USING LIFELINES FOR SPATIOTEMPORAL SUMMARIES

Anthony Stefanidis, Panayotis Partsinevelos, Peggy Agouris

Department of Spatial Information Engineering
National center for Geographic Information and Analysis
University of Maine, 348 Boardman Hall
Orono, ME 04469-5711, USA
{tony, panos, peggy}@spatial.maine.edu
http://www.spatial.maine.edu/~peggy/dgi.html

Abstract

In this paper we present a general framework for the analysis of spatiotemporal datasets. Our approach makes use of an analysis of spatiotemporal trajectories describing the path in space and time of objects in our dataset. We have developed a variation of self-organizing maps to generalize these object lifelines by permitting variable resolutions to represent complex segments compared to smoother ones. This allows us to better represent the complex spatiotemporal behavior of objects and to identify critical instances in their lifelines when certain aspects of their behavior change. The resulting generalized representation is the key concept that supports the generation of spatiotemporal summaries. The effectiveness of our hybrid SOM technique for spatiotemporal generalization is demonstrated with early experimental results.

1. Introduction

Geospatial applications tend to become increasingly spatiotemporal. The information that analysts seek often resides not in a single image or a map but in a multitemporal collection of images and maps instead. Various parallel advancements in sensor technology have allowed us to collect easily valuable multitemporal geospatial information. Reliable satellite, aerial and ground imagery can be readily available on demand or periodically. Advancements in video sensors are also providing us with the opportunity to capture dynamic events captured in video datasets. Even beyond imagery, the availability of cheap and compact positioning devices allows us to capture object trajectories, creating a novel type of spatiotemporal information. The processing and analysis of spatiotemporal data collections is introducing some interesting challenges, mostly associated with the size and complexity of the information space that has to be explored. It also introduces interesting challenges on metadata, as existing approaches are rather static in nature. This paper addresses the use of spatiotemporal lifelines to summarize the information content of multitemporal datasets. This summarized information can then be used to create multimedia summaries of dataset collections, serving as novel multimedia metadata for them. The paper deals primarily with monitoring applications, where few objects move and interact throughout a relatively stable background scene.

Interesting work on spatiotemporal summarization has been performed on video analysis. Work relevant to our topic includes the use of image templates, statistical features, and histogram-based retrieval and processing techniques [Chang et al., 1998]. Video summarization is the objective of work on the generation of a "skim" video to represent a synopsis of the original video taking into consideration both visual and speech properties [Smith & Kanade, 1995]. Video posters have also been proposed as alternatives to describe a video story content [Yeung & Yeo, 1997]. Work on topics usefule for summarization includes the approaches to identify different scenes within a video stream by analyzing a variety of properties [Vasconcelos & Lippman, 1998] and the work to extract and recognize moving

objects, and classify their motion [Medioni et al., 1998; Rosales & Sclaroff, 1999]. In addition, the generation of spatiotemporal synthetic datasets to simulate movement trajectories is presented in [Pfoser & Theodoridis, 2000]. In the database domain, we have the work temporal zooming of [Hornsby & Egenhofer, 2000].

In this paper we present a general framework for the analysis of spatiotemporal dataset. For reference we assume the use of video data as input for our work. We draw our motivation from monitoring applications, where objects move and interact throughout a relatively stable background scene. We present a framework to analyze such datasets in order to produce brief summaries of their content. Our approach makes use of an analysis of spatiotemporal lifelines to describe the path in space and time of objects in our dataset. A lifeline can be broadly defined as a sequence of the spatial locations $(x,y,z)_i$ of an object over a time interval (t_1,t_2) during which an object has moved from one location to another. In the case of a video dataset this interval is usually a video shot, and the sequence of spatial locations describes the object's location at every frame (or at fixed time intervals, e.g. every second). Alternatively the time interval of an object's lifeline may be a subset of a longer video shot, as t_1 and/or t_2 may be the time the object entered and/or exited the video camera field-of-view. Individual lifelines are analyzed to identify critical points, denoting instances where the object's behavior changes (e.g. accelerates, or makes an abrupt turn). These instances denote events in an object's trajectory, and are important information for subsequent analysis. We can move from single to multiple trajectories by spatial relations (e.g. topology, orientation, distance) and this allows us to identify groups of objects, or behavioral similarities in general. It should be noted that at this stage, our analysis focuses on the spatial properties of an object (e.g. location, shape), and we are not addressing thematic information (e.g. use, ownership).

While we focus on the analysis of video datasets, the framework can be generalized to function on any type of spatiotemporal datasets. Indeed, a video sequence could be substituted by a large collection of GIS layers representing different instances of the same area over many years, or by a collection of imagery gathered by a satellite as it passes over the same region in its lifetime, or even by the positional information of GPS-enabled agents roaming in a field. Similarly, an object in a video sequence could be substituted by an evolving geospatial entity (e.g. a crop field or a building complex).

This paper is organized as follows. In section 2 we present a general overview of our summarization framework. In section 3 we present our novel approach for single object generalization, while in section 4 we discuss issues related to the grouping of multiple objects. We conclude with early experimental results in section 5.

2. Proposed Framework

Our approach proceeds by analyzing the spatiotemporal trajectories of moving objects. The proposed framework is shown in figure 1. In the video dataset we assume a classification or coarse tracking that has identified in each frame (or in select frames) the approximate outline of mobile objects. This information may be corrupted by various types errors like occlusions, noise, and misclassified pixels. A spatiotemporal trajectory is produced by linking all this information and includes inherently all pertinent information for the behavior of a moving object over the corresponding time interval. However, some of this information is redundant and should be truncated for improved analysis, storage, and communication, while some of this information is significant and should be identified and emphasized. In order to capture the significant portion of this immense data flow, we require the efficient and accurate spatiotemporal generalization of these trajectories.

Trajectories can be perceived as paths in the spatiotemporal (ST) space. Therefore, we have extended a methodology originally intended to perform road extraction from satellite imagery to capture and generalize spatiotemporal trajectories. Our road extraction approach was based on the use of self-organizing maps (SOM, [Kohonen 1982]) to extract road centerlines [Doucette et al., 2001]. The SOM

technique not only links pixels into road centerlines, but also distributes nodes along the road to generalize the extracted centerline.

Common SOM solutions present some shortcomings, most notable of which are the high variability of the technique in various iterations, and the inability to densify the number of nodes in highly complex areas. To overcome this problem we introduce in this paper a hybrid approach, where SOM is further enhanced by a geometric analysis. This allows us to identify high variation areas (e.g. moments of rapid turns, or abrupt acceleration/deceleration in a video) and densify selectively the number of nodes over these intervals. Similarly, we can also identify areas of low variations (e.g. intervals of steady movement) and selectively thin the number of nodes over these intervals. This geometrically enhanced SOM solution allows us to extract generalized lifelines from the dense spatiotemporal trajectories of single objects. Multi-object trajectory analysis introduces relational considerations for the selection of additional nodes. Examples of important spatiotemporal object relations include topology, proximity, and user defined restrictions. An analysis of these properties can provide us with quantitative metrics to define qualitative concepts like 'group', 'convoy', or 'expected behavior', which are essential for summarizing spatiotemporal datasets. Such multi-object information provides us with additional input when identifying important instances in our datasets, and allows us to move to higher levels of abstraction.

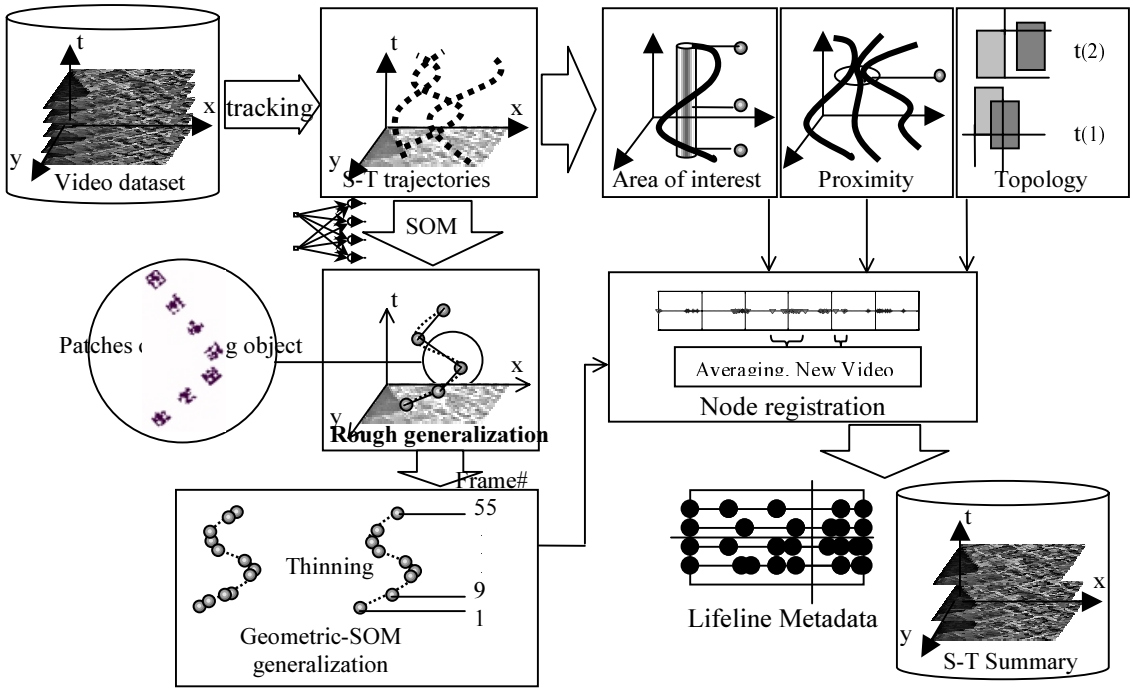


Figure 1: Outline of proposed framework

For video segmentation, storage, and retrieval applications, the most important aspect of our hybrid generalization approach is that it accommodates spatiotemporal zooming. This zooming is based on a dynamic, content-defined scale rather than a stable temporal increment-based scale. In other words, we can use more frames, closer to each other to represent abrupt motions, compared to fewer frames to represent smooth, regular movements. In addition, the resulting lifelines can be serve as object metadata

files that facilitate retrieving, querying and similarity matching. They dismantle the behavior of a moving object in its primary attributes, and thus enable behavior-based queries.

3. Single Object Analysis

The self-organizing map (SOM) algorithm is a nonlinear and nonparametric regression solution to a class of vector quantization problems. It belongs to a distinct class of artificial neural networks (ANN) characterized by unsupervised and competitive learning. It proceeds by distributing nodes to describe the spatial distribution of the input space. We use it as the basis of our spatiotemporal generalization.

Standard SOM algorithms tend to function well when linearizing road segments in aerial imagery, as these segments are reasonably smooth lines. However, spatiotemporal trajectories tend to include abrupt variations (e.g. an object may change its velocity and orientation very often in a limited area). This makes the use of standard SOM techniques inadequate for spatiotemporal generalization. To overcome this problem we have developed a hybrid geometry-enhanced technique. Compared to a standard SOM process, our hybrid trajectory analysis approach offers the advantages of:

- invariance to the selection of the initial number of nodes, and the
- ability to selectively densify or thin our node distribution, to better capture the complexity of content of the processed dataset.

This allows us to use multiple resolutions when generalizing a single lifeline. This is equivalent to selecting various scale factors based upon the complexity of the information included in the dataset. Complexity is defined in our context as the variation of an object's spatiotemporal location. In order to detect and quantify this variation we use as key metric for our analysis the 3-dimensional angle formed between three subsequent nodes in the 3-dimensional spatiotemporal space (x, y, t). Each pair of points describes a 'state' of spatiotemporal behavior. The angle between two consecutive states is indicative of the local spatiotemporal variation. High deviations of these angles from 180^o indicate extreme variations between subsequent states, and require more, densely distributed generalization nodes.

Our algorithm identifies locations where densification is needed and evaluates the number of additional representing nodes. Then, new, localized SOM solutions are performed, and thinning takes place in order to remove nodes that do not contribute to generalization. In this manner, the degree of generalization is based on the value of the 3-dimensional angle, and node repositioning and densification process takes place using this spatiotemporal angle information. A detailed description of the algorithm may be found in [Stefanidis et al., 2001]. The product of this procedure is a generalized spatiotemporal lifetime of the initial dataset. Further spatiotemporal zooming can take place by selecting less or more nodes to describe the movement attributes. The variables upon which densification is determined can be user defined and quantify the volume of the spatiotemporal zooming.

An error estimation is defined by measuring the distance between the actual and generalized line using both SOM and hybrid geometry-SOM generalization products. (fig.2)

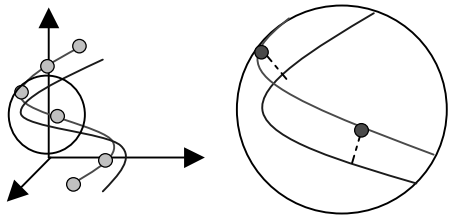


Figure 2. Lifeline distances in a 3-D ST space.

A measure of the spatiotemporal distance d_{ST} between a trajectory i and the generalized node poly-line j is provided by:

$$d_{ST} = \sqrt{\frac{\sum d_n^{ij}}{n}} \quad (1)$$

where n is the number of SOM nodes along trajectory j, and d^{ij} is the distance between the trajectory and the generalization nodes in the ST space (Fig. 3). Using both generalization techniques it is obvious as demonstrated in the experimental section that the proposed hybrid technique is significantly closer to the original dataset than the SOM one.

4. Multiple Object Analysis

The consideration of multiple objects brings forward the need to address two issues. First, we have to select specific time instances for our video summaries using independent nodes from multiple trajectories. Second, we have to consider the introduction of additional nodes when taking into account the relations of two or more trajectories. One can easily understand that the temporal coordinates of nodes describing the path of object i, and these describing the path of another object j, may be totally disjoint. The comparative analysis of these two lifelines may yield instances of interest, for example when they approach each other. Therefore, we introduce nodes termed "proximity nodes" when the distance between two or more objects drops below certain predefined threshold. In addition, the user may define areas of interest within which we can increase the monitoring resolution.

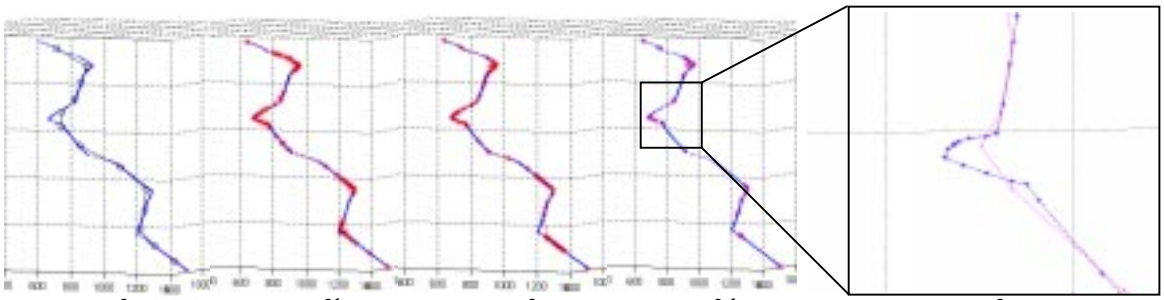
Topologic relations also define important information that is captured through additional nodes. For a temporal instance the elements included in a relational comparison between objects are direction and distance. When we refer to a temporal increment, differences in direction and distance are of interest. In addition, these attributes also refer to one object evolving in time. In this case cardinal direction or absolute distances do not contribute to the significance of the change. Their differences play the most important role. In order to capture the instantaneous relation between two objects we use one of the known topology-cardinality models [Egenhofer & Franzosa, 1991]. These relations are quantified through topologic models but they are beyond the scope of this paper.

In order to integrate and merge all the selected nodes from multiple trajectories, we introduce co-registration checks to handle overcrowded node areas. When grouping multiple trajectories, nodes separated by less than a minimum interval dt are merged and substituted by a single average node (averaging process). However, when adjacent nodes are describing long segments the summary is locally approximated by a brief near-video segment. Otherwise, it includes distinct sparse instances. Combined, these sparse instances and brief video segments produce a new representative and concise video summary, where temporal resolution changes dynamically, to best capture variations of spatiotemporal attributes.

5. Experiments

The approach described in this paper has been implemented in the MATLAB environment. We created synthetic datasets of moving objects to use in our experiments. We also generated random behaviors of spatiotemporal trajectories to describe objects moving with stable or variable velocity. The following figures show some experiments performed in order to exhibit the presented techniques.

In figure 3 we present a rough generalization of a spatiotemporal trajectory (a), the densification of the spatiotemporal trajectory with different generalization resolution (b,c), the product of thinning (d), and a detail of the generalization (e). The polygonic line in figure 3e demonstrates the standard SOM solution. We can easily see how the nodes from our hybrid approach describe the trajectory much better than the standard solution.



3a. 3b. 3c. 3d. 3e.

Figure 3: a) SOM generalization, b,c) proposed generalization with different densification parameters, d) thinning product, e) generalization detail.

In figure 4 the SOM generalization is described by "star" nodes while the hybrid generalization is depicted with triangular nodes. The RMS error as defined in section 3 is 89 pixels for the SOM result and only 9 pixels for the result of our hybrid technique.

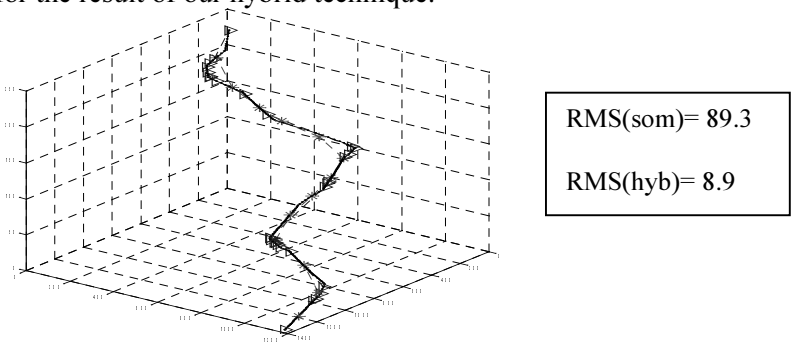
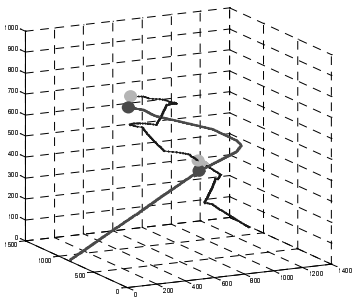


Figure 4: Precision comparison between SOM and hybrid techniques.

In figure 5 we demonstrate the selection of proximity nodes. We generated two trajectories traveling through roads in the scene shown in figure 5a. The proximity nodes are identified and shown as circles along the corresponding trajectories in figure 5b.

a.



b.

Figure 5: Moving datasets on image and proximity nodes.

Finally, we show the co-registration of two trajectories in figure 6. The nodes for these two trajectories are denoted by stars and triangles on the left-hand side. The structure of the corresponding summary is shown in figure 6 with elongated segments indicating short videos and small squares and circles indicating sparse instances.

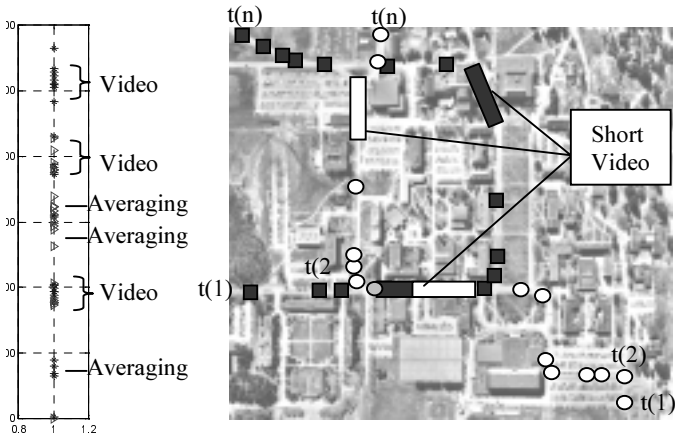


Figure 6: Registration and representation of product summary.

The above figures are early experimental results and indicate the performance of our proposed approach. Its main advantage is the potential for full automation, and the ability for objective yet meaningful analysis of spatiotemporal datasets to support their storing, browsing, and querying. These are issues that we will be addressing in the near future.

Acknowledgments

This work was supported by the National Science Foundation through Digital Government award number DGI-9983445.

References

[1] Chang W., G. Sheikholeslami, J. Wang, & A. Zhang, 1998. Data resource selection in distributed visual information systems, IEEE Transactions on Knowledge and Data Engineering, 10(6).

[2] Smith M. & T. Kanade, 1995. Video Skimming for Quick Browsing based on Audio and Image Characterization, Tech. Report CMU-CS-95-186, Carnegie Mellon University.

[3] Yeung M., & Boon-Lock Yeo, 1997. Video Visualization for Compact Presentation and Fast Browsing of Pictorial Content, IEEE Trans. on CSVT, 7(5), pp. 771-785.

[4] Vasconcelos N., & A. Lippman 1998. A Spatiotemporal Motion Model for Video Summarization, CVPR'98, Santa Barbara.

[5] Medioni G., R. Nevatia, & I. Cohen, 1998. Event Detection and Analysis from Video Streams, DARPA Image Understanding '98, pp. 63-72.

[6] Rosales R. & S. Sclaroff, 1999. 3D Trajectory for Tracking Multiple Objects and Trajectory Guided Recognition of Actions, CVPR '99.

[7] Pfoser D. & Y. Theodoridis, 2000. Generating Semantics-Based Trajectories of Moving Objects, Intern. Workshop on Emerging Technologies for Geo-Based Applications, Birkhaeuser.

[8] Hornsby K. & M. Egenhofer, 2000. Shifts in Detail through Temporal Zooming, Proceedings of the 10th International Workshop on Database & Expert Systems Applications.

[9] Kohonen T., 1982. Self-Organized Formation of Topologically Correct Feature Maps, Biological Cybernetics, pp. 59-69.

[10] Doucette P., P. Agouris, A. Stefanidis, & M. Musavi, 2001. Self-Organized Clustering for Road Extraction in Classified Imagery, ISPRS Journal of Photogrammetry and Remote Sensing, 55(5-6), pp. 347-358.

[11] Stefanidis A., P. Partsinevelos, P. Agouris & P. Doucette, 2000. Summarizing Video Datasets in the Spatiotemporal Domain, Proceedings DEXA2000, IEEE Press, Greenwich, pp. 906-912.

[12] Stefanidis A., P. Partsinevelos & P. Agouris, 2001. Automated Spatiotemporal Scaling for Video Generalization,. IEEE International Conference on Image Processing, ICIP'01 (in press).

[13] Egenhofer M. & R. Franzosa, 1991. Point-Set Topological Spatial Relations, Int. Journal of Geographical Information Systems, 5(2): 161-174.