DADOS DO PROJETO

Sumário (Português)

Os avanços tecnológicos recentes têm permitido recolher volumes de dados sobre a evolução de fenómenos espaciotemporais bastante superiores à capacidade existente para os analisar e extrair informação relevante em diversas áreas científicas. Por isso, são cada vez mais necessárias ferramentas capazes de automatizar processos de análise quantitativa de dados espácio-temporais, garantindo níveis de objetividade, precisão e reprodutibilidade compatíveis com a realização de trabalho científico. Atualmente, já existem ferramentas bem conhecidas para o processamento de dados espaciais estáticos (p. ex., os Sistemas de Informação Geográfica), mas o suporte à modelação de fenómenos dinâmicos é limitado, sendo muitas vezes necessário realizar um grande esforço na programação de algoritmos complexos e que são específicos a um determinado problema.

Este projeto centra-se no desenvolvimento de ferramentas avançadas para modelação e análise de dados espácio-temporais, usando modelos de representação contínuos no espaço e no tempo. O elemento chave será um sistema de gestão de dados capaz de modelar transformações espaciais genéricas (p. ex., mudança de forma, mudança de tamanho, translação, rotação, agregação ou fracionamento de entidades ou objetos) representando os fenómenos de interesse ao longo do tempo. Este sistema será acessível através de uma linguagem de interrogação disponibilizando funções para a gestão, consulta e processamento de grandes volumes de dados. Também serão desenvolvidos métodos para criar representações espácio-temporais a partir de sequências de imagens ou vídeos, e ferramentas de visualização de dados e interação com o utilizador. Será disponibilizado um conjunto integrado de ferramentas para simplificar a realização de estudos sobre fenómenos de natureza espácio-temporal. O objetivo é reduzir o tempo e o esforço que hoje em dia é necessário dedicar ao desenvolvimento de procedimentos complexos de gestão e processamento de dados, libertando assim recursos para a realização dos estudos propriamente ditos.

A prova de conceito baseia-se em dois casos de estudo envolvendo a modelação de fenómenos espácio-temporais com características distintas. O primeiro consiste na modelação da propagação de fogos florestais a partir de imagens aéreas, tendo em vista a realização de estudos sobre as emissões de carbono para a atmosfera. O segundo consiste na criação de uma base de dados caracterizando as alterações morfológicas que as células sofrem quando se movem no seu próprio ambiente. A quantificação destas características tem importância em processos biológicos como o desenvolvimento embrionário ou a formação de tumores. A origem dos dados são vídeos microscópicos. No futuro, pretende-se ainda que os resultados deste projeto possam ser aplicados noutras áreas, p. ex., em estudos sobre erosão costeira, assoreamento de rios ou outras.
As instituições participantes são a Universidade de Aveiro e o INESC-TEC Porto.

Sumário (Inglês)

Recent technological advances made it possible to collect data on the evolution of spatiotemporal phenomena far superior to the existing capacity to analyze and to extract relevant information from them in various scientific areas. Thus, there is a growing need of tools to automate processes of quantitative analysis of spatiotemporal data, guaranteeing levels of objectivity, precision and reproducibility compatible with the execution of scientific work. Nowadays, there are already well-known tools for the processing of static spatial data (e.g., Geographic Information Systems), but the support for the modelling of dynamic phenomena is limited, and it is often necessary to make a great effort in implementing complex algorithms that are specific to solve a single problem.

This project focuses on the development of advanced tools for modeling and analyzing spatiotemporal data using continuous models of space and time. The key element is a data management system designed to model generic spatial transformations (e.g., shape?s transformation, movement, rotation, aggregation or fragmentation of entities or objects) representing the evolution of the phenomena across time. The system will be accessible through a query language offering high-level functions for the management, retrieving and processing of large volumes of data. We will also develop methods to create spatiotemporal representations from sequences of images or videos, and spatiotemporal data visualization tools. The result will be an integrated system designed to perform studies about spatiotemporal phenomena, lessening the effort and the time required to implement complex data management and processing procedures, and thus releasing resources to accomplish the studies themselves.

The proof of concept is based on two case studies involving the modeling of spatiotemporal phenomena with highly distinct features. The first consists of the modeling forest fires propagation from aerial images, in order to carry out

studies on carbon emissions to the atmosphere. The second consists of the creation of a database characterizing the morphological changes of cells as they move, using data extracted from microscopic videos. The quantification of these characteristics is important in biological processes such as embryonic development or tumorigenesis. It is also expected that the results of this project will be applied in other fields in the future, such as, studies on coastal erosion, silting of rivers or others.

The participating institutions are the University of Aveiro and INESC-TEC Porto.

DADOS DO PROJETO

Sumário para publicação (Português)

Os avanços tecnológicos recentes têm permitido recolher volumes de dados sobre a evolução de fenómenos espaciotemporais bastante superiores à capacidade existente para os analisar e extrair informação relevante em diversas áreas científicas. Por isso, são cada vez mais necessárias ferramentas capazes de automatizar processos de análise quantitativa de dados espácio-temporais, garantindo níveis de objetividade, precisão e reprodutibilidade compatíveis com a realização de trabalho científico. Atualmente, já existem ferramentas bem conhecidas para o processamento de dados espaciais estáticos (p. ex., os Sistemas de Informação Geográfica), mas o suporte à modelação de fenómenos dinâmicos é limitado, sendo muitas vezes necessário realizar um grande esforço na programação de algoritmos complexos e que são específicos a um determinado problema.

Este projeto centra-se no desenvolvimento de ferramentas avançadas para modelação e análise de dados espácio-temporais, usando modelos de representação contínuos no espaço e no tempo. O elemento chave será um sistema de gestão de dados capaz de modelar transformações espaciais genéricas (p. ex., mudança de forma, mudança de tamanho, translação, rotação, agregação ou fracionamento de entidades ou objetos) representando os fenómenos de interesse ao longo do tempo. Este sistema será acessível através de uma linguagem de interrogação disponibilizando funções para a gestão, consulta e processamento de grandes volumes de dados. Também serão desenvolvidos métodos para criar representações espácio-temporais a partir de sequências de imagens ou vídeos, e ferramentas de visualização de dados e interação com o utilizador. Será disponibilizado um conjunto integrado de ferramentas para simplificar a realização de estudos sobre fenómenos de natureza espácio-temporal. O objetivo é reduzir o tempo e o esforço que hoje em dia é necessário dedicar ao desenvolvimento de procedimentos complexos de gestão e processamento de dados, libertando assim recursos para a realização dos estudos propriamente ditos.

A prova de conceito baseia-se em dois casos de estudo envolvendo a modelação de fenómenos espácio-temporais com características distintas. O primeiro consiste na modelação da propagação de fogos florestais a partir de imagens aéreas, tendo em vista a realização de estudos sobre as emissões de carbono para a atmosfera. O segundo consiste na criação de uma base de dados caracterizando as alterações morfológicas que as células sofrem quando se movem no seu próprio ambiente. A quantificação destas características tem importância em processos biológicos como o desenvolvimento embrionário ou a formação de tumores. A origem dos dados são vídeos microscópicos. No futuro, pretende-se ainda que os resultados deste projeto possam ser aplicados noutras áreas, p. ex., em estudos sobre erosão costeira, assoreamento de rios ou outras.
As instituições participantes são a Universidade de Aveiro e o INESC-TEC Porto.

Sumário para publicação (Inglês)

Recent technological advances made it possible to collect data on the evolution of spatiotemporal phenomena far superior to the existing capacity to analyze and to extract relevant information from them in various scientific areas. Thus, there is a growing need of tools to automate processes of quantitative analysis of spatiotemporal data, guaranteeing levels of objectivity, precision and reproducibility compatible with the execution of scientific work. Nowadays, there are already well-known tools for the processing of static spatial data (e.g., Geographic Information Systems), but the support for the modelling of dynamic phenomena is limited, and it is often necessary to make a great effort in implementing complex algorithms that are specific to solve a single problem.

This project focuses on the development of advanced tools for modeling and analyzing spatiotemporal data using continuous models of space and time. The key element is a data management system designed to model generic spatial transformations (e.g., shape?s transformation, movement, rotation, aggregation or fragmentation of entities or objects) representing the evolution of the phenomena across time. The system will be accessible through a query language offering high-level functions for the management, retrieving and processing of large volumes of data. We will also develop methods to create spatiotemporal representations from sequences of images or videos, and spatiotemporal data visualization tools. The result will be an integrated system designed to perform studies about spatiotemporal phenomena, lessening the effort and the time required to implement complex data management and processing procedures, and thus releasing resources to accomplish the studies themselves.

The proof of concept is based on two case studies involving the modeling of spatiotemporal phenomena with highly distinct features. The first consists of the modeling forest fires propagation from aerial images, in order to carry out

studies on carbon emissions to the atmosphere. The second consists of the creation of a database characterizing the morphological changes of cells as they move, using data extracted from microscopic videos. The quantification of these characteristics is important in biological processes such as embryonic development or tumorigenesis. It is also expected that the results of this project will be applied in other fields in the future, such as, studies on coastal erosion, silting of rivers or others.

The participating institutions are the University of Aveiro and INESC-TEC Porto.

Descrição Técnica / Revisão da Literatura

This project focuses on the modelling and processing of spatiotemporal (ST) data. While many exist proposals focusing on the management and processing of data about the location of the so-called moving objects (e.g., cars or ships) [MeCh10,SiClMa14], research on the modelling of phenomena with extent (e.g., propagation of forest fires or coastal erosion) across time is limited.

The representation modes to deal with ST data are raster (discrete) and vector (continuous), and the choice depends on the applications? requirements [KoSeFr03]. The former use data models based on the discretization of time, space or both. The general approach is to assign timestamps to a variety of formats such as feature classes (geometries) [Yu09,GePe14,JaSaDa15,HaAk16], mosaic datasets [CaKi06, MeCa12], and regular or irregular gridded spatial fields [ViLo07, BoYu10, ViPi14,ViVi16]. Discrete representations are easier to implement but the accuracy, storage requirements and performance are highly influenced by temporal and spatial resolution. Conversely, continuous representations allow for more compact data structures, as well as to model features of definable entities or events (e. g., the size and shape of a cell), which are implicit in discrete models.

The most successful ST data model using continuous models in time and space is based on Abstract Data Types (ADT) [GuBoEr00,LeFoGu03]. The main reason is that ADTs can be smoothly built into extensible DBMS, such as the Secondo prototype [GuBeDu10] or object-relational DBMS [ZhJiZh11,MaMoCa12,PeFrGi14]. ST data is represented as an ordered collection of motion units and each motion unit represents a ST phenomenon between two consecutive observations. Hence, methods are needed to model the phenomena spatial behavior between observations.
The first algorithm proposed to create the so-called moving regions from observations in ST databases is [ToGu01]. This continues to be the solution used in recent work [McWe10,McFr15,HeGu16] and the focus is on creating representations that are topologically valid at all times, while keeping compatibility with the underlying spatial DB. However, the methods using this algorithm cause deformation of the geometries estimated during the interpolation and the approximation errors are too big to be neglected in scientific work: when the moving regions rotate, the geometries tend to inflate at the middle point of the interpolations and the methods used to deal with concavities either do not perform well with noisy data [ToGu01,HeGu16,MoDiAm16] or make them approximately convex during interpolation [McWe10,McFr15]. Moreover, these solutions have been evaluated using only synthetic data and they are not being used in real studies.

Our strategy to deal with the problem of creating ST data from observations is to bring the knowledge about morphing techniques, which are widely used in animation software packages, into ST databases research. First, we have investigated on creating ST data from sequences of images [MoJe13,MoDiPa14] using different interpolation methods [MoDiMe14]. We were able to create ST data with small errors of approximation, but compatibility with current spatial DBMS was not achieved. So, we started looking for other morphing techniques and the best solution we found is to use methods based on triangulation [MoDiAm16]. Besides being able to create ST data with small errors of approximation, the decomposition of complex geometries into triangles helps on the simplification of the algorithms required to deal with ST data, as well as on detecting and solving of topological issues during interpolation. This is the starting point of this project proposal and the next steps consist of working on data models, query languages, algorithms, user interaction and visualization of continuous ST data.

The background of the members of this project covers the three key areas of this project: spatial and ST databases, 2D (and 3D) modeling, user interaction and data visualization.
José Moreira?s (Assist. Prof., UA/IEETA) main research area is on spatial and ST databases using continuous models of time and space. His first works focused on moving objects databases [MoRiSa99,MoRiAb00,MoRi03,MoRiSaSc08] and benchmarking [SaMo01]. Next, he has also investigated on discrete ST data models and query languages [FrCaMo09,MaMoCa12b]. He participated in research projects on ST data and he has been recently invited to teach a course on Spatial and ST databases at the University OTH Regensburg, Bavaria, Germany.
Paulo Dias (Assist. Prof., UA/IEETA) research includes 3D Reconstruction, Visualization and Human Machine Interaction. In recent years, he has participated in 5 national or international projects in related areas. He is also author or co-author of more than 60 papers.
Alexandre Carvalho (Assist. Prof., INESC TEC) research is related to ST data management and ST information visualization. In recent years, participated as Work Package Leader in FP7 LeanBigData, towards an innovative visualization system for database technology. At INESC-TEC, since 1998, he participated in more than 20 R&D projects.

Marco Oliveira (INESC TEC) is a Phd student in informatics engineering. His R&D interests are in complex information, ST information systems and interoperability. At INESCTEC, since 2000, he worked as researcher and project manager in several R&D, auditing and knowledge transfer projects.
Ana Isabel Miranda (Full Professor, UA/GEMAC) sub-director of the Department of Environment and Planning and coordinator of the Research Group Emissions, Modelling and Climate Change. Her research work includes over 650 publications on topics comprising air quality modelling and management, and health effects of air pollution. Her participation in this project will add great value given the extensive experience in research projects related to forest fires. She had participated in 20 European projects, being the national scientist responsible in 2 of them.

Descrição Técnica / Objetivos / Plano de Investigação e Métodos

The number and diversity of technologies to monitor and record the progress of real-world phenomena along time is increasing fast. Despite recent advances, the support provided by existing solutions to model, manage and analyze ST data is insufficient and do not cover the whole spectrum of potential applications. Indeed, it is often required to implement time-consuming and complex programs tailored to solve a specific problem from a particular domain, which cannot be easily applied to problems from the same or from other domains. In addition, these solutions often have limited functionality because dealing with ST data is complex and costly. Thus, the study of novel features to deal with large ST datasets using continuous models of time and space has the potential to enable easier implementation and boost the development of applications in several scientific domains.

Few database systems (prototypes or commercial) supporting time dependent and continuously changing data modeling are available for practical use today. Secondo is by far the most complete one [LeFoGu03,GuBeDu10]. This is a prototype developed for teaching and computer science research, but there are issues that need further investigation in order to used in scientific work. Indeed, Secondo has not yet been evaluated using real data (only synthetic data was used), and, despite recent advances , there are known issues regarding the precision (fidelity) of the ST data representations, particularly when the objects or the events being modelled rotate or in the presence of concavities [McWe10,McFr15,HeGu16,]. However, the ST data management system is a key component to make it possible to provide a complete infrastructure to execute tasks such as the acquisition, modelling, management and analysis of ST data.

The aim of this project is to develop an integrated set of tools for modeling, management and analysis of ST data represented using continuous models in space and time. The main objectives are:
1.To further investigate on the creation of ST data from observations in order to find a solution that minimizes deformation and creates a realistic transformations during interpolation and, at the same time, is compatible with existing spatial DBMS.
2.To develop generic methods and techniques that can be used in different application domains.
3.To show that it is possible to develop generic systems for modeling and processing ST data and achieve results compatible with scientific or industrial work.
4.To develop an integrated system to support the implementation of ST data analysis tasks, including extracting information from raw data, data management, visualization and information retrieval. The system must be modular, so that each tool can be easily replaced with new versions.

Our strategy is to model the ST behavior of real-world phenomena using morphing algorithms. Despite these techniques are widely used, particularly in animation or video editing software packages, their use to model ST phenomena in scientific applications is unusual or nonexistent. Particularly, we are interested on using methods based on compatible triangulation [BaBaAn08]. This choice is based on the results presented in [MoDiAm16], which show that it is possible to create more reliable representations of the phenomena than the in previous work. In addition, the use of triangle meshes allows to create compact representations of ST data. The algorithms to deal with ST will be easier to implement since they will operate on triangles rather than on complex geometries. In addition, the morphing of each triangle can be represented by a simple formula or a transformation matrix, and so, it will be possible to estimate the position or the shape representing a phenomena of interest at any given time using a single formula. The use of simple geometries such as triangles will also help to solve the issues related with the compatibility of the ST data management system with the underlying spatial DBMS.

The project is organized into seven Work Packages.
*WP1 (Spatiotemporal data preprocessing) focuses on the development of efficient methods to prepare raw data, e.g., sequences of images or video-frames, to be imported into a ST database. Image processing techniques will be used. Methods for outlier detection and the video-frame sampling (data reduction) will be investigated.

*WP2 (Spatiotemporal data management system) focuses on the development of a ST data management system. This includes the design of a data model and the development of algorithms to implement the most common ST operations, namely, data selection, projections and topological operations. These features will be iintegrated into a spatially-enabled DBMS, such as Secondo, PostgreSQL/PostGIS, or another extensible DBMS.

*WP3 (ST data visualization and user interaction) further extends previous work from the team towards efficient data retrieval methods for large spatiotemporal datasets. Previous work addressed appropriate data structures and algorithms that cache retrieved data based on the principle of temporal coherency/consistency of data [CaOlRo13]. Current work will also address spatial consistency and spatiotemporal consistency as the user interactively navigates in the dataset regarding both space and time. The aim is to reduce the turnaround time of retrieved data and thus, increasing the availability of data at the application level.

*WP4 (Case study 1: Testing the fireline propagation in smoke dispersion applications) consists of the implementation of a case study requiring modelling the fireline propagation in controlled burn experiments. GEMAC is a research group that produces annual reports on emissions caused by forest fires in Portugal (Kyoto Protocol commitments). These are calculated using a software (Farsite) to simulate fires spread, but WP4 will enable using real data.

*WP5 (Case Study 2: Modelling cells movement and morphological changes in biological processes) consists of creating a database to model the ST behavior of live cells. It will not only include time-lapse videos (raw data) but also ST data about cells movement and morphological changes across time. A frontend application for data retrieval based on ST features such as cell size variation or proximity between cells, will be developed.

*WP6 (Benchmarking). The case studies of this project will provide real data and information requirements demanding for the modelling of a wide range of ST features. Thus, we will prepare and publish a benchmark with workloads (ST data extracted from real datasets), query sets, metrics and reporting guidelines. This addresses one of the important shortcomings in previous work, where only synthetic data was used.

*WP7 (Dissemination of results).

It is expected that the results of this project will contribute for the following goals that so far are incompatible: to be able to model ST phenomena realistically and with small errors of estimation, using robust methods to guarantee that the data are valid at all times and compatible with the current spatial DBMSs. The achievement of these research goals is a key step towards the development of generic systems for the processing of ST data (analogous to a temporal GIS, but not restricted to geographical data) and could boost the development of new studies in other scientific and industrial domains, besides those studied in this project, e.g., studies on coastal erosion or river siltation.

In addition, this project will develop an integrated set of tools to create, manage, retrieve and visualize ST data that will be ready for practical use and has a strong component of real-world case studies, which are also advances relatively to previous work. It is expected that the continued use of this kind of tools in the future would encourage interoperability and data sharing between scientists and organizations. Currently, each research team develops its own tools, which is a major obstacle for data sharing.

This project has three members from the University of Aveiro (UA) and two members from INESC-TEC, with complementary fields of expertize. The research units from the UA that participate in this project are: IEETA and GEMAC. The IEETA, represented by José Moreira and Paulo Dias, is the leader of the work packages focusing on image processing (WP1), ST data management (WP2) and case study #2 (WP5). GEMAC is represented by Ana Miranda, an expert on smoke emissions and fire propagation, and leads WP4. The members of the INESC-TEC are Alexandre Carvalho and Marco Oliveira. INESC-TEC leads the work packages on visualization (WP3) and benchmarking (WP6).
INESC-TEC members share research interests in the area of visualization with Paulo Dias (IEETA) and in GIS with José Moreira (IEETA), so the communication between the members of the team will be easier.

The project has two consultants.
* Dr. Carlos Ortiz Solórzano, is director of the Cancer Imaging Laboratory and the Image Unit at CIMA, and a lecturer in the University of Navarra's Faculty of Medicine, School of Engineering (TECNUN) and CEIT-IK4, where he served as director of the PhD programme in Biomedical Engineering. He will be the consultant for WP5.

URL: http://ceit.es/en/press-room/news-and-events/1135-dr-carlos-ortiz-de-solorzano-full-professor-in-signal-theory-and-communications

* Ralf Hartmut Güting is a full professor in Computer Science at the University of Hagen, Germany, since 1989, and he is an expert on extensible database systems and spatiotemporal or moving objects databases. He is a Senior Associate Editor of the ACM Transactions on Spatial Algorithms and Systems and an Editor of GeoInformatica. He has previously served as an Associate Editor of the ACM Transactions on Database Systems and as an Editor of the VLDB Journal. He is the most cited author is ST databases and his participation as consultant will be of great value for this project.
URL: http://dna.fernuni-hagen.de/gueting/home.html

Descrição da Estrutura de Gestão

The University of Aveiro (IEETA and GEMAC) and the INESC-TEC (Porto) teams are committed to collaborate very closely on this research. José Moreira (IEETA) is the PI of the project; Alexandre Carvalho is the Co-PI and the leader of the project at INESC-TEC. They have participated in a former FCT project (PTDC/EIA/73531/2006) and since then, have continued to collaborate in research and academic activities. They are responsible for an effective development of the project, reinforcing the collaboration among the partners on the project goals, working methods and dissemination strategies. They will:
* Coordinate the implementation and the timeline of the Work Packages considering the given budget and the rules of FCT;
* Assess the production of deliverables and the accomplishment of the milestones, annual reports and the dissemination plan;
* Coordinate the inputs of external consultants and plan the visits to Hagen and Navarra.
* Promote the communication with stakeholders and the participation of students in the project.
* Present and discuss progress reports with partners.

The project consists of seven work packages. Each work package has a leader. The four leaders are senior researchers with experience in supervising students and scholarship holders, and experienced in various types of management and leadership activities, namely, as project manager, B.Sc. and M.Sc. programme director, research group director and department director. They are responsible for:
* Scientific and technical guidance of research and development activities.
* Assess the production of deliverables.
* Share information and discuss transversal issues with the other work package leaders.

This project has seven work packages. The work package dependences are shown in the project management diagram in annex and the articulation between work packages is included in the work package descriptions. The milestones are:
M1 Draft version of a dataset is ready for use (to be used in WP2 and WP3).
M2 Datasets and application for ST data preprocessing are ready & ST operations implemented (the tools for ST data preprocessing are ready and all raw data has been prepared to be used in WP2-WP6; the ST data model and query language are designed and a selection of operations is ready to be used in WP3).
M3 Prototype for ST data management and retrieval is ready & Demo for ST visualization is ready (the prototypes and demo applications are ready to be used in WP4-WP6).
M4 Final report is ready.
The deliverables by work package are listed in the timeline in annex.