ASTROINFORMATICS: IMAGE PROCESSING AND ANALYSIS OF DIGITIZED ASTRONOMICAL DATA WITH WEB-BASED IMPLEMENTATION

N. Kirov[1,2]**, O. Kounchev**[2]**, M. Tsvetkov**[3]

[1]Computer Science Department, New Bulgarian University
[2]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences
[3]Institute of Astronomy, Bulgarian Academy of Sciences
nkirov@nbu.bg, kounchev@gmx.de, milcho@skyarchive.org

ABSTRACT

The newly born area of Astroinformatics has arisen as an interdisciplinary domain from Astronomy and Information and Communication Technologies (ICT). Recently, four institutes of the Bulgarian Academy of Sciences launched a joint project called "Astroinformatics". The main goal of the project is the development of methods and techniques for preservation and exploitation of the scientific, cultural and historic heritage of the astronomical observations. The Wide-Field Plate Data Base is an ICT project of the Institute of Astronomy, which has been launched in 1991, supported by International Astronomic Union [1]. Now over two million photographic plates, obtained with professional telescopes at 125 observatories worldwide, are identified and collected in this database. Digitization of photographic plates is an actual task for astronomical society. So far 150 000 plates have been digitized through several European research programs. As a result, a huge amount of data is collected (with about 2TB in size). The access, manipulation and data mining of this data is a serious challenge for the ICT community. In this article we discuss goals, tasks and achievements in the area of Astroinformatics.

1   THE WIDE-FIELD PLATE DATA BASE

The basic information for all wide-field ($> 1^{o}$) photographic astronomical plate archives and for their content, stored in many observatories is collected in the Wide-Field Plate Database (WFPDB, `skyarchive.org`), established in 1991 and developed in the Institute of Astronomy, Bulgarian Academy of Sciences. All known plate archives (with total number 442) are separated by the telescope with which they were produced. In these archives there are more than two million plates according to the information from the last version of the Catalogues of the Wide-Field Plate Archives [2], which is an integrated part of the WFPDB. Another part of the database is the Catalogue of Wide-Field Plate Indexes, where now the descriptive information from 123 archives with more than 500 000 plates. The digitized plate images – with low resolution for quick plate

visualization and easy online access, as well as with high resolution scans for astronomical investigations, are object of established Data Bank with quickly increased content. The WFPDB provides also links to online services and cross-correlation with existing catalogues and journals. The development of the WFPDB aims as its enlargement with new catalogues and digital plate images, organized in a database, as well as in providing direct access of the WFPDB user to the logbooks carried out during the observations by the observer.

One of the main applications of such database is motivated by the recently increased interest in the search and identification of Potentially Hazardous Asteroids (PHAs) which might cause catastrophic events by their collision with Earth (ESA and NASA have established a special Near-Earth Object survey program, (see `spaceguard.esa.int` and `impact.arc.nasa.gov`) devoted to the Earth impact threat by PHAs). In this context the determination of the orbits of PHAs through historical astronomical data is one of the main approaches, and this is one of the main motivations nowadays for the preservation of the huge amount of astronomical plates by means of scanning and digitization, followed by compression and image processing methods.

2 ASTROINFORMATICS

We can find definitions of the new scientific area of Astroinformatics in Internet:

"*Astroinformatics includes a set of naturally-related specialties including data organization, data description, astronomical classification taxonomies, astronomical concept ontologies, data mining, machine learning, visualization, and astrostatistics.*" Kirk D. Borne, "Astroinformatics: A 21st Century Approach to Astronomy" (`arxiv.org/abs/0909.3892`)

"*This project (astroinformatics.de) is founded to explore the limits of astronomical image processing.*" Thilo Bauer, `astroinformatics.de`

The newly born area of Astroinformatics has emerged from a synthesis between Astronomy and modern information and communication technologies, i.e. Astroinformatics is a new interdisciplinary area based on advanced Internet applications (see [3]). Recently, four institutes of the Bulgarian Academy of Sciences – Institute of Mathematics and Informatics, `www.math.bas.bg`, Institute of Astronomy, www.astro.bas.bg, Institute of Information Technologies, `www.iinf.bas.bg`, and Central Laboratory for Geodesy, `clg.cc.bas.bg` – have launched a joint Astroinformatics project (`www.astroinformatics.eu`).

The main problems addressed by the Astroinformatics have the following components:

- Improvement and development of existing algorithms for search in astronomical data bases, as well as in the WFPDB (`www.skyarchive.org`);

- Set up of data base giving interactivity and inter-operability in making improvement and development of new methods for compression, representation and visualization of the images extracted from the photographic plates (see [4]), which methods are problem-orientated to the theory and practice of astronomy in the frames of the project;

- Set up of Web access to the digitized original log-books, containing original data of the observers with applying a technology for objects localization in scanned binary images;

- Improvement of the access to the modern astrometric catalogues created and upgraded last decades on the base of the Earth observations and from the Hipparcos and Tycho catalogues.

The main goal of the project is usage of information communication methods and technologies for representation, processing and preservation of astro-images in specialized Plate Data Base and effectively search in WFPDB. Other objectives are:

- Creation of the platform independent distributed database, supplying unified access to accumulated digital images of wide-field astronomical photographic plates constituting a constantly growing volume willing to approach 2PB.

- Adapting of methods of fast access to image data base, image recognition, and signal and image processing and their application to the WFPDB;

- Creation of problem-oriented method for astronomical images compression;

- Content Based Image Retrieval (CBIR) of astro-images from astronomical data base of images;

- Development of problem-oriented CBIR methods to WFPDB/ABDI;

- Application of OCR algorithm of information extracting from telescope registers;

- Implementation of Virtual Observatory Standards for data description and processing;

- Development of WEB-based access to the digitized wide-field photographic astronomical plates;

- Creation of WEB-based tools of:
    - Fast search of stars and specific astronomical objects from digitized star images;
    - Star catalogues improving by means of WFPDB

- Adapting of methods for:
    - Analysis of parameters of variable star;
    - Pattern recognition;
    - Statistical tests of Pleiades cluster data from WFPDB;

- Dissemination of the scientific team experience.

Part of the objectives are connected with another implication of the digitized astronomical plates for star catalogues improvement and their use by other sciences (such as geosciences, space physics etc.)

The scientific researches in both Information and Communication Technologies and Astronomy, as well as in the interdisciplinary research area Astroinformatics are the core working packages of the project. They form the three main objectives of the project and consist of 8 working packages:

WP1 User Requirements
The user requirements study targets collecting information from the potential project users and related organizations necessary to formulate strong conceptual software architecture as well the key entities for the WFPDB development.

WP2 Market Survey & State of the Art
This task provides an overview on the available contemporary technologies for large data storage, preservation, web-based access and analysis useful for planned WFPDB development. It gives insight on the available specialized software technologies for manipulating digitized astronomical images, data analysis, and the ones describing the celestial coordinate system. It helps the project management by proposing useful software and hardware updates, when needed, for the future work on the project.

WP3 Data storage & preservation
This is the most important working package of the program as it is planned to give a system for storage and preservation of the important astronomical data that cover a large time span of 130 years professional astronomical observation with photography methods. This package is the basis for the implementation of the web-access to the data, and in the same time it provides tools for the scientific research and development in ICT, Astronomy, and Geodesy. Astronomical plate digitization is a specific and time-consuming work (for instance the digitizing the Rozhen National Observatory plate archive, holding about 9000 plates taken with 2 m RCC and 50/70 cm Schmidt telescopes will take 20 p/yr and the data storage estimated about 3Tb). Since the plates are aging, it is important that their preservation is done on time, avoiding the potential data loss in time. Having in mind the planned future growth of the digitized images collection we need new technologies for reliable data storage.

WP4 Virtual Observatory standards compatibility
This WP aims the implementation of the world-wide accepted standards of IVOA for information and communication access and analysis protocols. The main objectives are: implementing of EuroVO presentation and access protocols to the digitized astronomical plates; setting up a procedure for data extraction from the active by now parts of the WFPDB – Catalogue of Wide-Field Plate Archives, Catalogue of Wide-Field Plate Indexes, and the created within WP3 image databases (low and high resolution) according the EuroVO data format standards.

WP5 Web-based access
Providing tools for web-based access and analysis of the stored digitized astronomical plate observation is the main goal of the WP. The obtained results will benefit both astronomical and computer science institutes by giving access to archives with 130 years old observations. Within this WP a specialized search engine integrated into the web application will be developed, which will be able to operate in the large database with astronomical images, prepared in the previous working packages.

WP6 Scientific research in ICT
The purpose of this WP is to transfer and adapt appropriate ICT methods, approaches, algorithms and tools in resolving various problems, primary related to astronomical image analysis, access methods to image databases, data visualizations, searching efficiency, detection of known constellations, compression methods etc. The digitized astronomical plates residing in the WFPDB, represent unique astronomical material covering the sky observations since 130 years ago, which provides valuable resource of information for the researchers in the astronomical area. The scientific research in ICT aims from one side to provide information technologies serving the astronomers in detecting new events, sky objects and phenomena during this long period and from the other to expand the computer science with new generalizations, serving other application areas.

WP7 Scientific Research in Astronomy

- Searching observational data for specific astronomical objects (small Solar system bodies, quasars, supernova, FU Ori stars, etc.)
- Position of the observed astrometrical stars in the Celestial reference frame during the period 1890-1992 and separation of the potential double stars from the single observed stars.

- Searching of the places of the astronomical plates, containing the single stars, observed during the period 1890-1992.
- Method for determination of celestial coordinates of stars from digitized astronomical plates.
- Determination of celestial coordinates of stars by means of WFPDB.
- Investigation of the accuracy of astrometrical catalogues improvement by model of simulated observations.
- Statistical test method of the observational data, based on a long term "monitored" (approx. 50 years) Pleiades cluster
- Evaluation of the accuracy estimation of the star proper motion, determined by means of WFPDB
- Wavelet methods for investigation of light curve variations of variable stars.
- Time series analysis of light curve variations of variable stars.
- Investigation of light curve variations of variable stars.

WP8 Dissemination

A main purpose of the dissemination measures is to make the international community aware of the urgent necessity to digitize the accumulated heritage of wide-field astronomical photographic plates. Collaboration with international partners will help for the implementation of the project results within the institution keeping archival plates.

3 SUMMARY

The main objective of the new scientific area Astroinformatics is the development of ICT methods and instruments for digitization, processing, representation and storage of astronomical images in specialized databases for efficient data retrieval. The methods may be characterized as modern ICT for databases management and for WEB-based information system, in particular for image analysis, compression, digitization and recognition, image retrieval, Fourier analysis and wavelet analysis.

The main expected result of the described project is the creation of an ICT system for data preservation and Web-based access to digitized astronomical plates, and its future inclusion into the Euro-VO for contemporary astronomical and interdisciplinary research.

REFERENCES

[1] M. K. Tsvetkov, *Wide-Field Plate Archives*, International Astronomical Union Commission 9 Working Group on Wide-Field Imaging, Newsletter No. 1, (1991), 17.

[2] K. P. Tsvetkova, M. K. Tsvetkov *Catalogue of Wide-Field Plate Archives: version 5.0* In: Virtual Observatory: Plate Content Digitization, Archive Mining and Image Sequence Processing, Eds. M. Tsvetkov, V. Golev, F. Murtagh, R. Molina, Heron Press, Sofia, 2006, 45-53.

[3] O. Kounchev, et al., *Astroinformatics: A Synthesis between Astronomical Imaging and Information & Communication Technologies*, Modern Trends in Mathematics and Physics ed. S.S. Tinchev, Heron Press, Sofia, 2009.

[4] J.-L. Starck, F. Murtagh, *Astronomical Image and Data Analysis*, Springer, 2003.