Automatic Street View System Synchronized with TV Program using Geographical Metadata from Closed Captions

Yuanyuan Wang
Kyoto Sangyo University
circl.wang@gmail.com

Daisuke Kitayama
Kogakuin University
kitayama@cc.kogakuin.ac.jp

Yukiko Kawai
Kyoto Sangyo University
kawai@cc.kyoto-su.ac.jp

Kazutoshi Sumiya
University of Hyogo
sumiya@shse.u-hyogo.ac.jp

ABSTRACT

Various TV programs, such as travel and educational programs, often introduce tourist spots, or historical places. However, viewers find it difficult to grasp the surroundings of these spots or places, how the locations are related, and distances between them, when instantaneously moving between the scenes by switching video streams. Therefore, we built an interface using the geographical metadata of the video streams, called TV-Milan, that automatically presents geographic contents (i.e., maps, photos, Street View, etc.) synchronized with the TV program.

Categories and Subject Descriptors

H.5 [[Information Interfaces and Presentation]]: Multimedia Information Systems, User Interfaces, Hypertext/Hypermedia

General Terms

Design, Human Factors

Keywords

TV program, geographical metadata, Street View System

1. INTRODUCTION

TV programs, such as travel and educational programs, provide geographical information of tourist spots or historical places in video streams. However, when the video stream scene instantaneously changes from one place to another, viewers find it difficult to grasp the surroundings of such places and the positional relationships and distances between them. For instance, suppose a TV program introduces tourist spots in Kyoto, Japan. When one scene shows a restaurant in Gion and the next scene shows tourist attractions in Yasaka, the viewers cannot follow the movement between Gion and Yasaka or understand their relative locations.

In this study, we built a TeleVision Map Interface for Location AwareNess, called TV-Milan. This system synchronizes geographic information (i.e., maps, photos, Street View, etc.) with a TV program on the basis of the intentions of a video stream by using geographical metadata from closed captions, which aids the viewer's understanding of geographical information (i.e., tourist spots, routes, etc.) in the video stream. As depicted in Figure 1, the proposed model, TV-Milan, can be implemented by first extracting the names of places and their time of appearance in the video stream, and then extracting the geographical relationships between the places.

When scenes are switched instantaneously in the video stream, TV-Milan presents a route between the two places by using Google Maps Street View. These two places appear in the closed captions supplement for the TV program. In addition, to aid the viewer's understanding of the surroundings of the places and their relative locations, TV-Milan presents not only Google Maps Street View but also Google Maps or photographs in Google Earth uploaded through Panoramio. Therefore, as users view the TV program, TV-Milan enables them to easily and efficiently grasp the geographical information of the various places from the closed captions of the video stream.

2. EXTRACTING GEOGRAPHICAL METADATA OF VIDEO STREAMS

We extract geographical metadata, including temporal and geographical relationships, by extracting place names from the closed captions of video streams.

2.1 Extracting Temporal Relationships

We first extract the temporal relationship as order of appearance that the appearance time of the place names from the closed captions of a video stream. Then, we extract the geographical relationship between any two places in the order of their appearance in each scene of the video stream. Specifically, we consider that one video stream often consists of various topics and classify scenes according to the topic

AVI '14, May 27 - 30, 2014, Como, Italy
ACM 978-1-4503-2775-6/14/05.
http://dx.doi.org/10.1145/2598153.2600028

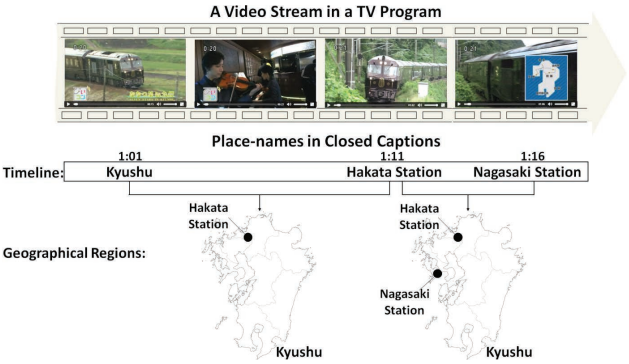

Figure 1: Conceptualdiagramof a TV video stream

by setting a time width for each video stream.

2.2 Extracting Geographical Relationships

We extract geographical and regional relationships between two places on the basis of their temporal relationships in each scene of a video stream. We utilize 9-intersection[1] to extract six types of geographical relationships: *Equal* means that the same place name appears repetitively. *Disjoint* means the regions of two places are separated. *Meet* means the regions of two places have an overlapping boundary. *Overlap* means the regions of the two places are overlapping. *Contains*(*inside*) means the region of one place includes (or is included in) the region of another place.

3. PRESENTING GEOGRAPHIC CONTENTS SYNCHRONIZED WITH TV PROGRAM

3.1 Extracting Intentions of Video Streams

We extract the intentions of the video streams on the basis of geographical metadata; i.e., the temporal and geographical relationships between given places, A and B, that appear in closed captions. When A *equal* A, we assume that the intention is to introduce A. When A *contains* B, we assume that the intention is to introduce A and B. When A *disjoint* B and each place has its own wide region (i.e., prefectures), we assume that the intention is to compare A and B. When A *meet* B, we assume that the intention is to compare A and B. When A *disjoint* B and each place has its own narrow region (i.e., municipalities), we assume that the intention is to describe the route between A and B. When A *overlap* B or A *inside* B, we assume that the intention is to introduce the surrounding areas of A and B.

3.2 Presenting Geographic Contents

We built a novel interface called TV-Milan (see Figure 2). This system synchronizes geographic content (i.e., maps, photos, Street View) with the TV program on the basis of the derived intentions of the video streams, which aids viewers to better understand the geographical information (i.e., tourist spots, regions, routes, etc.) presented. The interface is implemented by a video player window, a list of place names, a map window, and a Street View window.

[1]Y. Kurata. An Overview of the Research on Topological Relations and Future Issues in GIScience (in Japanese). Theory and Applications of GIS, 18(2):41-51, 2010.

Figure 2: Screenshot of TV-Milan

For example, suppose three place names, Kyushu (1:01), Hakata Station (1:11), and Nagasaki Station (1:16) appear in a video stream (see Figure 1), which is a travel program introducing Kyushu in Japan, with total run-time of 23 minutes. We first extract geographical metadata: Kyushu *contains* Hakata Station, and Hakata Station *disjoint* Nagasaki Station. When the video stream shows Kyushu to Hakata Station in the video player window, the map screen presents the region of Kyushu and zooms-in to Hakata Station. Simultaneously, a Street View window first presents the photos related to Kyushu. Then, when the video stream shows Hakata Station, the Street View window changes to present photos related to Hakata Station. When the video stream shows Hakata Station to Nagasaki Station, the map window presents both Hakata Station and Nagasaki Station and highlights the route between them. Simultaneously, the Street View starts at Hakata Station and proceeds along the route to Nagasaki Station. In this manner, TV-Milan enables viewers to easily and efficiently grasp the geographical information of places that appear in the video streams.

4. CONCLUSIONS

In this paper, we built an interface, called TV-Milan, which presents geographic content (i.e., maps, photos, Street View) synchronized with a TV program. We used actual travel programs in Japan with our developed prototype system, and we were able to confirm that TV-Milan helps users to easily and efficiently view the TV program while simultaneously viewing geographic information.

In the future, we intend to improve the method for presenting geographic contents by considering cinematography and film languages. Further, we plan to expand TV-Milan to allow user interactions with the video streams for controlling the geographic content to be displayed.

Acknowledgments

This work was supported in part by Strategic Information and Communications R&D Promotion Programme (SCOPE), the Ministry of Internal Affairs and Communications of Japan.