MSVA: Musical Street View Animator: An Effective and Efficient Way to Enjoy the Street Views of Your Journey

Yin-Tzu Lin, Po-Nien Chen, Chia-Hu Chang, Ja-Ling Wu
National Taiwan University
{known,dudebrotim,chchang,wjl}@cmlab.csie.ntu.edu.tw

ABSTRACT

Google Maps with Street View (GSV) provides ways to explore the world but it lacks efficient ways to present a journey. Hyperlapse provides another ways for quick glimpsing the street-views along the route; however, its viewing experience is also not comfortable and could be tedious when the route is long-distance. In this paper, we provide an efficient and enjoyable way to present street view sequences of long journey. Street view journey video accompanied with locally listened music will be produced by the proposed approach. During the move between locations, we use the speed control techniques for animation production to improve the viewing experience. User evaluation results show that the proposed method increases the satisfaction of users in viewing the street view sequences.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems

General Terms

Design, Human Factors

Keywords

musical video; google street view; animation; musical medley

1. INTRODUCTION

All over the world, there are lots of remarkable places well worth visiting once in our lifetime. Nowadays, making a tour around the world becomes more and more convenient with the increasingly rich travel information and well-developed transportation systems. While traveling abroad, we can see the attractive scenes, the historical buildings, the street-views with different cultures, and hear the music played locally. All of these stimuli of fresh perceptions can relax our body and free our mind from burdens and stresses.

MM'14, November 3–7, 2014, Orlando, Florida, USA.

http://dx.doi.org/10.1145/2647868.2654967 .

Google Maps with Street View (GSV) provides map-based web services to let users explore places on the map and see the 360-degree images of correspondent scenery around most of the world. These spherical panoramas were created by stitching the photos taken from various Google-adapted vehicles. Then, each spherical panorama point was aligned to the correspondent coordinate on the map based on their geo-information. Therefore, users can click arrow in GSV to move toward the image closest to a specifically chosen location. While transiting from current panorama to the nearest place, GSV utilizes a rough 3D model, where the panorama photos are mapped onto, to create an animated perspective change on the crossfaded areas of images. In this way, the transitions between distinct panoramas will be smoothed and produce a walk-through effect as well, although the point-to-point panoramas are sparse. However, if we try to continuously view all the street views along a trip route, it would be tedious because one have to keep clicking the arrow continuously. In addition, all the generated views cannot be easily shared with others. Although users can search for a series of attractions in web-based travel communities and explore the paths between them via GSV, all the planned and/or interested path cannot easily be organized and recorded into the touring videos.

Teehan+Lax Labs's Google Street View Hyperlapse tool [11] provides the time-lapse photography effect in viewing GSV. The images are played with constant and fast speed, but the video becomes long and tedious when long-distance street view sequence is the target, which makes the viewer feel dizzy when the viewing time is long.

The collections of GSV panoramas keep increasing and publicly available. How to well make use of such tremendous but sparse and noisy street view panoramas to generate a pleasant journey video accompanied with the locally played music, for promoting tourism and sharing travel experiences, is an important but challenging issue. In this paper, we provide an efficient and enjoyable way to present street view sequences, as shown in Figure 1. The flow chart of the proposed system is shown in Figure 2. The processes for each step will be detailed in the following sections.

2. INCIDENTAL MUSIC CREATION

There are many ways to match music with visual contents, e.g. low level content-aware matching (color↔chord, motion↔tempo [7]) or high level semantic matching (emotion/mood [1], context [3]). In our application, we would like to provide users the music played locally at certain time. Thus, we take the geo-information as the cue in choosing music to accompany street view sequence. In order to han-

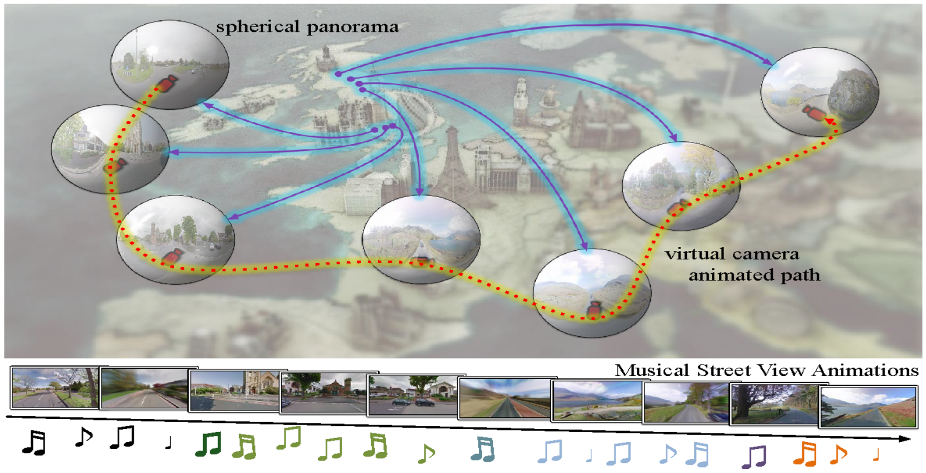


Figure 1: This figure depicts the essential idea of the musical street view animation system. Given a set of attractions' locations and the viewing orientations, our system can dynamically estimate the virtual camera orientations and pacing according to the proposed animated transition schemes, which are inspired by 12 basic principles of Disney's animation [4], for each Google street view panoramas along a navigating path. Then, the street view image sequences including the specified input key frames and the drive-through frames are generated. In addition, the music, which is popularly listened by the people at or near the locations, will be mined from the social media (e.g., Twitter) and be highlighted to create an incidental music. With these, the system automatically generates the musical street view animation to provide an efficient and enjoyable way to dramatically glide through the attractions.

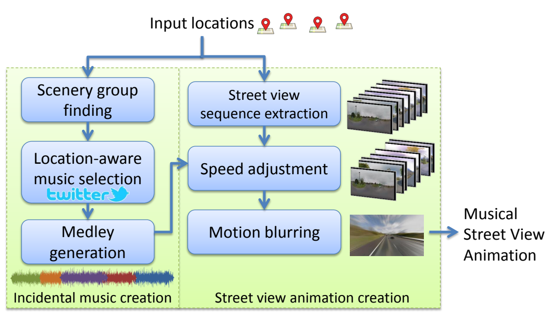


Figure 2: Flowchart of the proposed system

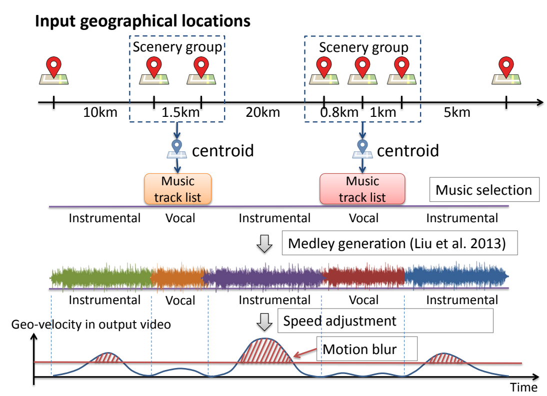


Figure 3: Schematic diagram of our system

dle different music tracks picked by multiple locations while keeping the completeness and smoothness of the incidental music, we use the method proposed in [8] to generate musical medleys as our incidental music. A medley is a kind of music piece that is composed of existing music excerpts. With Liu et al. [8]'s method, it is able to create medleys according to the specified conditions, such as the medley structure or the must-use clips. The pre-described flexibility makes us easy to create incidental music that could match to the conditions in our application. We design an incidental music creation process as shown in Figure 3. First, we group locations with the distances less than a certain value (e.g., 2.5km) as a scenery group. Then, we specify our target medley structure according to the number of the scenery groups. We assign the positions of scenery groups as vocal segments and other parts as instrumental segments. The reason is that vocal parts of a popular song are often thought to be topics of a song and listeners may pay more attention to vocal parts. The scenery group contains neighboring tourist attractions would also grab higher attention from the audience. Take Figure 3 as an example, the medley structure is Instrumental-Vocal-Instrumental-Vocal-Instrumental. After specifying the medley structure, the specified vocal segments should be chosen from the music rank list that is generated according to the centroid of each scenery group. Then, the other parts of the medley will be auto-completed by [8].

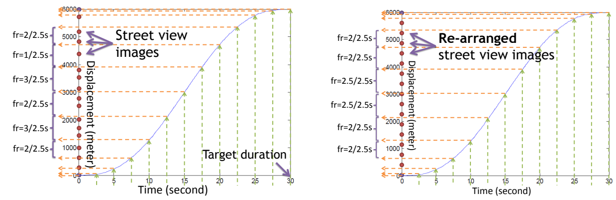
2.1 Location-aware Music Selection

In this section, we describe the details of selecting music tracks according to the given GPS coordinates of a scenery group. There are several methods to recommend music according to location cues, for example, choosing traditional music with respect to the input location [9], or as Kaminskas et al. [5] proposed, a place of interest (POI) can be matched to music tracks by their relationships on the ontology graph, or by their overlap ratio of emotional and physical tags. Here we used musicmicro dataset provided by Schedl [10] for its easy accessibility. In the dataset, Schedl collected 594,306 music-related microblogs (e.g., containing hash tags #nowplaying, #itunes, etc.) by 136,866 users in 722 different cities in 180 countries on Twitter from November 2011 to September 2012. Such popular musics from the past can bring audience back in time and evoke their old memories. Each record r in the dataset reports a user was listening to a music track t_r at certain geo-location p_r (by GPS coordinates) at a certain time (month, weekday). We would like to pick music tracks that are listened by users near to the query location and the number of listened times are also high enough (popular around the area). So, we first take the listening records at locations with distances less than a given distance D to the query point as candidates (We could also choose the records in the same city or country of the query location as candidates). For each candidate record r, we define a weight score $w(q,r) = 1 - \frac{gdist(q,p_r)}{D}$, where q is the GPS coordinate of the query point, and $gdist(q,p_r)$ is the geo-distance of q and p_r computed by Haversine formula[1]. Then, we accumulate the weight of each record according to the music track listened. That is, the score of each track t is defined as $score(q,t) = \sum_{r \in R_D(q,t)} w(q,r)$, $R_D(q,t) = \{r | gdist(p_r,q) \le D \text{ and } t_r = t\}$. Finally, we add another weight to the final score to eliminate the effect of repeatedly choosing widely listened songs. That is, $scoreW(q,t) = \log \frac{|R|}{|R(t)|} \cdot score(q,t)$, where, R is the set of all records, and $R(t) = \{r | t_r = t\}$. This weighting mechanism is similar to the concept of inverse document frequency (IDF) widely adopted in the field of text retrieval.

3. STREET VIEW ANIMATION CREATION

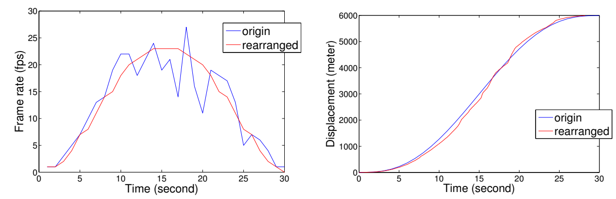
To generate vivid street view animations, we utilize Google's API to estimate a routing path connecting the given points, and then retrieve all the panoramas along the path. After projecting the panorama images onto a spherical model, the street view images can be rendered according to specified viewing angles. Then, the system can create animations by taking the street view images as the source materials to be displayed in a rapid succession. Generally, the time intervals between adjacent, yet distinct, street view images are set as a constant value. However, such kind of constant animation pacing tends to be dull especially for viewing large-scale street views. Inspired by 12 basic principles of Disney' s animation [4], we proposed the following animated transition scheme. When moving toward next attraction far from current location, the transition pacing will be gradually increased at the starting point, reached an extreme in the middle, and then be gradually decreased at the end. With such a slow-in/slow-out effect (SI/SO), the street view animations can direct viewers to carefully watch the remarkable attractions but quickly pass through the distant views between them. As for the attractions that are close to each other, the transition pacing will be kept constant. In this way, the proposed system can create a "scenic car driving" metaphor, which is similar to Smart Player [2], and provide an effective and efficient way to display the street view images.

[1]`https://en.wikipedia.org/wiki/Haversine_formula`



(a) Original (b) Rearranged

Figure 4: Schematic diagram for calculating the frame-rate.



(a) fr-t plot (b) D-t plot

Figure 5: Fr-t and D-t plot of original and rearranged animations.

3.1 Speed Adjustment

By conducting preliminary tests with 9 users, we found that the actual speed perception is not simply frame-rate dependent, but is highly affected by the geo-velocity (calculated via geo-displacement over time interval between street view images). However, frame-rate variation may still cause unsmoothness in visual motion perception. On the basis of these observations, we design the following playing speed adjustment method to reduce the effect of speed variations. A five-order Bézier curve is used to approximate the easing function for simulating SISO effects. The relation of geo-displacement of each street view image over time can be shown in Equation (1), where $\mathbf{s}(t)$ is the geo-displacement of the current street view image (at time t) to the start point, and $\mathbf{P}_2$, $\mathbf{P}_3$ are control points of the chosen Bézier curve[2].

$$\mathbf{s}(t) = 10t^2(1-t)^3\mathbf{P}_2 + 10t^3(1-t)^2\mathbf{P}_3 + 5t^4 - 4t^5. \quad (1)$$

We let $\mathbf{P}_2 = \frac{d_s}{S}$, $\mathbf{P}_3 = 1 - \frac{d_e}{S}$, where S is the geo-distance between the start and the end points, d_s and d_e are the distances of the positions that we wish to start speeding up and slowing down, respectively. Then, a practical method to compute the frame-rate of the street view sequence is illustrated in Figure 4(a). The vertical and the horizontal axes of the generated curve are first scaled to the real geo-displacement and the target duration that we wish the street view animation to play, respectively. The target duration can be set according to the duration of the incidental music created in Section 2. Then, the street view images can be plotted on the vertical axis, as illustrated by the red dots in Figure 4(a). Afterwards, we divide the target duration into equal length intervals, and for each time interval, we calculate the frame-rate of that interval by counting how many street view images lie in the corresponding displacement interval. The resulting frame-rate to time (fr-t) plot and the geo-displacement to time (D-t) plot are shown in Figure 5.

[2]$\mathbf{P}_0 = \mathbf{P}_1 = 0$, $\mathbf{P}_4 = \mathbf{P}_5 = 1$

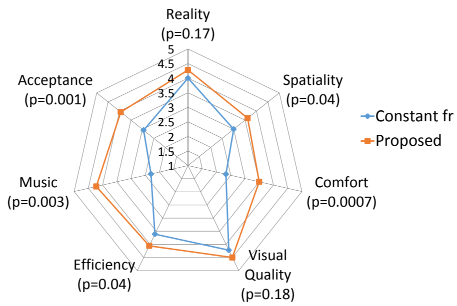


Figure 6: The average score of each performance evaluation dimension.

However, we found that the frame-rate moves up and down along the time because the geo-distance between consecutive street view images varies with the situation encountered while the photos were taken, e.g., the driving speed. This effect causes users to perceive unsmooth motions. Fortunately, the variation of the geo-distance intervals between consecutive street view images is not large. As a result, we rearrange the street view images to make the number of images lay in the time intervals uniformly distributed along the time axis, as shown in Figure 4(b). Then, the resulting fr-t plot (Figure 5(a)) can be smoothed along the time axis while not causing too stiff penalties for the D-t plot (Figure 5(b)).

3.2 Motion Blur

To further reinforce the impression of high speed and enhance the smoothness in visual motion perception, a velocity-based motion blur filter is applied to the street view images when the geo-velocity is higher than a threshold (as shown in the bottom part of Figure 3). While moving forward, the vanish point of each street view, which indicates the motion direction, will be estimated by using the method proposed in [6] and be set as the center of the applied zoom blur effect. The higher the speed is, the more the blur amount is applied. By leveraging the motion blur effect, the viewing experience can be enriched in a less dizzy manner.

4. EVALUATION

In the user evaluation, we compared two street view animation schemes: constant rate [11], and the proposed one. In the constant playing scheme, the street view image sequences are played with a constant frame rate of 30 fps (frame per second) without incidental music. Seven locations around Sutton, London and Buttermere, North West England are selected as inputs to create street view animations[3]. Eighteen users (16 males and 2 females) aged around 20-60 participated in the evaluation. Evaluators were asked to grade from 1 to 7 on both street view animation to show their satisfactions (higher scores for better satisfaction) with respective to each of the following perspectives:

- *Reality*: How do they feel about the reality of "scenic car driving" in the street view animation?
- *Spatiality*: How strong sense of the geographical relation does the street view animation offer?
- *Comfort*: How comfortable while viewing the street view animation? (dizzy feeling means less comfort)
- *Visual Quality*: How smooth is the visual motion perception in the street view animation?
- *Efficiency*: How efficient is the way of previewing scenic spots via viewing the street view animation?
- *Music*: How is the incidental music matched with the street view animation?
- *Acceptance*: How do they feel about the overall system?

The average scores are illustrated in Figure 6. The differences between the proposed and the constant frame-rate methods in terms of reality and visual quality are relatively small while in all the other perspectives, the proposed method statistically significantly outscored the constant rate counterpart. The result indicates that adding animation effect will not affect the reality and visual quality of the street view, but can increase efficiency and comfort. Besides, the evaluators are satisfied with the match of incidental music and the street view animation with the average score higher than four.

5. CONCLUSIONS

In this paper, we provide an efficient and enjoyable way to create street view journey videos, by accompanying with locally listened musics and adding animation effects to it. The user evaluations show that the proposed method is effective in enhancing the feelings of efficiency and comfort. We believe our work can assist people easily create and share attractive journey videos with music according to their own trip path. In addition, more tourism promotion videos can be produced to benefit the travel industry. In the future, we plan to conduct more user tests in deciding a proper compromise between the geo-velocity and the smoothness of frame-rate to further improve the comfort of the video. Besides, other music selection methods are also worthy of exploring to enhance the entire viewing experience.

6. REFERENCES

[1] C.-H. Chen, M.-F. Weng, S.-K. Jeng, and Y.-Y. Chung. Emotion-based music visualization using photos. In *Proc. MMM*, pages 1–11, 2008.

[2] K. Cheng, S. Luo, B. Chen, and H. Chu. SmartPlayer: user-centric video fast-forwarding. In *Proc. SIGCHI*, pages 789–798, 2009.

[3] P. Dunker, P. Popp, and R. Cook. Content-aware Auto-soundtracks for Personal Photo Music Slideshows. In *Proc. ICME*, 2011.

[4] O. Johnston and F. Thomas. *The Illusion of Life: Disney Animation.* 1981.

[5] M. Kaminskas, F. Ricci, and M. Schedl. Location-aware music recommendation using auto-tagging and hybrid matching. *Proc. RecSys*, pages 17–24, 2013.

[6] H. Kong, J.-Y. Audibert, and J. Ponce. General road detection from a single image. *IEEE Trans. IP*, 19(8):2211–20, Aug. 2010.

[7] Y.-T. Lin, T.-H. Tsai, M.-C. Hu, and W.-H. Cheng. Semantic based Background Music Recommendation for Home Videos. In *Proc. MMM*, pages 283–290, 2014.

[8] I.-T. Liu, Y.-T. Lin, and J.-L. Wu. Music Cut and Paste: A Personalized Musical Medley Generating System. In *Proc. ISMIR*, Curitiba, PR, Brazil, 2013.

[9] C. Parker, D. Joshi, P. Lei, and J. Luo. Finding geographically representative music via social media. *Proc. MIRUM*, page 27, 2011.

[10] M. Schedl. Leveraging microblogs for spatiotemporal music information retrieval. In *Proc. ECIR, LNCS 7814*, Moscow, Russia, 2013.

[11] Teehan+LaxLabs. Google Street View Hyperlapse. http://hyperlapse.tllabs.io/.

[3]The generated animations can be found at `http://www.cmlab.csie.ntu.edu.tw/~known/street/results/`