Compressing Trajectories Using Inter-Frame Coding

Cheng-Yu Lin and Ling-Jyh Chen
Institute of Information Science, Academia Sinica, Taiwan

***Abstract*—In this study, we address the scalability issue in trajectory data management for emerging location-based applications. We propose a scheme called Inter-Frame Coding (IFC) for lossless compression of trajectory data. We evaluate the IFC scheme using real-world trajectory dataset, and the results show that IFC can achieve a compression ratio of 58%. The IFC scheme is simple, efficient, and lossless; thus, it has the potential to facilitate trajectory-based data storage, compression, and computation.**

I. Introduction

Location-aware services are rapidly permeating every part of our living environments. The difference between the new genre of applications and conventional ones is that they are driven by moving objects with the location information as a function of time [6]. A substantial amount of research effort has been invested in the areas of trajectory data management. However, the amount of data increases dramatically over time, leading to storage, transmission, and computation problems. Consequently, an effective trajectory data compression solution that can improve the scalability of moving object databases is highly desirable.

In this study, we propose a novel algorithm, called Inter-Frame Coding (IFC), for trajectory compression. The IFC scheme exploits the spatial and temporal localities between contiguous data points on a trajectory, and compresses data by reducing the amount of redundant information in the raw spatial-temporal data points. Unlike existing approaches [3–5], the proposed scheme is simple and *lossless*; and it has the potential to facilitate large-scale storage, compression, and computation of trajectory data.

Using trajectory datasets collected by the two real-world systems, namely the Taipei eBus system [2] and the Microsoft GeoLife Project [1], we evaluate the proposed scheme in terms of the data compression ratio. The results show that the IFC scheme achieves a high compression rate in large-scale databases. In our evaluations, the IFC scheme achieved compression ratios of 49% and 58% on the TPE eBus and Microsoft GeoLife datasets respectively. Moreover, IFC is simple and ready for immediate real-world deployment. In addition, it can be implemented easily in conjunction with unequal erasure protection and data prioritization schemes for operations in lossy or resource-constrained environments.

II. Inter-Frame Coding

The rationale for this study is based on the observation that *spatial and temporal localities are common in a trajectory*. We propose a novel trajectory compression algorithm, called *Inter-Frame Coding* (IFC), to exploit the spatial and temporal localities of contiguous spatial-temporal data points, thereby reducing the amount of redundant information in the raw trajectory data.

There are two types of data points in the IFC scheme: I frames, which contain the *index* data points of a trajectory; and O frames, which contain the *offsets* of the subsequent data points that correspond to the I frames. Let $\mathcal{I}_i^u$ denote the i-th I frame that represents the v-th spatial-temporal data point of the u-th trajectory (i.e., $\mathcal{T}_v^u$); and let $\mathcal{O}_{i,j}^u$ denote the j-th O frame associated with $\mathcal{I}_i^u$, i.e., the offset of $\mathcal{T}_{v+j}^u$ to $\mathcal{T}_v^u$.

Specifically, $\mathcal{I}_i^u = (sn, u, lng_i^u, lat_i^u, t_i^u)$, where sn is the sequence number of $\mathcal{I}_i^u$; u is the trajectory identifier; and lng_i, lat_i, and t_i are the longitude, latitude, and timestamp of $\mathcal{T}_v^u$ respectively. Meanwhile, $\mathcal{O}_{i,j}^u = (i\_sn, lngOff_{i,j}^u, latOff_{i,j}^u, tOff_{i,j}^u)$, where i_sn is the sequence number of the I frame that $\mathcal{O}_{i,j}^u$ is associated with (i.e., $i\_sn = \mathcal{I}_i^u.sn$); and $lngOff_{i,j}^u$, $latOff_{i,j}^u$, and $tOff_{i,j}^u$ represent, respectively, the longitude, latitude, and time offsets of $\mathcal{T}_{v+j}^u$ to $\mathcal{T}_v^u$.

Generally, an I frame is associated with n O frames. The value of n is a system parameter tunable based on several factors, such as the sampling rate of the trajectory data, the speed of the moving object, and the data compression ratio required for the application (which we discuss in detail in the next subsection). However, when the offset values exceed the range allowed in an O frame, a new I frame must be created, even though the number of O frames associated with the former I frame is less than n.

III. Evaluation

We evaluate the IFC scheme, in terms of the data compression ratio, using two real-world trajectory dataset as outlined in Table I, on the open-source PostgreSQL database (version 8.4.4) and the PostGIS spatial database extension (version 1.5.1). For I frame data in the IFC scheme, we store the location information (lng and lat) using the *Point* data type (16 bytes), the time information (t) using the *Timestamp* data type (8 bytes), and the sequence number (sn) and the trajectory identifier (u) using the *Integer* data type (4 bytes). For O frames, we store the I frame sequence number (i_sn) using the *Integer* data type; and the longitude, latitude, and time offsets $lngOff$, $latOff$, and $tOff$) all use the *Short Integer* data type (2 bytes).

Figure 1 compares the theoretical compression ratio and the ratio achieved in the two scenarios with different n values under the IFC scheme. We observe that the GeoLife curve

This research was supported in part by the National Science Council of Taiwan under Grants: NSC 99-2631-S-003-002 and NSC 99-2219-E-001-001.

TABLE I
THE BASIC PROPERTIES OF THE TWO DATASETS USED IN THIS STUDY

	TPE	GeoLife
Data source	Taipei e-bus System	GeoLife Project, Microsoft Research Asia
Duration	22 days (2010/04/01 - 2010/04/23)	1,129 days (2007/04/13 - 2010/05/15)
Coverage	greater Taipei area, Taiwan	mostly in Beijing, China
# of distinct moving objects	4,028	104
trajectory types	bus	bus, car, bike, and walk
avg # of trajectories per day	3,865	12
avg # of data points per day	3,235,460	24,020
data resolution	about 1 minute	about 5 seconds

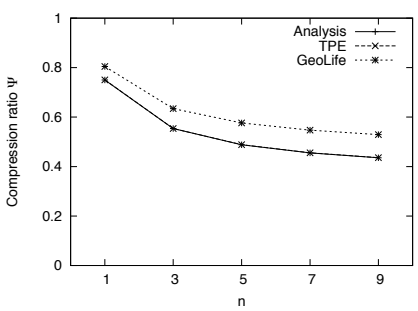

Fig. 1. Comparison of the theoretical compression ratio and the compression ratio achieved using the two datasets under IFC with different n values.

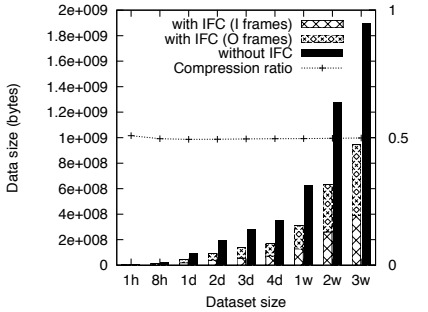

Fig. 2. Comparison of the size of the TPE dataset (for different time periods) with and without the IFC scheme ($n = 5$).

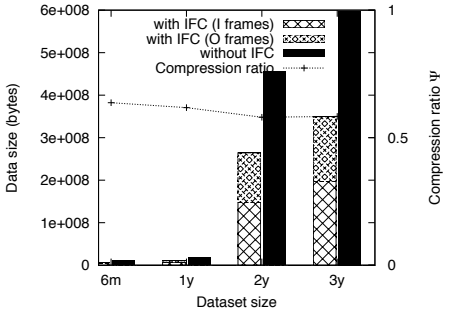

Fig. 3. Comparison of the size of the GeoLife dataset (for different time periods) with and without the IFC scheme ($n = 5$).

is consistently higher than the other two curves, and the TPE curve and the theoretical compression ratio curve are overlapped completely. The reason is that the GeoLife dataset is comprised of a large number of short trajectories contributed by volunteers on a daily basis, while the TPE dataset contains a set of uninterrupted trajectories contributed by the Taipei e-buses on a day-and-night basis. Consequently, the GeoLife dataset has a higher percentage of I frames that have less than n O frames due to the truncation of the trajectory data collection, resulting in a higher compression ratio than the theoretical value. By contrast, in the TPE dataset, most of the I frames have complete n frames, and only a limited number of I frames have less than n O frames due to the truncation of the data collection. Therefore, the TPE dataset can achieve a comparable compression ratio to the theoretical ratio.

Figures 2 and 3 compare the storage required, in terms of the byte size and the compression ratio, before and after applying the IFC scheme ($n = 5$) with different lengths of the TPE and GeoLife datasets. We observe that the proposed scheme can achieve a compression ratio of approximately 49.5% in all test cases, which represents a 50.5% storage saving. The compression ratio is slightly higher than the theoretical value (i.e., 48.81%) because it is inevitable that some I frames have fewer than n O frames in the compressed dataset. For instance, the number of data points in a trajectory may not be a multiple of $(n+1)$, so there will be fewer O frames for the last I frame. Moreover, most trajectories are related to urban areas, which have poor GPS reception due to buildings and other obstacles; therefore, an I frame may not have exactly n O frames if one data point floats away from the maximum offset value allowed for an O frame. Since the I frame is larger than the O frame, the compression ratio becomes larger if some I frames do not have n O frames in the dataset.

IV. CONCLUSION

In this paper, we propose the Inter-Frame Coding algorithm (IFC) for lossless compression of trajectory data. Using realistic datasets compiled by two real-world systems, we evaluated the proposed IFC scheme and verified that it can achieve a compression ratio of 58%. The scheme is simple, lossless, efficient, and extensible with advance features (e.g., unequal erasure protection and data prioritization). Thus, we believe that it could facilitate the development of trajectory databases and future location-aware services.

REFERENCES

[1] GeoLife: Building social networks using human location history. http://research.microsoft.com/en-us/projects/geolife/.

[2] Taipei e-bus system. http://www.e-bus.taipei.gov.tw/index.htm.

[3] J. Gudmundsson, J. Katajainen, D. Merrick, C. Ong, and T. Wolle. Compressing spatio-temporal trajectories. In *ICAC*, 2007.

[4] N. Meratnia and R. A. de By. Spatiotemporal Compression Techniques for Moving Point Objects. In *EDBT*, 2004.

[5] F. Schmid, K.-F. Richter, and P. Laube. Semantic Trajectory Compression. In *ISASTD*, 2009.

[6] A. P. Sistla, O. Wolfson, S. Chamberlain, and S. Dao. Modeling and Querying Moving Objects. In *IEEE ICDE*, 1997.