

Double-Bounding R-tree: A Novel Approach to Index Digital Camera's Viewable Scene

He Ma, Yue Zhao, Guohui Wei, Xiaoyu Cui

Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang, 110169, China

email: {mahe, cuixy, zhaoyue}@bmie.neu.edu.cn

Keywords: Video Indexing, Meta-data Management, Query Processing

Abstract. Geographic Information System (GIS) applications now increasingly make use of geo-located multimedia data such as images and videos. Furthermore, the widespread use of smartphones (and increasingly tablets) and the rapid improvement of hardware has enabled the acquisition of high-definition user-generated videos that are annotated with geo-properties. The sensor meta-data (*e.g.*, GPS and digital compass values), which associate a continuous stream of location and viewing direction information with the collected videos, are considerably smaller in size than the visual content and are helpful in effectively and efficiently manage and search through large repositories of videos. In this paper, we proposed the Double-Bounding R-tree (DBR-tree), which utilizes the inner-bounding rectangle and outer bounding rectangle to manage the camera's field of view. Experimental results show that this structure can prune the unnecessary overlap calculation and improve the spatial query performance.

Introduction

Recent developments in video recording technology have enabled user-generated video (UGV) production on a daily basis. For example, smartphones (and increasingly tablets), which are carried by users all the time, have become extremely popular for capturing and sharing online videos due to their convenience, enhanced quality of images and wireless connectivity. The multimedia community nowadays has increasing interest in GIS applications, *e.g.*, Flickr and Panoramio allow users to upload images and videos with attached geo-locations of the cameras. Furthermore, embedded sensors (*e.g.*, GPS and compass units) have been cost-efficiently deployed on mobile devices. Consequently, some very useful meta-data of videos, especially geographical properties of video scenes, can easily be collected during video capturing. This association of video scenes and their geographic meta-data has led to interesting research topics in the multimedia community. For instance, the recorded sensor meta-data can be utilized to aid in modeling, indexing and searching of geo-tagged videos at the high semantic level preferred by humans.

Current video search techniques that annotate videos based on the visual content are struggling to achieve satisfactory results in online user-generated videos (UGVs), particularly in accuracy and scalability. In this work, we utilized the camera's Field-of-View (FOV for short) to convert the visual content into a series of spatial objects. Compared to visual content, the meta-data occupies much less space, which makes searching among large scale of videos practical. To improve the performance of our approach, a meta-data indexing algorithm is presented. Experimental results show that the proposed methodology outperforms well in both effectiveness and efficiency.

Related Work and Preliminaries

The camera viewable scene is what a camera in geo-space captures. This region is referred to as camera field-of-view (FOV in short) with the shape of a pie-slice shape^[1]. As shown in Fig. 1, the FOV coverage in 2D space can be defined with four parameters: the camera location P , the camera viewing direction vector \vec{d} , the viewable angle α , and the maximum visible distance R_V . The camera location P is the (latitude, longitude) coordinate read from a positioning device (e.g., GPS and/or Cricket coordinates^[2]). The viewing direction vector can be acquired from a digital compass. The camera's viewable angle α is calculated based on the lens properties for the current zoom level^[3]. The visible distance R_V is the maximum distance at which a large object within the camera's FOV can be recognized. Thus, the camera viewable scene at time t is denoted by the tuple $\text{FOV}(P \langle \text{lat}, \text{lng} \rangle, \vec{d}, \alpha, R_V, t)$. The tuple values are acquired from embedded sensors during video recording.

Associating geographic (e.g., location and orientation) information with a camera's video data for retrieval purposes has become an active research area^[4, 5]. Authors^[1, 6] proposed a model to represent a camera viewable scene with a shape of a pie-slice, based on the GPS and digital compass data as well as the visible distance. Ma *et al.*^[7, 8] proposed a grid-based index structure to speed up the searching process when dealing with directional or radius queries. They also use R-tree to index geo-referenced video with a two-level structure, *i.e.*, the first level is a basic R-tree structure while the second level is a probability map attached to each leaf node of the R-tree.

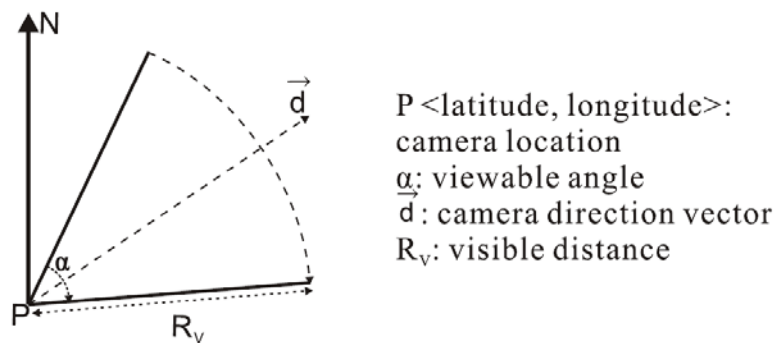


Fig. 1. Illustration of the field-of-view (FOV) model in 2D space.

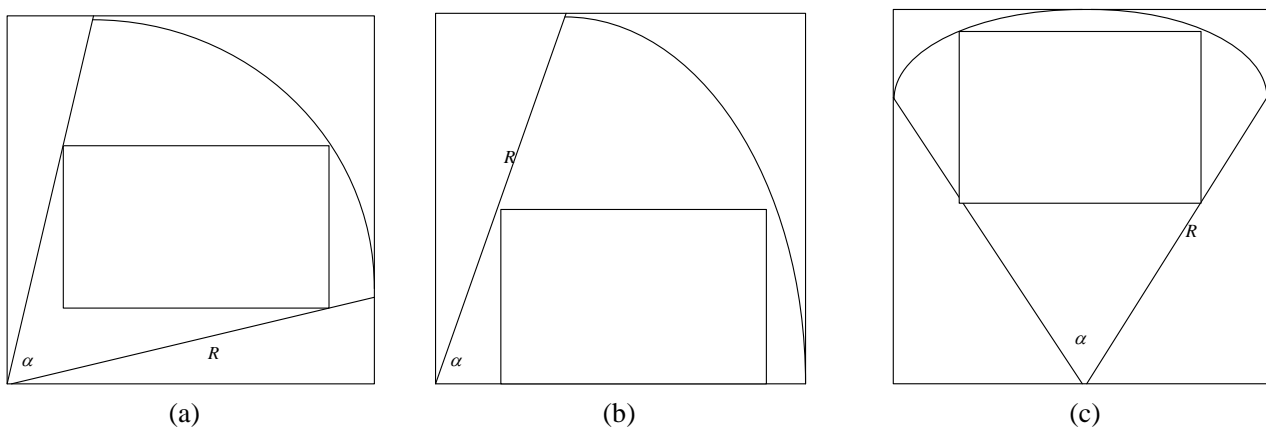


Fig. 2. Illustration of the inner and outer bounding rectangle of the DBR-tree.

Design of the Double-Bounding R-tree

The R-tree (including R-tree [9], R⁺-tree [10], R^{*}-tree [11] and their variations) is the most widely used multidimensional spatial data indexing structure. To reduce the calculation complexity, R-tree family uses minimum bounding rectangle (MBR for short) to partition the metric space. When dealing with the camera's maximum viewing distance, this large number always makes the searching regions locating within the FOV. To avoid unnecessary searching, we design the DBR-tree and the structure of the DBR-tree is shown in Fig. 2. Fig. 2(a) shows a general case of the inner and outer bounding rectangle. Fig. 2(b) and 2(c) show two special cases respectively. The outer bounding rectangle is with the same idea as typical R-tree, which the inner bounding rectangle is connected either on the line or inside the FOV. In this case, a two phase pruning algorithm is used: FOVs with the inner rectangle overlapping with the query region can be directly treated as queried results, while those with the outer rectangle not overlapping with the query region can be pruned at the first step. This means: if the query region overlaps with the inner bounding rectangle, the FOV can definitely capture the query region; if the query region does not overlap with the outer bounding rectangle, the FOV cannot capture the query region. Only the rest FOVs need calculating overlap with a refinement step. The refinement step is calculated according to the real overlap region.

Experimental results

We performed our experiments on one large synthetic dataset. Due to the difficulties of collecting a large set of real-world videos, a synthetic dataset of moving cameras with positions inside a 75 km × 75 km region is used to test the performance with large-scale of data. We generated camera trajectories using the Georeferenced Synthetic Meta-data Generator [12]. The produced synthetic meta-data exhibits equivalent characteristics to real-world data. We selected 100 randomly distributed center points within the test area and generated 5, 500 moving cameras with trajectories near these center points. Each camera was traced for 1, 000 seconds, with a sampling rate of 1 per second for the GPS and compass. Thus, the resulting dataset contained about 5.4 million FOVs. To simulate a real-world case, we set the maximum speed of the moving cameras to 60 km/h, with an average speed of 20 km/h. We also set the maximum camera rotation limit to 30° per second, which ensures that the camera rotates smoothly and does not jump from one direction to another, emulating real user behavior. The experiments were then conducted on a server with two quad core Intel[®] Xeon[®] X5450 3.0 GHz CPUs and 32 GB memory running Linux 2.6.18. We then generated 10, 000 range queries of 500 m * 500 m rectangles within the 75 km × 75 km test region and counted the cumulative processing time. Fig. 3 shows the comparison between the typical R-tree and the DBR-tree. Overall, the DBR-tree can achieve around 30% speed up than typical R-tree. As the dataset size grows, the processing time of both structures increases, but increasing rate of DBR-tree is smaller than that of typical R-tree.

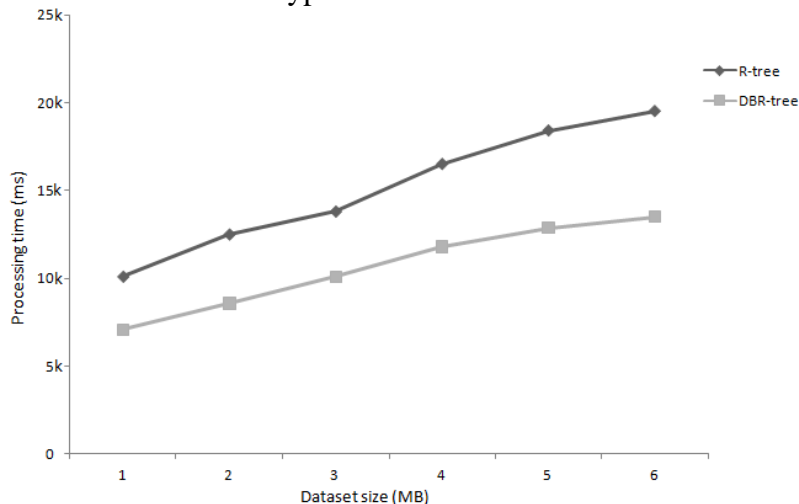


Fig 3. Comparison of processing time between typical R-tree and DBR-tree.

Conclusion

In this manuscript, we investigated the problem of indexing the FOV using sensor meta-data. The DBR-tree is designed with both inner bounding rectangle and outer bounding rectangle. Both the inner and outer rectangle can prune the unnecessary computation. Experimental results show that the DBR-tree can achieve around 30% improvement in processing time compared to typical R-tree.

Acknowledgement

In this paper, the research was sponsored by the National Natural Science Foundation of China (Project No. 61501101 and Project No. 61472266), the Fundamental Research Funds for the Central Universities (N130319002), the general program of the education department of Liaoning province (L2014086) and the International cooperation projects of the technology division of Shenyang (F13-319-6-17).

References

- [1] S. Arslan Ay, R. Zimmermann, and S. H. Kim. Viewable Scene Modeling for Geospatial Video Search. In SIGMM, ACM International Conference on Multimedia, pages 309–318, 2008.
- [2] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket Location-Support System. In MobiCom, ACM International Conference on Mobile Computing and Networking, pages 32–43, 2000.
- [3] C. H. Graham, N. R. Bartlett, J. L. Brown, Y. Hsia, C. C. Mueller, and L. A. Riggs. Vision and visual perception. 1965.
- [4] N. O'Connor, T. Duffy, C. Gurrin, H. Lee, D. Sadlier, A. Smeaton, and K. Zhang. A Content-Based Retrieval System for UAV-like Video and Associated Metadata. In Airborne Intelligence, Surveillance, Reconnaissance Systems and Applications, 2008.
- [5] Z. Zhu, E. Riseman, A. Hanson, and H. Schultz. An Efficient Method for Geo-referenced Video Mosaicing for Environmental Monitoring. In Machine Vision and Applications, volume 16, pages 203–216. Springer, 2005.
- [6] X. Liu, M. Corner, and P. Shenoy. SEVA: Sensor-Enhanced Video Annotation. In SIGMM, ACM International Conference on Multimedia, pages 618–627, 2005.
- [7] Ma, H., Ay, S. A., Zimmermann, R., & Kim, S. H. Large-scale geo-tagged video indexing and queries. *Geoinformatica*, 18(4), 1-27, 2014.
- [8] Ma, H., Ay, S. A., Zimmermann, R., & Kim, S. H. A grid-based index and queries for large-scale geo-tagged video collections. *Database Systems for Advanced Applications*, 7240, 216-228, 2012.
- [9] Antonin Guttman. R-Trees: A Dynamic Index Structure for Spatial Searching. In SIGMOD, ACM International Conference on Management of Data, pages 47-57, 1984.
- [10] N. Roussopoulos, C. Faloutsos and S. Timos. The R⁺-tree: A Dynamic Index for Multi-dimensional Objects. In VLDB International Conference on Very Large Databases, pages 507-518, 1987.
- [11] N. Beckmann, H.P. Kriegel, R. Schneider and B. Seeger. The R*-tree: An Efficient and Robust Access Method for Points and Rectangles. In SIGMOD, ACM International Conference on Management of Data, pages 322-331, 1990.

[12] S. Arslan Ay, R. Zimmermann, and S. H. Kim. Generating Synthetic Meta-data for Georeferenced Video Management. In ACM SIGSPATIAL GIS International Conference on Advances in Geographic Information Systems, pages 280–289, 2010.