A Virtual Telescope for Augmented Reality Application

EE368 Digital Image Processing Final Project Proposal, 2012 Spring

Group Members: Qiyuan Tian (qytian@stanford.edu)
Qingyi Meng (qingyim@stanford.edu)
Android Phone-based: Yes

Description

We propose an Android phone-based application "Virtual Telescope" for viewing distant buildings by displaying to the users the replacement images retrieved from the Internet.

People wish to more closely and clearly view specific buildings when they have a overhead and panoramic view of the city, standing on the mountain or in a high building. A telescope does great help but it is often bulky, usage specific and inconvenient to carry. People could also turn to smart phones and continuously zoom in the phone cameras to make distant buildings appear larger and nearer. Unfortunately, average phone cameras have very limited optical power, on which level the enlarge effect is not satisfying. Based on the fact that most people do not care whether the view is of the moment, our application acquires images of the same building taken by others from the Internet and displays them to replace the actual viewfinder image. The further implementation could even show users indoor scenes of the interested building as guiding them around. To some extent, our "virtual telescope" is a typical augmented reality application.

The application includes three steps:

1, In order to retrieve images containing the same building, the application has to perform descriptor extraction algorithm, such as SIFT [1] and SURF [2] on the viewfinder image, and send the descriptors to the server to match the known images available in the database.

2, The server compares the descriptors with known objects in the database, and choose the match with the greatest number of equivalent descriptors. It then returns the matched image to the phone.

3, The phone displays the returned image to the users.

There are several technical difficulties for each step:

1, When to start search or which viewfinder image to use for image retrieval. If the application starts search very early, the search process is likely to be accomplished

before users reach the limitation of the optical system and the found result can be shown instantly without any delay. However, the early-stage viewfinder images contain other objects, presenting challenges to the matching process. On the other hand, if the application starts search late until to the stage that viewfinder images only contain the interested building, users may have to wait for extra time, which worsens user experience. This requires our method to be extremely computationally efficient.

2, The enlarged viewfinder images often have so low resolution that the distant buildings may only occupy several pixels. It is reduces the accuracy of image retrieval. We have to choose a powerful descriptors extraction algorithm [3] and enhance the query image in some sense.

3, More than one photos of most buildings could be matched by the server, but which one should it return to the phone? It is related to the problem how to display the returned image to the users. The ultimate goal is to naturally incorporate the returned image to the zooming in process, making it unnoticeable to the users [4]. Therefore, the chosen image has to be as similar as possible to the image sent to the server on criteria like perspective, illumination (especially day and night).

It takes time to overcome the aforementioned difficulties. For the course final project, we decide to conduct a simple version of the completed project to prove the idea. To be specific, we plan to manually store in the phone a high resolution overhead and panoramic image (like image of Stanford campus taken on the Dish Mountain). The user interface allows users to zoom in and zoom out image to observe different buildings in details (like Hoover Tower). When some building is enlarged to some level, the application computes the descriptors and send them to the server, which has already stored several known images (like all building images of Stanford Campus). The server conducts the comparing process and sends the matched image to the phone, which simply displays the returned image on the viewfinder.

References

[1] Lowe, D. Object recognition from local scale-invariant features. In International Conference on Computer Vision 2 (1999), pp. 1150-1157.

[2] Bay, H., Tuytelaars, T., and Gool, L. Surf: Speeded up robust features. In European Conference on Computer Vision (May 2006), no. 9.

[3] D. Q. Huynh, A. S. Saini, and W. Liu. Evaluation of Three Local Descriptors on Low Resolution Images for Robot Navigation. Image and Vision Computing New Zealand (IVCNZ), Wellington, New Zealand, pp. 113-118, Nov 2009.

[4] R.T. Azuma, A survey of augmented reality, Presence-Teleoperators and Virtual Environments, Vol. 6, No. 4, pp. 355--385