ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 8, August 2016

Data Mining and Knowledge Innovation Tools for Managing Big Earth Observation Images

[1]B. Arunamma, [2]P.Balaji
[1] M.Tech CSE ,Student [2] Associate professor
Siddharth Institute of Engineering and Technology, Puttur

Abstract: **The continuous increase in the size of the archives and in the variety and complexity of Earth-Observation (EO) sensors require new methodologies and tools that allow the end-user to access a large image repository, to extract and to infer knowledge about the patterns hidden in the images, to retrieve dynamically a collection of relevant images, and to support the creation of emerging applications (e.g.: change detection, global monitoring, disaster and risk management, image time series, etc.). In this context, deals with knowledge discovery from Earth-Observation (EO) images, related geospatial data sources and their associated metadata, mapping the extracted low-level data descriptors into semantic classes and symbolic representations, and providing an interactive method for efficient image information mining. with providing a platform for data mining and knowledge discovery content from EO archives. The platform's goal is to implement a communication channel between Payload Ground Segments and the end-user who receives the content of the data coded in an understandable format associated with semantics that is ready for immediate exploitation. It focuses on the design and implementation of methods for the extraction of relevant descriptors (features) of EO images, specifically Terra SAR-X images, physical integration (fusion) and combined usage of raster images and vector data in synergy with existing metadata. It will provide the user with automated tools to explore and understand the content of highly complex images archives. The challenge lies in the extraction of meaningful information and understanding observations of large extended areas, over long periods of time, with a broad variety of EO imaging sensors in synergy with other related measurements and data.**

Index Terms**— Systems to manage Earth-Observation images, data mining, knowledge discovery, query engines**

I. INTRODUCTION

In recent years the ability to store large quantities of Earth Observation (EO) satellite images has greatly surpassed the ability to access and meaningfully extract information from it. The state of- the-art of operational systems for Remote Sensing data access (in particular for images) allows queries by geographical location, time of acquisition or type of sensor. Nevertheless, this information is often less relevant than the content of the scene (e.g. specific scattering properties, structures, objects, etc.). Moreover, the continuous increase in the size of the archives and in the variety and complexity of EO sensors require new methodologies and tools - based on a shared knowledge - for information mining and management, in support of emerging applications (e.g.: change detection, global monitoring, disaster and risk management, image time series, etc.). Along the years, several solutions were presented for accessing the Earth-Observation archives as for example queries of the image archive using a small number of parameter like: geographical coordinates, acquisition times, etc.(Wolfm¨uller et al., 2009).

Later, the concept of query by example allowed to find and retrieve relevant images taking into account only the image content, provided in the form of primitive features, several systems following this principle appeared for instance (Agouris et al., 1999), (Mu˜noz and Datcu, 2010), (Datcuet al., 2003),(Shyu et al., 2007). However, later the problems of matching the image content (expressed as primitive features in the low level of processing) with semantic definitions adopted by human were evident; causing the so-called semantic gap (Smeulderset al., 2000). With the semantic gap, the necessity of semantic definition was clearly demonstrated. In an attempt to reduce the semantic gap, more systems including labeling or definition of the image content by semantic names were introduced. In this contex ,here concerned ithprovidingaplatformfordataminingandknowledgediscoverycontentfromEOarchives.The platform's goal is toimplementacommunicationchannelbetweenPayloadGroundSegmentsandtheend-userwhoreceivesthecontentofthe data coded in an understandable format associated with semantics that is ready for immediate exploitation. Here, it highlighted the importance of integration of interactive geospatial data visualization with statistical data mining algorithms. A 3D visualization and interactive exploration of large relational data sets through the integration of several multidimensional data visualization techniques and for the purpose of visual data mining and exploratory data analysis was presented in (Yang, 2003). Here, the experiments were done using more than a million records in a relational database. Recently, as an advanced example of visual data mining system implementation, the system called Immersion Information Mining was introduced in (Babaee et al., 2013). This system uses virtual reality and is based on visual analytic approach that enables knowledge discovery from EO archives.

II. SYSTEM ARCHITECTURE

The Knowledge Discovery (KDD) component of system intends to implement a communication channel between the Earth Observation (EO) data sources and the final user (end-user) who receives the content of the data sources coded in an understandable format associated with semantics and ready for its exploitation. Figure 1 shows the different KDD components and the interaction between them. Here, it can be observed that all the components rely on a Data Base Management System. The end-user interacts with three components, while the data sources are processed by the data model generation

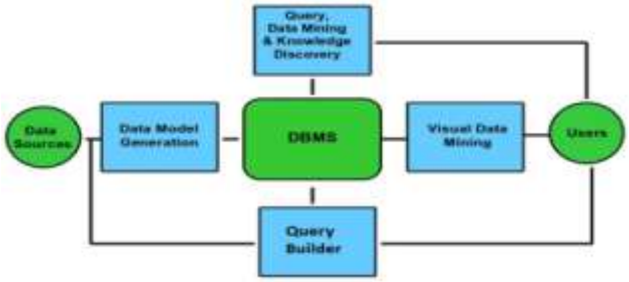

Fig 1: Components of knowledge Discovery

The KDD component implementing the Query, Data Mining & Knowledge Discovery, Visual Data Mining, and Interpretation and Understanding modules process the Earth-Observation Data Model in interaction with the user, update it, and generate as output the information required by the user in a format adapted to the application context.

The system can help the user to deal with large image collections by accessing and extracting automatically its content (Data Model Generation), querying by means of image content and semantics and mining relevant information (Query Data Mining and Knowledge Discovery), inferring knowledge about patterns hidden in the images (Interpretation and Understanding), and visualizing the image collections (Visual Data Mining).

II. SYSTEM MODULES: HERE DESCRIBING FOR EACH MODULE THE IMPLEMENTED FUNCTIONALITY.

Data Model Generation (DMG) module

The *image content analysis* of SAR images is composed of the generation of patch quick-looks together with image descriptor extraction. Feature extraction methods such as gray level co-occurrence matrix and Gabor filters, and compression based on dictionaries are used in the case of Terra SAR-X image ingestion. The *metadata extraction* is composed of the parsing the xml annotation file in the case of Terra SAR-X. In the case of optical images; the metadata is extracted from the GeoTIFF header. The *image content analysis* of optical images is composed of the quick-look generation together with image descriptor extraction. The color histogram and compression based on dictionaries are used as feature extraction methods. The *ingestion of EO sources* is composed of the creation of collections and selection of the EO products to be processed.

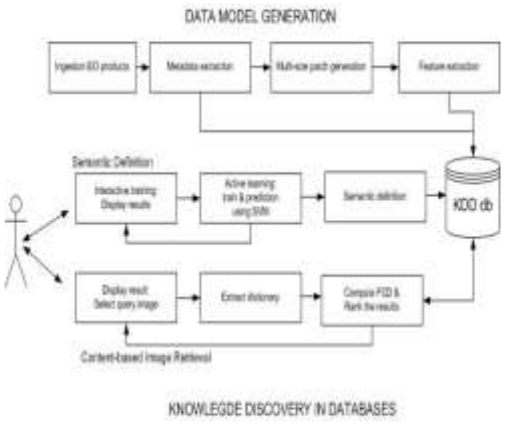

Fig 2: Knowledge Discovery in Database

DBM module

An updated version of the database schema. The mapping of the different tables into Java classes together with a class for manipulating the database operations such as connections, queries, etc. Implementation of user-defined functions based on SQL for extending the database functionality. The Data Model generated by the DMG is wrapped into tables and columns, and stored into a relation database system. DBMS provides the tools for accessing and manipulating the tables of the data base as well as enabling and granting the access to the data mining methods and queries. The main entities in the data model are: EO product, image, meta data, patch, primitive features ,semantic labels and annotation.

Query, Data Mining and Knowledge Discovery (KDD) module

In the **KDD module**, since the features, patches and all related information are available in the database, the repository is accessed in order to provide data mining functionality, semantic definition and Content-based Image Retrieval (CBIR). This module has query, data mining and knowledge discovery components, providing an integrated system where the user strongly interacts with the system to perform advanced content-based image query and retrieval (e.g. label definition, query by example, etc.)

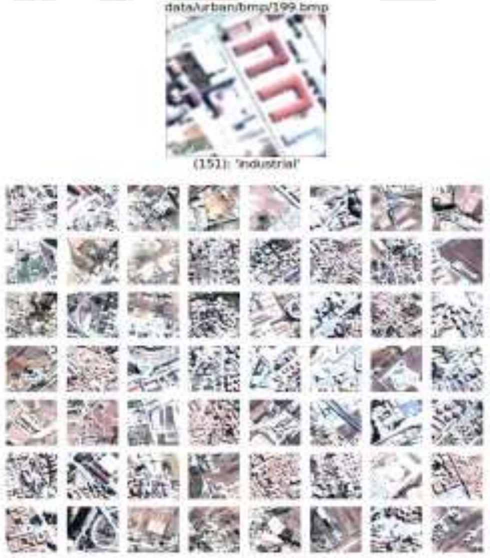

Figure 3: GUI of Content-based Image Retrieval

In this module distinguish three main functionalities:

- ❖ Semantic definition using active learning methods and relevance feedback tools.
- ❖ Content-Based Image Retrieval using the concept of query based on example. It is implemented using compression based techniques to describe the image content, and distance measures to discover similarity between patches in the database.
- ❖ Queries using metadata and semantics using SQL statements and combining the information stored in the tables.

Query builder and Image Understanding (IU) module

The image interpretation and understanding process is based on semi-supervised learning methods complemented with active learning and relevance feedback tools. This application involves four main functions.

The primitive feature extraction is done in the data model generation. Later, the feature vectors together with the patch quick-looks are exported in order to be loaded into this tool.

- ***Create application projects:*** The final user can create different application scenarios by creating different projects. In order to create a project is needed to specify the path where the quick-looks and the primitive features are stored. The primitive features are stored as binary files and the quick-looks are jpg files.
- ***Semantic definition*****:** The semantic definition is based on an adaptation of the support vector machine. The method is based on the training and feedback by giving positive and negative examples. It is an iterative loop, where the user gives the positive and negative examples and the system trains the data and retrieves the results.
- ***Image visualization*****:** In order to verify the semantic definition, the different patches are shown with different colors in the image visualization panel. Figure 13 presents an example of visualization of the image content using this tool.

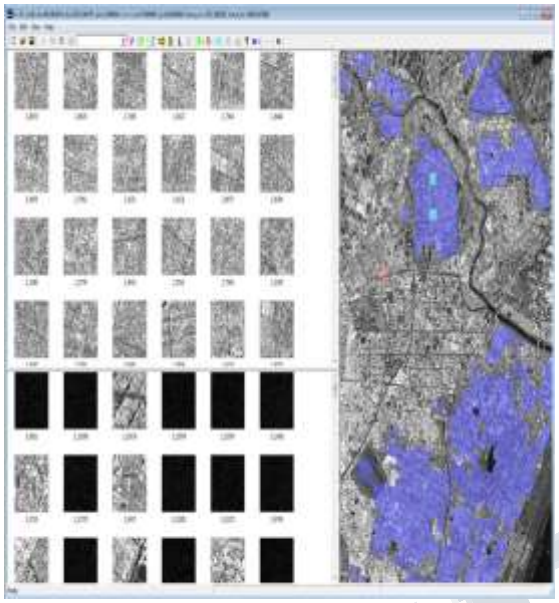

Fig 3: Image content with positive and Negative example

E. Visual Data Mining (VDM) module

The Visual Data Mining component is a tool which is able to visualize feature spaces of large repositories of images interactively. Visual Data Mining relies on powerful Human-Machine Interfaces (HMI) and Graphical User Interfaces (GUI) with functionalities such as browsing, querying, zooming, etc, thus, enabling the end-user to exploit the image database. This tool allows interactive exploration and analysis of very large, high complexity, and non-visual data sets stored into the database. It provides to the end-user an intuitive tool for Data Mining by presenting a graphical interface, where the selection of different images and/or image content in 2-D or 3D space is achieved through visualization techniques, data reduction methods, and similarity metrics to group the images.

The image database. This tool allows interactive exploration and analysis of very large, high complexity, and non-visual data sets stored into the database. It provides to the end-user an intuitive tool for Data Mining by presenting a graphical interface, where the selection of different images and/or image content in 2-D or 3D space is achieved through visualization techniques, data reduction methods, and similarity metrics to group the images.

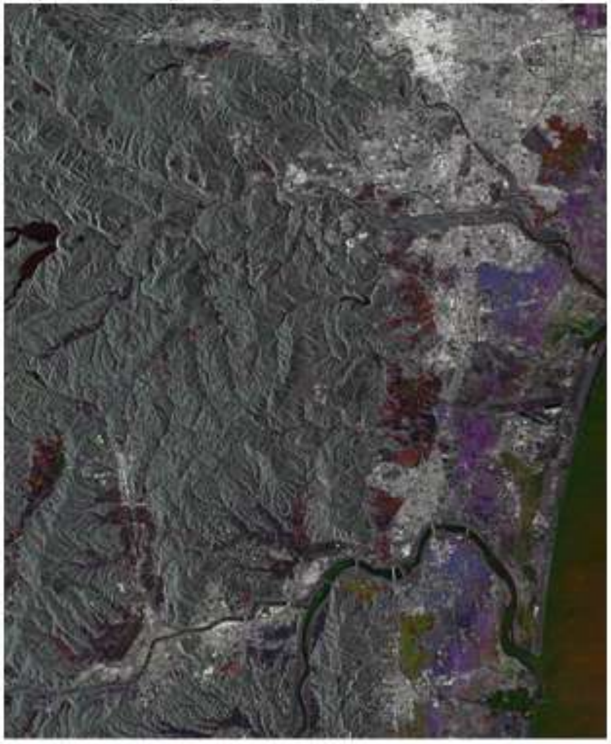

Fig 4.Visualization of the image content using image Understanding tool

The Visual Data Mining offers the following functionalities:

1.Dimensionality reduction deals with converting the n dimension feature vector into vector of much lower dimensionality (i.e 3 dimension) such that each dimension convey much more information. In VDM module the Gabor feature vectors with 48 dimensions is converted into vector of 3 dimensions.

2. ***Visualization*** of the collection content in a 3D-space.

The following steps are performed in the visual data mining workflow:

1. The Visual Data Mining reads the input features and its associated metadata from a text file(s).

2. It reads the JPEG image quick-looks

3. It projects the n-dimensional space into a 3-dimensional space by choosing 3 coordinates out of the total number per line (e.g. 48) of read records from the input features file.

4. It visualizes the thumbnails in a 3-dimensional space associated to their description projections; scaling, zooming and rotation navigation functions are provided.

5. Quick-looks could be pinned in order to display the semantic label, the identifier of the patch, and the number of quick-looks covered by the respective quick-look.

The graphical user interface of Visual Data Mining is presented in Figure 5

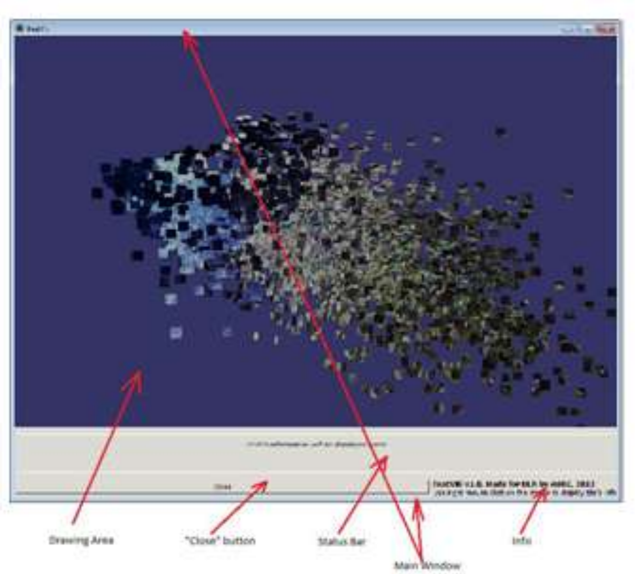

Fig 5: Graphical User Interface of Visual Data Mining tool.

III. CONCLUSION

Here presented several tools for data mining and knowledge discovery in order to exploit big Earth-Observation image archives in a fast and efficient way. The architecture is presented as a modular system integrating several components with well-defined functionality. The main operation of the system starts with the ingestion of different EO data sources during the data model generation. ,the system components can help the end-user to deal with large image collections by accessing and extracting automatically their content(Data Model Generation),allowing querying(by means of image content and semantics),mining relevant information (Query Builder),inferring knowledge about patterns hidden in the image archive(Data Mining and Knowledge Discovery),and visualizing the complete image database.

REFERENCES

[1] Agouris, P., Carswell, J. and Stefanidis, A., 1999. An environment for content-based image retrieval from large spatial databases. ISPRS Journal of Photogrammetry and Remote Sensing 54(4), pp. 263 – 272.

[2] Babaee, M., Rigoll, G. and Datcu, M., 2013. Immersive Interactive Information Mining with Application to Earth Observation

Data Retrieval. In: Availability, Reliability, and Security in Information Systems and HCI, Lecture Notes in Computer Science,Vol. 8127, Springer Berlin Heidelberg, pp. 376–386.

[3] Bifet, A., 2013. Mining big data in real time. Informatica (Slove-nia) 37(1), pp. 15–20.

[4] Chang, C.-C. and Lin, C.-J., 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, pp. 27:1–27:27

[5].Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X. and Gao, W., 2010. WLD: A Robust Local Image Descriptor. IEEE Transactions on Pattern Analysis and Machine Intelligence32(9), pp. 1705–1720.

[6] Costache, M., Maitre, H. and Datcu, M., 2006. Categorization based relevance feedback search engine for earth observation images repositories. In: IEEE International Conference on Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006.,pp. 13 –16.

[7] Cui, S.,Dumitru, C. and Datcu, M., 2013a. Ratio-Detector-Based Feature Extraction for Very High Resolution SAR Image Patch Indexing. IEEE Geoscience and Remote Sensing Letters 10(5), pp. 1175–1179.

[8]Cui, S., Dumitru, O. and Datcu, M., 2013b. Semantic annotation in earth observation based on active learning. International Journal of Image and Data Fusion pp. 1–23.

[9] Datcu, M., Daschiel, H., Pelizzari, A.Quartulli, M., Galoppo, A. Colapicchioni, A., Pastori, M., Seidel, K., Marchetti, P. and D'Elia, S., 2003. Information mining in remote sensing image archives: system concepts. IEEE Transactions on Geo science and Remote Sensing 41(12), pp. 2923 – 2936.

[10] de Oliveira, M. F. and Levkowit, H., 2003. From visual data exploration to visual data mining: a survey. IEEE Trans. Visual. Comput. Graphics 9(3), pp. 378–394.
[11]DLR, 2007. TerraSAR-X, Ground Segment, Level 1b Product Data Specification, TX-GS-DD-3307. http://sss.terrasar-x.dlr.de/pdfs/TX-GS-DD-3307.pdf.

[12]Espinoza-Molina, D. and Datcu, M., 2013. Earth-Observation Image Retrieval Based on Content, Semantics, and Metadata.IEEE Transactions on Geoscience and Remote Sensing 51(11),
pp. 5145–5159.

[13]Espinoza-Molina, D., Datcu, M., Teleaga, D. and Balint, C.,2014. Application of visual data mining for earth observation use cases. In: ESA-EUSC-JRC 2014 - 9th Conference on Image Information Mining Conference: The Sentinels Era, pp. 111–114.

[14]Fan, W. and Bifet, A., 2013. Mining big data: Current status, and forecast to the future. SIGKDD Explor. Newsl. 14(2), pp. 1–5.

[15]Keim, D., Panse, C., Sips, M. and North, S., 2004. Visual data mining in large geospatial point sets. IEEE Comput. Graph. Appl.24(5), pp. 36–44.