwarding at a uniform speed. The second one is a storyboard-style summarization which displays uniformly sampled frames from the video. The numbers of storyboard frames are proportional to the length of the given video. We sampled a frame every five second from the given video. To evaluate the effectiveness of crosswalk detection, we carried out a method using only ego-motion classification without crosswalk detection as well.

We asked the subjects to vote the best summary as a walking route guidance among the storyboard and the videos generated by three kinds of summarization methods for each of the egocentric video. As a results, the proposed method gathered the best votes on average over three test videos, which means the proposed method based on ego-motion and crosswalk detection was effective compared to the two baselines and the method without crosswalk detection.

4. Conclusions and Future Work

In this paper, we proposed a new method to summarize walking ego-centric video for generating walking route guidance.

For future work, we plan to treat with a bike and a car egocentric video. In such cases, other object cues are expected to be importance for egocentric video summarization. We plan to add detection methods on other important objects to the system.

References

[1] D. Hoiem, A. Efros, and M. Hebert. Recovering surface layout from an image. *International Journal of Computer Vision*, 2006. 1, 2

[2] M. Okamoto and K. Yanai. Summarization of egocentric moving videos for generating walking route guidance. In *Proc. of Pacific-rim Symposium on Image and Vision Technology (PSIVT)*, 2013. 1, 2