

Photo Tourism: Exploring Photo Collections in 3D

In their paper, the authors propose a computer vision algorithm in which pictures of famous places are processed to create a photo tourism application. The system uses images collected from the Internet and processed in a way to account for being noisy in information. The spectrum of the images, the authors mention, is greatly varied in terms of illumination, season, weather condition and similar challenges that computer vision algorithm needs to account for. Overall, the system is capable of visualizing famous scenes in 3D, object-based tourism (it is important for tourism applications; some objects are more important than their vicinity), understanding where the photographer was when the picture was taken and distinguish what is being photographed.

Results

The authors provided a great detail about the set of computer vision algorithms has been utilized for the system, while providing some details about the experimentation. The system was evaluated with two different data sets, controlled and uncontrolled the authors refer. In the controlled experimentation, the system was tested with two datasets collected in two different cities under controlled settings such as single camera-lens with single photographer. As for the uncontrolled experimentation, the authors used images collected from the Internet. The authors claim, the average reprojection error was about 1.5 pixels.

Conclusions

The authors imply that the structure from motion technique utilized in the proposed system was the slowest part of the pipeline. Even though, the paper published 10 years ago and the amount of data being handled is vast (more than 2500 images), the authors noted that it took two weeks for system to utilize Structure from Motion, which sounds far too slower than rest of components of the system. A quick search resulted that some cameras released in 2008 were referred as early geotagging cameras, I believe a similar system can rely on EXIF data more than the authors could. One extension could be made in regards to that point is to incorporate the output of the structure from motion and geo-registration algorithm with the EXIF data in a Bayesian Framework.

The novelty of the proposed system is an end-to-end computer vision system capable of providing photo tourism experience. The video provided by the authors was impressive, and I am glad to see the website forming the user interface is still alive. Moreover, it is historically important to see when computer vision applications started affecting human life in more tangible ways than being part research topics in labs. However, a point that I should mention is that it would be better if more parametric study was provided. It would be more clear if The system heavily depends on the theoretical of computer vision field and only evaluation metric provided was the reprojection error. Even that metric was averaged out, was not compared after two different experimentation done.