Automated Data Mining of the Reference Stars From Astronomical CCD Frames

Sergii Khlamov[1], Vadym Savanevych[1], Tetiana Trunova[1], Zhanna Deineko[1], Oleksandr Vovk[1] and Roman Gerasimenko[1]

[1] *Kharkiv National University of Radio Electronics, Nauki avenue 14, Kharkiv, 61166, Ukraine*

Abstract

In astronomical images obtained using telescopes and cameras, there are from 1 to 100 thousand or more stars depending on the resolution and exposure time. These objects are fixed against the background of the frame and have constant positions in the celestial sphere. To determine which part of the sky corresponds more accurately to a given frame, it is necessary to associate the frame with known astronomical astrometric and photometric catalogs. These catalogs contain millions of position values of various stars as static objects. Having such information in the form of big data, as well as a huge amount of classified and clustered data in the form of databases, computational methods for fast extraction of the necessary data from them need to be developed. For this purpose, classical methods of "knowledge discovery in databases" (KDD) and Data Mining exist. However, for their proper application, it is necessary to classify the input data set for subsequent analysis and rejection. The implementation of these methods is closely related to the developed mathematical computational methods for automatic selection of reference stars in astronomical images. The result is implemented in the Lemur software of the CoLiTec (Collection Light Technology) project for the astronomical data processing using the data mining methods.

Keywords

Reference stars, data mining, big data, knowledge discovery in databases, astronomical catalogues, astrometry, photometry, recognition patterns, image processing, series of images, CCD-frames

1. Introduction

Technological progress in the production of cameras as charge-coupled devices (CCD) [1] and telescopes demonstrates continuous acceleration, reflecting modern trends in scientific and engineering developments. New materials and more efficient components, such as photodetectors and optical systems, contribute to the creation of cameras and telescopes with increased resolution and sensitivity.

Today's digital cameras possess impressive characteristics, including resolutions exceeding 100 megapixels, significantly surpassing those of previous models. Telescopes are also keeping pace: it is expected that the Large Synoptic Survey Telescope (LSST) (Figure 1) will have a resolution of around 3.2 gigapixels [2]

The speed of data acquisition and processing has also increased noticeably. Modern cameras are capable of recording and processing data orders of magnitude faster than their predecessors, achieving shooting speeds of up to 20 frames per second while maintaining full resolution. The development of parallel computational algorithms contributes to more efficient processing of such a big data [3] obtained from cameras and telescopes, opening up new possibilities for scientific research and practical applications in various fields including astronomy [4].

COLINS-2024: 8th International Conference on Computational Linguistics and Intelligent Systems, April 12–13, 2024, Lviv, Ukraine
sergii.khlamov@gmail.com (S. Khlamov); vadym.savanevych1@nure.ua (V. Savanevych); tetiana.trunova@nure.ua (T. Trunova); zhanna.deineko@nure.ua (Zh. Deineko); oleksandr.vovk@nure.ua (O. Vovk); roman.herasymenko@nure.ua (R. Gerasimenko)
0000-0001-9434-1081 (S. Khlamov); 0000-0001-8840-8278 (V. Savanevych); 0000-0003-2689-2679 (T. Trunova); 0000-0003-0175-4181 (Zh. Deineko); 0000-0001-9072-1634 (O. Vovk); 0009-0008-7948-5851 (R. Gerasimenko)

As the resolution of cameras and telescopes has increased over time, allowing for the registration and analysis of more detailed images of the night sky, this process directly impacts the ability to detect and study various objects in space [5]. Additionally, wide-angle cameras allow for an expanded field of view, significantly increasing the number of objects captured in a single frame. This is particularly useful when studying areas with high stellar density or when searching for distant galaxies and cosmic objects [6].

Figure 1: Large Synoptic Survey Telescope (LSST) under construction

However, even with the increase in resolution and field of view, the quality of astronomical imaging of objects, especially those located at great distances and with low brightness, remains a challenge for modern cameras and telescopes. One of the main factors affecting image quality is the level of noise. When imaging faint objects, even a small amount of noise on the camera sensor can significantly distort the image and hinder its interpretation. Despite significant improvements in noise reduction in modern cameras, this aspect remains a problem when dealing with very faint and distant objects [7].

Another important factor is atmospheric conditions. Atmospheric turbulence and atmospheric distortions can significantly affect the quality of images and its typical shape or form [8], especially when working with high magnifications. This limitation can be overcome by using space telescopes, which are located beyond the Earth's atmosphere.

It is also worth considering the technical limitations of cameras and telescopes, such as limited sensitivity to certain wavelengths or limitations on dynamic range. Research and development efforts continue with the aim of overcoming these limitations; however, this remains a relevant area of research in astronomy and optics.

With these improvements in cameras and telescopes comes an increase in the volume of data in astronomical catalogs [9]. This is due to the increasing number of observations conducted, the use of more advanced instruments, and the expansion of the coverage area of the celestial sphere. So, it is necessary to develop the mathematical methods for the automated data mining [10] of the reference stars [11] from astronomical CCD-frames received from a such big volume of astronomical data.

2. Related Works

2.1. Astronomical big data sources

In recent decades, astronomy has undergone a data processing and analysis revolution due to the implementation of large astronomical projects and the use of advanced observational tools. These modern instruments collect and process vast amounts of data, extracting valuable information about celestial objects and phenomena [12].

One of the main sources of data is ground-based telescopes and observatories. These facilities, located around the world, gather data on stars, galaxies, and other celestial objects using various observation methods, including optical, infrared, and ultraviolet spectra. Thanks to ground-based telescopes and Virtual Observatories [13], astronomers have access to a wide range of objects and can study their properties and evolution [14].

New telescopes are capable of generating huge volumes of high-resolution data. This includes not only images but also spectroscopic data and information on temporal changes in the brightness of objects. Modern big data processing techniques and machine learning allow for efficient extraction of information from data arrays, identification of interesting objects, and automatic compilation of astronomical catalogs [15]. Additionally, international projects and collaborations are a contributing factor. Astronomers from around the world join forces to conduct joint observations and create extensive catalogs, leading to more comprehensive coverage of the celestial sphere and enrichment of data [16]. This growth of data in astronomy provides unique opportunities for scientific research.

Pattern recognition of both celestial objects and patterns themselves is an important tool for analyzing data obtained from astronomical surveys and observations [17]. Pattern recognition involves the process of analyzing and identifying astronomical objects in images obtained from telescopes and space observatories. This process may include the detection and classification of stars, galaxies, cosmic objects, and other astronomical phenomena based on their characteristics such as position [18], brightness [19], shape, spectral features, and others. Initially, data collection and preparation occur – this may include observational information from telescopes, including images, spectra, and time series. Then, data processing, including noise filtering, background alignment [20], image quality enhancement, and identification of objects of interest.

Space telescopes and missions are also important sources of astronomical data. Space telescopes such as Hubble [21], Kepler, and the upcoming James Webb Space Telescope provide high-quality data on distant galaxies, exoplanets, and other objects in the Universe. These missions allow astronomers to explore cosmic objects without the influence of Earth's atmosphere and expand our knowledge of the Universe [22].

Radioastronomical observations also play a crucial role in astronomical research. Using radio telescopes such as the Very Large Array (VLA) and ALMA, astronomers study radio sources and processes occurring in the radio spectrum. These observations provide information on various phenomena, including active galactic nuclei, radio pulsars, and cosmic microwave background radiation [23].

Recently, the detection of gravitational waves has become a new source of astronomical data. Using interferometers such as LIGO and Virgo, astronomers detect events such as black hole mergers and neutron star mergers. These observations provide new data on cosmic phenomena that were previously inaccessible to observation.

All astronomical data sources even small play a key role in scientific research, providing astronomers with valuable information about cosmic objects and processes. Modern observational tools and data processing technologies allow researchers to expand our knowledge of the Universe and address important scientific questions.

2.2. Data mining in astronomy

Data mining, in the context of astronomical research, is a method of data analysis that relies on computational algorithms to discover patterns, trends, and structures in space data [24]. With the increasing volume of astronomical data generated by advancements in observational technologies and expansion of spatial coverage, data mining becomes a crucial tool for extracting valuable information from these data arrays.

Astronomical data is often characterized by high dimensionality, complex structure, and a significant amount of noise. Data mining enables efficient processing and analysis of such data, revealing hidden patterns that may not be apparent at first glance. Data mining methods encompass various approaches such as clustering, classification, regression, associative rules, and others.

Applied to astronomical data, data mining can serve various purposes, including [25]:

- discovery of new object classes: using clustering and classification methods, data mining can unveil new classes of astronomical objects hidden within the dataset.
- identification of correlations and dependencies: by analyzing multiple parameters and characteristics of astronomical objects, data mining can help identify correlations and dependencies among them, leading to new scientific discoveries and understanding of physical processes.
- prediction of temporal changes: using regression and time series methods, data mining can be employed to predict temporal changes in luminosity or other characteristics of astronomical objects.

Thus, data mining represents a powerful tool for the analysis of astronomical data [26], playing a significant role in unraveling the mysteries of the Universe and deepening scientific understanding of space.

2.3. Knowledge discovery in databases in astronomy

Pattern recognition of celestial objects [27] and patterns is closely related to concepts such as data mining and knowledge discovery in databases (KDD) in the context of astronomical research.

Data mining is the process of automated analysis of large volumes of data to identify interesting and non-obvious patterns, templates, and trends. In astronomy, where data on astronomical objects are extremely extensive and multiparametric, data mining plays a key role in processing and analyzing this data. Pattern recognition of objects and stars is one of the stages of this process, where machine learning algorithms [28] and statistical methods [29] are applied for classification and clustering of objects in images or observational data.

On the other hand, KDD encompasses a wide range of methods and techniques for identifying and interpreting patterns and new knowledge from databases. In astronomy, where data often have a complex structure and may contain noise, KDD helps researchers identify hidden relationships between various parameters of astronomical objects, leading to the discovery of new physical laws and understanding of the Universe [30]. Thus, for pattern and star recognition, the use of techniques such as data mining as well as KDD in astronomy is an important aspect as they help researchers gain valuable knowledge from cosmic astronomical data. Knowledge discovery in databases in astronomy is a methodology for analyzing and interpreting extensive datasets collected from various observational sources in space (Figure 2).

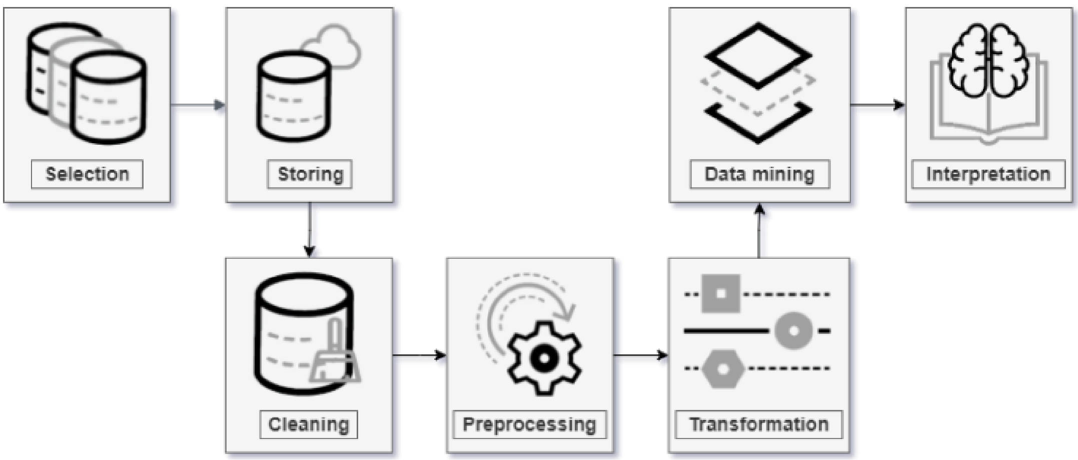

Figure 2: Knowledge discovery in databases in astronomy

This approach involves the application of various algorithms and models to identify important patterns, trends, and structures within astronomical data, enabling astronomers to extract new knowledge and draw scientific conclusions. Knowledge discovery in databases in astronomy leads to a number of scientific outcomes, including:

- classification of stellar spectra: utilizing machine learning methods and data analysis, astronomers can classify stars based on their spectral characteristics. For example, clustering spectral data enables the identification of various types of stars and determining their evolutionary stages.
- discovery of new classes of galaxies: analyzing astronomical catalogs using knowledge discovery algorithms can lead to the discovery of new types of galaxies, such as galaxies with unusual shapes or structures, which require further investigation and explanation.
- prediction of gamma-ray bursts: time series methods and statistical analysis can be employed to forecast temporal changes in the activity of gamma-ray bursts, enabling astronomers to prepare for observations and studies of such phenomena.
- identification of gravitational lenses: Analyzing large astronomical databases using knowledge discovery algorithms can assist in identifying and classifying gravitational lenses, which is crucial for studying dark matter and furthering our understanding of cosmology.

These examples illustrate the significance of knowledge discovery in databases in astronomy, as it plays a crucial role in scientific research and helps expand our understanding of the Universe.

2.4. Data mining of the reference stars

The uniformity of the standard form of the image of objects [8] is an important factor influencing the subsequent process of identification with the astronomical catalogue [31]. Therefore, it is necessary to conduct an in-depth analysis of literature data to compare methods for preparing images for the identification process itself. Such methods are expected to reduce the shift in the positional coordinates of the frame center between the frames themselves in the series.

For example, classical methods of computer vision [32] and object image recognition [16] are not able to provide the required level of processing speed. These methods require the analysis of all pixels of potential objects to determine their typical shape. However, when the standard form is heterogeneous, objects are confused, which increases the processing and identification time. Methods for estimating image parameters [33] are based on the analysis of only those pixels that potentially belong to the object under study. Their disadvantage is the inability to determine specific pixels and reject those whose intensity exceeds a specified limit value initially accurately.

In the study [34], the authors use automatic selection of a reference point to select calibration frames. However, this is not a requirement for the identification process itself. Because if there are artifacts in the image, these control points may be false. Thus, the accuracy of identification with real objects from the astronomical catalog decreases. The works [35] propose segmentation method. However, it only work with single images of objects. That is, in the case of a variety of standard shapes (stroke, extended, circular), this method will not provide the necessary accuracy due to the ambiguity in the number of brightness peaks.

This variety of typical shapes also influences various methods of Wavelet transform [36] and time series analysis [37]. The disadvantage of these methods is that they can only work with "pure" measurements, so image heterogeneity will greatly spoil the overall indicator. Another implementation is presented in the study [38] in the form of an additional calibration procedure to avoid the internal coma of the telescope's secondary mirror. But, to equalize brightness and remove "highlights", there is a brightness method that is more improved in accuracy and quality using an inverse median filter [39]. However, the disadvantage of these implementations is the poor accuracy of positional coordinate estimates during the process of identification between frames of the same series.

The matched filtering procedure is also known [40], but it uses only an analytical image model. The disadvantage of this procedure is the inaccuracy of identification when the typical image of an object is different in different frames of the series. The classical method of adding frames [41, 42] to improve the "super" frame is also ineffective in the case when the SSO image does not have clear boundaries on all digital frames of the series. Therefore, it is necessary to develop the mathematical computational methods for automatic selection of reference stars in astronomical images, which will take into account the peculiarities of digital frame formation.

3. Methods

3.1. Determining an estimate of the equatorial coordinates of astronomical objects in CCD-frame

The preliminary identification procedure [43] allows us to obtain linear plate constants $(a_{pl1}, b_{pl1}, c_{pl1})$ and $(a_{pl2}, b_{pl2}, c_{pl2})$, which will determine the relationship between the coordinate system (CS) of the CCD-frame and the tangential (ideal) coordinate system of CCD-frame:

$$\begin{cases} \xi = a_{pl1} \cdot x + b_{pl1} \cdot y + c_{pl1}; \\ \eta = a_{pl2} \cdot x + b_{pl2} \cdot y + c_{pl2}, \end{cases} \tag{1}$$

where ξ and η – ideal (tangential) coordinates of reference stars;

x, y – measured coordinates of reference stars in the coordinate system of CCD-frame.

The calculated linear constants of the plate allow us to obtain estimates of the equatorial coordinates of objects in the frame using the following expression:

$$\begin{cases} \alpha = \alpha_{00} + arctg\left(\dfrac{-\xi}{\cos\delta_{00} - \eta\sin\delta_{00}}\right); \\ \delta = \arcsin\dfrac{\eta\cos\delta_{00} + \sin\delta_{00}}{\sqrt{1+\xi^2+\eta^2}}, \end{cases} \tag{2}$$

where α_{00}, δ_{00} – equatorial coordinates of the optical center of the CCD-matrix.

In the final conversion from CCD-frame coordinates to equatorial coordinates, a cubic model of plate constants is used, which ensures reliable identification and measurement of positions throughout the frame.

3.2. Uniform distribution of candidates for the reference stars in astronomical CCD-frame

Practice shows that the concentration of bright measurements in a certain area of the CCD-frame (for example, in the center) can increase the identification accuracy in this area by reducing it in other areas of the same CCD-frame (Figure 3, left).

To ensure almost equal accuracy of object coordinate measurements throughout the entire CCD-frame, it is advisable to distribute candidates for reference stars evenly throughout the frame.

Thus, a uniform distribution of identified pairs throughout the entire CCD-frame will ensure the necessary uniformity in the accuracy of determining the equatorial coordinates of objects throughout the entire CCD-frame (Figure 3, right).

Therefore, it is necessary to fragment the frame into $M_{reg} \times M_{reg}$ areas (sections) of equal area for uniform distribution of identified pairs on the CCD-frame when selecting candidates for reference stars. In each frame fragment, the same number of objects with a bright image (stars) is selected.

The number of measurements of the frame N_{mea} and stars from the forms of the astronomical catalog N_{st} obtained during observations and intra-frame processing is divided by the number of frame sections.

Next, in each such area $N_{mea}\ /\ M_{reg}^2$ of the brightest measurements of the frame and $N_{st}\ /\ M_{reg}^2$ of the brightest stars in the catalog are selected.

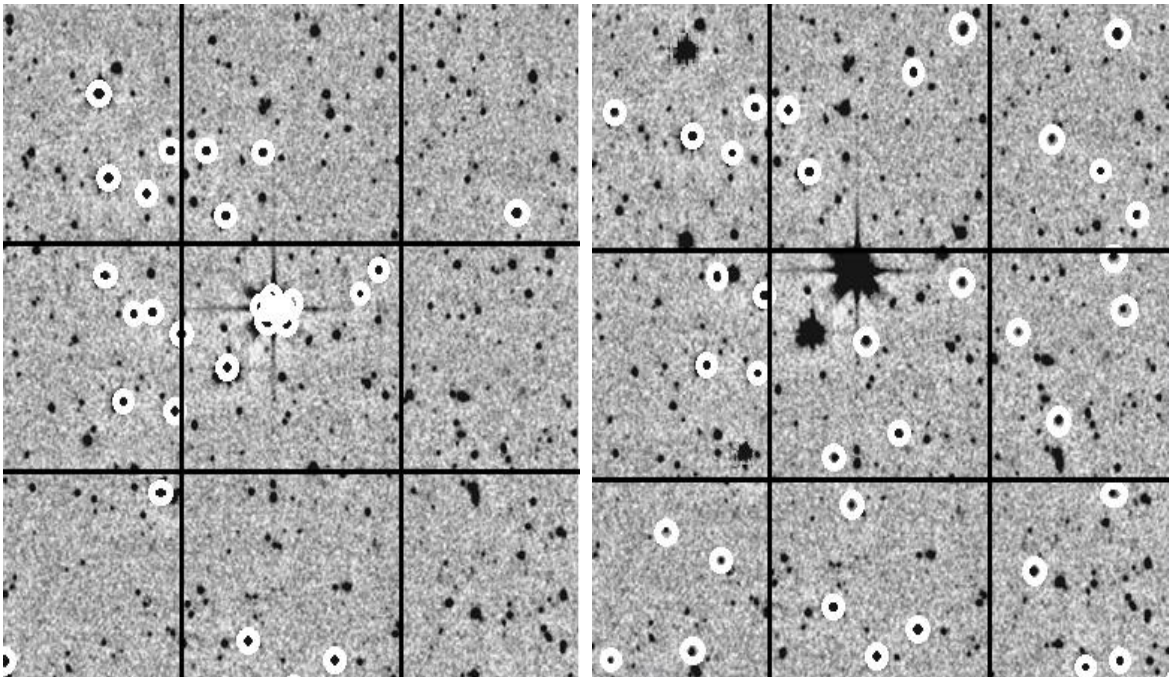

Figure 3: Astronomical frame: left – brightest measurements; right – uniform distribution of guide stars

3.3. Mathematical method for automated selection of the reference stars in astronomical CCD-frame

At each stage of selecting guide stars, measurements of nearby objects are excluded from the set of candidates. This means that the distance between them does not exceed the previously specified value $r_{(mea\_group)}$. That is, the i-th and m-th measurements of the CCD-frame are excluded from candidates for reference stars if the following condition is met:

$$\sqrt{(x_{meainfr} - x_{meam\,nfr})^2 + (y_{meainfr} - y_{maem\,nfr})^2} \leq r_{mea\_group}, \tag{3}$$

where $x_{meamnfr}$ and $y_{meamnfr}$ are the positional coordinates of the measurement of a nearby object in the SC of CCD-frame.

Like (3), measurements of nearby stars from clusters/compact groups of stars in the astronomical catalog are excluded from consideration. This is the case when a nearby object has a comparable or greater brilliance. The criterion for such membership is the presence of a nearby star at a distance less than a predetermined value $r_{(mea\_group)}$:

$$\sqrt{(\alpha_{catj} - \alpha_{cat\ell})^2 + (\delta_{catj} - \delta_{cat\ell})^2} \leq r_{star\_group}, \tag{4}$$

where α_{cat} and δ_{cat} – positional coordinates on the celestial sphere in the astronomical catalog form.

Another important criterion for rejecting candidates is the absence of a brightness peak in the image of the object on the CCD-frame. The criterion for such absence of a peak can be considered the approximate equality of the brightness of the potential peak and the brightness of the pixels A_{ik} from the region Ω_{peak} of size $C_{peak} \times C_{peak}$ of pixels centered in the potential peak. Approximate equality is the difference between the brightness of the pixels of a potential peak and the brightness of the region by no more than N_{Apeak} brightness units.

$$(A_{peak} - A_{ik}) \leq N_{Apeak}, \text{ for } \forall\, i,k \in \Omega_{peak} \tag{5}$$

In this way, sets of measurements are formed from the side of the CCD-frame and the astronomical catalog, which in the subsequent stage will take part in vaporization and testing of hypotheses about the correspondence of the “measurement-formula” identification pair.

3.4. Data mining process of the automated reference stars selection

The architecture of data mining process of the automated reference stars selection includes the following sequence of operations (Figure 4).

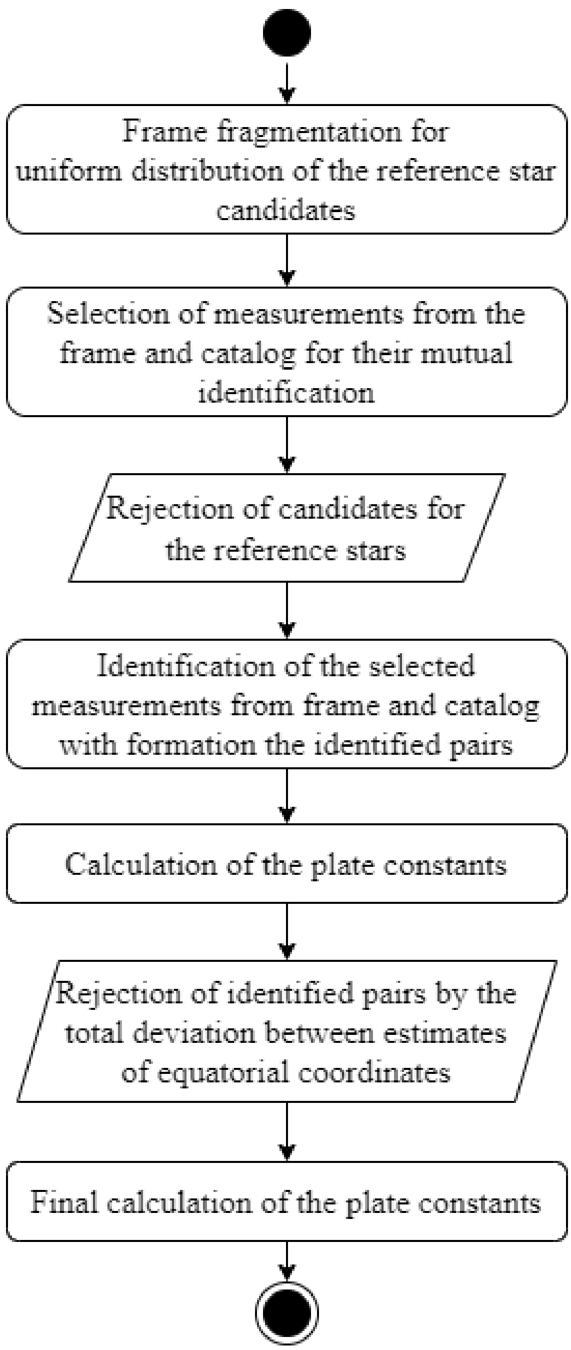

Figure 4: Architecture of data mining process of the automated reference stars selection

1. Calculation of linear plate constants $(a_{pl1}, b_{pl1}, c_{pl1})$ and $(a_{pl2}, b_{pl2}, c_{pl2})$ (1).
2. Obtaining an estimate of the equatorial coordinates of objects (2).

3. Fragmentation of the CCD-frame into a set of $M_{reg} \times M_{reg}$ areas of equal area for uniform distribution of candidates for reference stars.
4. At each stage of the method, a sequence of operations is performed.
   4.1. Selection of sets of measurements of the CCD-frame and catalog forms for their mutual identification.
      4.1.1. At the first stage, N_{mea}/M_{reg}^2 of the brightest measurements of the CCD-frame are selected in each section of the CCD-frame. N_{st}/M_{reg}^2 of the brightest stars in the catalog of the corresponding part of the celestial sphere are also selected.
      4.1.2. At the second and third stages, in each section of the CCD-frame, the next $\Delta N_{mea}/M_{reg}^2$ and $\Delta N_{st}/M_{reg}^2$ of the brightest measurements of the CCD-frame and catalog forms, respectively, are additionally selected.
   4.2. Rejection of selected candidates for guide stars is performed.
      4.2.1. Rejection of measurements of objects close to each other (3).
      4.2.2. Rejection of measurements of stars in the astronomical catalog if they belong to clusters or compact groups of stars (4).
      4.2.3. Rejection of measurements of objects with images without brightness peaks (5).
   4.3. Identification of selected frame measurements and catalog forms with the formation of identified pairs.
   4.4. At each iteration, calculation/refinement of linear constants of the plate with a higher degree model.
   4.5. Rejection of identified pairs based on the total deviation $\Delta_{\alpha\delta ijk}$ between estimates of equatorial coordinates in the identified «measurement-form» pair:

$$\Delta_{\alpha\delta ijk} > K_{rej}\hat{\hat{\Delta}}_{\alpha\delta}, \tag{6}$$

where $\hat{\hat{\Delta}}_{\alpha\delta} = \sqrt{\frac{1}{N_{cou}}((\sum_{k=1}^{N_{cou}} \alpha_{catj(k)} - \alpha_{meainfr(k)})^2 + (\delta_{catj(k)} + \delta_{meainfr(k)})^2)}$

– the average modulus of deviation of an identified pair in equatorial coordinates on the set of selected identified pairs;

K_{rej} – coefficient of the condition for rejecting «measurement-form» pairs from the set;

k – number of the identified «measurement-form» pair;

N_{cou} – number of identified «measurement-formula» pairs used to calculate plate constants.

5. Final calculation of linear constants of the plate.

4. Experiment

The object of study are the images of the Solar System objects (SSO) (like stars, asteroids, comets) and any other space objects (like space robots [44], drones [45], satellites [46]) detected in a series of CCD-frames. The initial series for the study were obtained from a variety of telescopes installed at observatories in Ukraine and around the world. Namely, the ISON-NM observatory, the SANTEL-400AN telescope (New Mexico, USA); Vihorlat Observatory, VNT telescope (Humenne, Slovakia); Odesa-Mayaky Observatory, OMT-800 telescope (Mayaki, Ukraine); Cerro Tololo observatory, PROMPT-8 telescope (La Serena, Chile) [47]. All mentioned above observatories were approved and confirmed by the Minor Planet Center (MPC) as an official organization for the observing and reporting on minor planets or SSOs under the auspices of the International Astronomical Union (IAU) [48].

To verify the developed mathematical computational methods for automatic selection of reference stars in astronomical images, testing was carried out on a series of frames containing 27,352 measurements. Such a total number of measurements was successfully identified with the astronomical catalog.

The USNO B1.0 catalog was used as a photometric catalogue. The catalog contains angular positional coordinates and magnitudes of more than one billion SSOs, formed over 3.6 billion measurements.

When conducting research, the following values of the parameters of the developed methods were assumed:

- the number of the brightest measurements of the CCD-frame was $N_{mea} = 400$;
- the number of the brightest measurements of the catalog forms for selecting candidates for reference stars was $N_{st} = 600$;
- the number of fragments into which the CCD-frame is divided for uniform distribution was $M_{reg} = 4$;
- the number of measurements of the CCD-frame $\Delta N_{mea} = 300$;
- the number of measurements of the catalog form $\Delta N_{st} = 500$ with increasing iteration;
- the criterion for the absence of a peak is the deviation of the brightness of the object image pixels by no more than $N_{Apeak} = 4$ in the region $C_{peak} \times C_{peak}$ ($C_{peak} = 5$) centered at the peak;
- the maximum permissible distance between measurements on a CCD-frame of close group objects was $r_{mea\_group} = 20$ pixels;
- the maximum permissible distance between measurements in the form of catalogs of nearby group objects was $r_{star\_group} = 5$ pixels;
- the coefficient of the rule for rejecting «measurement-formula» pairs from a set of reference stars was considered $K_{rej} = 1$.

The parameters of the procedure listed above were obtained empirically.

The following statistical indicators of the accuracy of reference star measurements were studied: estimates of the average deviation of estimates of equatorial coordinates between the catalog and measured values, $\bar{\Delta}_\alpha, \bar{\Delta}_\delta$; standard deviation (RMS) $\sigma_\alpha, \sigma_\delta, \sigma_m$ and an estimate of the mean deviation of the gloss estimate between the catalog and measured values $\bar{\Delta}_m$.

Histogram of the distribution of deviations of the equatorial coordinate (right ascension (RA)) of reference stars from the brightness and coordinates of objects in the rectangular coordinate system of the CCD-frame is presented in Figure 5.

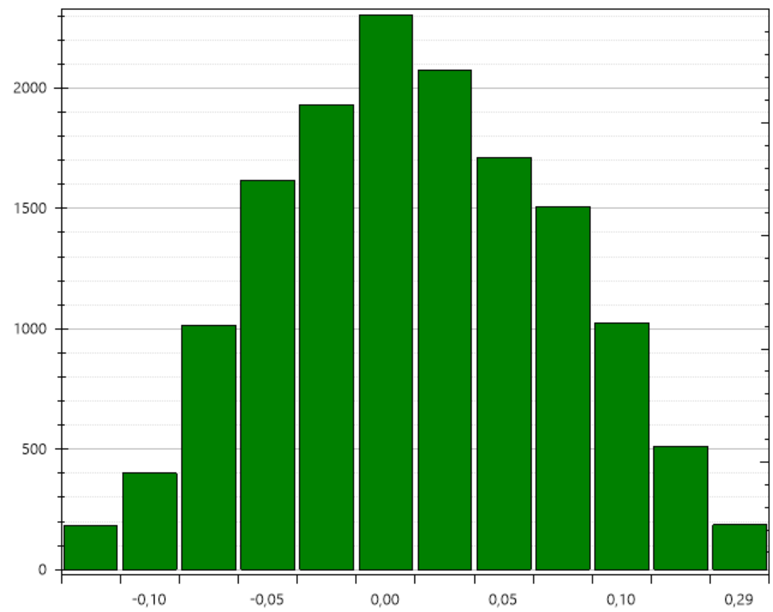

Figure 5: A histogram of distribution of the deviations by RA measurements of reference stars

Histogram of the distribution of deviations of the equatorial coordinates (declination (DE)) of reference stars from the brightness and coordinates of objects in the rectangular coordinate system of the CCD-frame is presented in Figure 6.

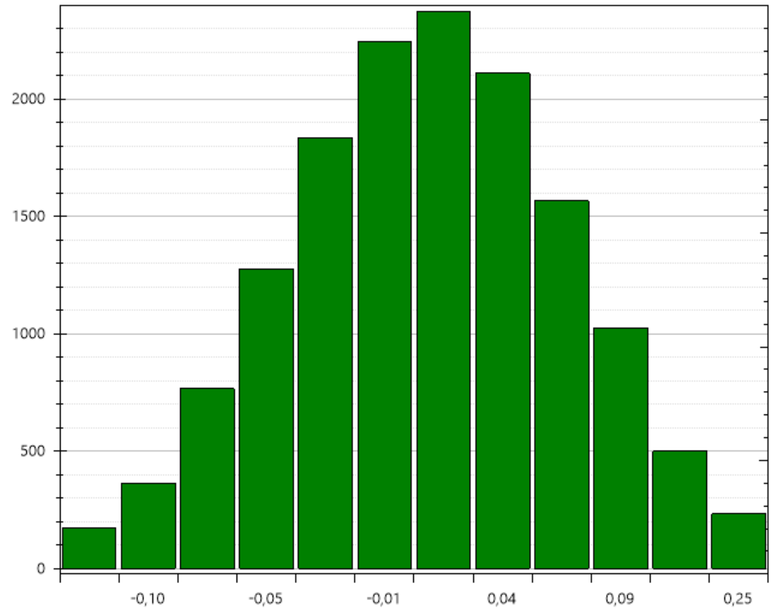

Figure 6: A histogram of distribution of the deviations by DE measurements of reference stars

The received dependence of deviations of equatorial coordinates from the position of reference stars in frame is presented in Figure 7.

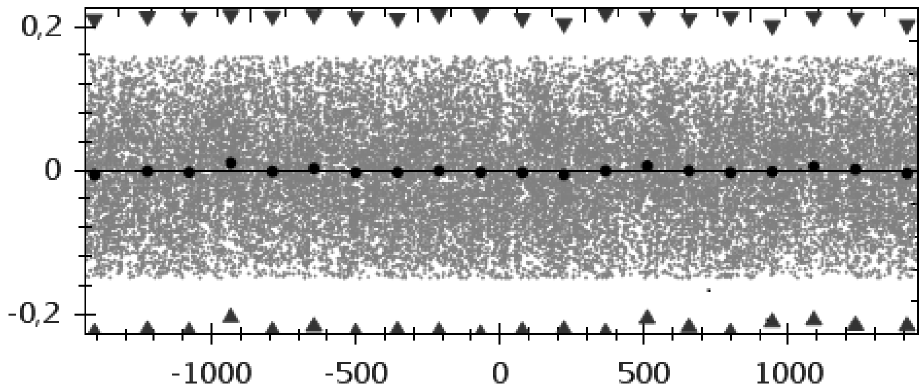

Figure 7: Dependence of deviations of equatorial coordinates from the position of reference stars

The received dependence of deviations of equatorial coordinates from the brightness assessment of objects in frame is presented in Figure 8.

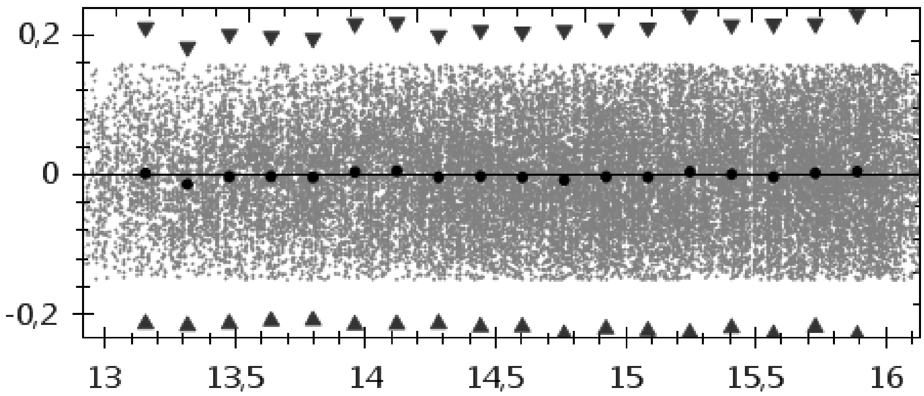

Figure 8: Dependence of deviations of equatorial coordinates from the brightness of objects

The research result based on series of frames with 27,352 measurements is presented in Table 1.

Table 1
Statistical indicators of the accuracy of reference star measurements

Statistical indicator	Value before	Value after
Average deviation RA, arc. sec.	0,014	0,001
Average deviation DE, arc. sec.	0,017	0,001
Average deviations brightness, mag.	0,36	0,03
Max. deflection module RA, ang. sec.	0,95	0,13
Max. deflection module DE, ang. sec.	0,91	0,12
Min. gloss deflection modulus, mag.	0,009	0,001
Max. gloss deflection modulus, mag.	3,13	0,36
RMS deviations according to RA, ang. sec.	0,57	0,07
RMS deviation according to DE, ang. sec.	0,51	0,06
RMS deviations in brightness, mag.	0,75	0,35

Presented in Table 1 indicators show the successful application of the developed methods. The standard deviation of frame identification errors in this case is 5–7 times less than without using the developed methods.

5. Results

Existing methods for basic image processing [41] and computer vision [32] were analyzed. However, the speed and accuracy of identification by such methods directly depends on the characteristics of the formation of a series of digital frames. There is also a dependence on the completeness of the astronomical catalog with data and on the constancy of the typical image [8] of the object in all frames of the series. Therefore, to develop the methods for automated data mining of the reference stars from astronomical CCD-frames and certain rules and criteria for rejecting candidates at each iteration were proposed.

The obtained research results, as well as the developed mathematical computational methods for automatic selection of reference stars in astronomical images, were implemented in the C++ programming language. This code was implemented at the stage of intra-frame processing of the Lemur software package (Ukraine) [49] for the automated detection of new and maintenance of known objects within the CoLiTec project [50]. The developed mathematical computational methods, implemented in Lemur software (Ukraine), was used during the successful identification of CCD-frames, which contained a total of more than 800,000 SSOs. Their measurements were also successfully identified with known astronomical catalogs.

Obtained in Table 1, the results are determined by the uniform distribution of candidates for reference stars, as well as correctly selected conditions and rejection criteria. It clearly indicates that the assigned tasks have been successfully completed. The research showed that the usage of the developed methods reduces identification errors with cataloged (reference) objects by 5–7 times. This significantly affects the quality and accuracy of a few tasks for detecting the trajectories of objects.

6. Conclusions

In recent decades, astronomy has undergone a data processing and analysis revolution due to the implementation of large astronomical projects and the use of advanced observational tools. These modern instruments collect and process vast amounts of data, extracting valuable information about celestial objects and phenomena. In astronomical images obtained using telescopes and cameras, there are from 1 to 100 thousand or more stars depending on the resolution and exposure time. These objects are fixed against the background of the frame and have constant positions in the celestial sphere. To determine which part of the sky corresponds more accurately to a given frame, it is necessary to associate the frame with known astronomical astrometric and photometric catalogs.

These catalogs contain millions of position values of various stars as static objects. Having such information in the form of big data, as well as a huge amount of classified and clustered data in the form of databases, computational methods for fast extraction of the necessary data from them need to be developed. For this purpose, classical methods of "knowledge discovery in databases" (KDD) and Data Mining exist. However, for their proper application, it is necessary to classify the input data set for subsequent analysis and rejection. The implementation of these methods is closely related to the developed mathematical computational methods for automatic selection of reference stars in astronomical images.

We presented the developed Lemur software of the CoLiTec (Collection Light Technology) project, which is implemented as a client-server application for the processing of astronomical data using the data mining and KDD methods. As described in this article the KDD with the data mining step is very useful for the data optimization to receive only the helpful data with reference stars.

Acknowledgements

The research was performed with help of all observatories, useful tools and observers who provided astronomical data for testing the developed Lemur software with implementation of the data mining methods. The research was supported by the Ukrainian project of fundamental scientific research “Development of computational methods for detecting objects with near-zero and locally constant motion by optical-electronic devices” #347 in 2024-2026 years.

References

[1] F. Chierchie, et al., Detailed modeling of the video signal and optimal readout of charge-coupled devices, International Journal of Circuit Theory and Applications, vol. 48, issue 7, pp. 1001-1016, 2020. doi: 10.1002/cta.2784.

[2] D. J. Hoover, D. Z. Seligman, and M. J. Payne, The Population of Interstellar Objects Detectable with the LSST and Accessible for In Situ Rendezvous with Various Mission Designs, The Planetary Science Journal, vol. 3, issue 3, p. 71, 2022. doi: 10.3847/psj/ac58fe.

[3] D. Peralta, S. del Rio, S. Ramirez-Gallego, I. Triguero, J. Benitez, and F. Herrera, Evolutionary feature selection for big data classification: A map reduce approach, Mathematical Problems in Engineering, vol. 2015, 246139, pp. 11, 2015. doi: 10.1155/2015/246139.

[4] M., Khalil, et al., Big data in astronomy: from evolution to revolution, International Journal of Advanced Astronomy, vol. 7, issue 1, 2021. doi: 10.14419/ijaa.v7i1.18029.

[5] D. Oszkiewicz, et al., Spins and shapes of basaltic asteroids and the missing mantle problem, Icarus, vol. 397, 115520, 2023. doi: 10.1016/j.icarus.2023.115520.

[6] V. Troianskyi, V. Kashuba, O. Bazyey, et al., First reported observation of asteroids 2017 AB8, 2017 QX33, and 2017 RV12, Contributions of the Astronomical Observatory Skalnaté Pleso, vol. 53, pp. 5-15, 2023. doi: 10.31577/caosp.2023.53.2.5.

[7] V. Troianskyi, P. Kankiewicz, and D. Oszkiewicz, Dynamical evolution of basaltic asteroids outside the Vesta family in the inner main belt, Astronomy and Astrophysics, vol. 672, A97, 2023. doi: 10.1051/0004-6361/202245678.

[8] V. Savanevych, et al., Formation of a typical form of an object image in a series of digital frames, Eastern-European Journal of Enterprise Technologies, vol. 6, issue 2-120, pp. 51–59, 2022. doi: 10.15587/1729-4061.2022.266988.

[9] V. Akhmetov, et al., Fast coordinate cross-match tool for large astronomical catalogue, Advances in Intelligent Systems and Computing, vol. 871, pp. 3–16, 2019. doi: 10.1007/978-3-030-01069-0_1.

[10] H. Yang, et al., Data mining techniques on astronomical spectra data–II. Classification analysis, Monthly Notices of the Royal Astronomical Society, vol. 518, issue 4, pp. 5904-5928, 2023. doi: 10.1093/mnras/stac3292.

[11] V. Savanevych, et al., Selection of the reference stars for astrometric reduction of CCD-frames, Advances in Intelligent Systems and Computing, vol. 1080, pp. 881–895, 2020. doi: 10.1007/978-3-030-33695-0_57.

[12] F. Genova, Data as a research infrastructure CDS, the Virtual Observatory, astronomy, and beyond, EPJ Web of Conferences, vol. 186, EDP Sciences, 2018. doi: 10.1051/epjconf/201818601001.

[13] P. Hasan, and S. N. Hasan, Astronomy data, virtual observatory and education, Proceedings of the International Astronomical Union, vol. 15, issue S367, pp. 151-154, 2019. doi: 10.1017/S174392132100034X.

[14] D. Oszkiewicz, et al., Spin rates of V-type asteroids, Astronomy and Astrophysics, vol. 643, A117, 2023. doi: 10.1051/0004-6361/202038062.

[15] V. Akhmetov, et al., New approach for pixelization of big astronomical data for machine vision purpose, IEEE International Symposium on Industrial Electronics, pp. 1706–1710, 2019. doi: 10.1109/ISIE.2019.8781270.

[16] S. Khlamov, I. Tabakova, T. Trunova, Recognition of the astronomical images using the Sobel filter, Proceedings of the 29th IEEE IWSSIP 2022, Sofia, Bulgaria, June 1st – 3rd, 4 p., 2022. doi: 10.1109/IWSSIP55020.2022.9854425.

[17] M. K. Cavanagh, K. Bekki, and B. A. Groves, Morphological classification of galaxies with deep learning: comparing 3-way and 4-way CNNs, Monthly Notices of the Royal Astronomical Society, vol. 506, issue 1, pp. 659-676, 2021. doi: 10.1093/mnras/stab1552.

[18] L. Mykhailova, et al., Method of maximum likelihood estimation of compact group objects location on CCD-frame, Eastern-European Journal of Enterprise Technologies, vol. 5, issue 4, pp. 16-22, 2014. doi: 10.15587/1729-4061.2014.28028.

[19] Š. Parimucha, et al., CoLiTecVS – A new tool for an automated reduction of photometric observations, Contributions of the Astronomical Observatory Skalnate Pleso, vol. 49, issue 2, pp. 151-153, 2019. doi: 2019CoSka..49..151P.

[20] S. Khlamov, et al., Development of the matched filtration of a blurred digital image using its typical form, Eastern-European Journal of Enterprise Technologies, vol. 1, issue 9-121, pp. 62–71, 2023. doi: 10.15587/1729-4061.2023.273674.

[21] H. E. Bond, et al., Hubble Space Telescope Imaging of Luminous Extragalactic Infrared Transients and Variables from the Spitzer Infrared Intensive Transients Survey, The Astrophysical Journal, vol. 928, issue 2, p. 158, 2022. doi: 10.3847/1538-4357/ac5832.

[22] J. Bennett, S. Shostak, N. Schneider, and M. MacGregor, Life in the Universe. Princeton University Press, 2022.

[23] D. S. Madgwick, Correlating galaxy morphologies and spectra in the 2dF Galaxy Redshift Survey, MNRAS, vol. 338, issue 1, pp. 197-207, 2003. doi: 10.1046/j.1365-8711.2003.06033.x.

[24] Ž. Ivezić, et al., Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data, Princeton University Press, 2019.

[25] R. Mor, et al., Expanding Big Data mining for Astronomy, XIV Scientific Meeting of the Spanish Astronomical Society, p. 235, 2020. doi: 2020sea..confE.235M.

[26] Y. Zhang, and Y. Zhao, Astronomy in the big data era, Data Science Journal, vol. 14, 2015.

[27] S. Khlamov, V. Savanevych, I. Tabakova, and T. Trunova, The astronomical object recognition and its near-zero motion detection in series of images by in situ modeling, Proceedings of the 29th IEEE IWSSIP 2022. doi: 10.1109/IWSSIP55020.2022.9854475.

[28] C. J. Fluke, and C. Jacobs, Surveying the reach and maturity of machine learning and artificial intelligence in astronomy, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 10, issue 2: e1349, 2020. doi: 10.1002/widm.1349.

[29] V. Shvedun, et al., Statistical modelling for determination of perspective number of advertising legislation violations, Actual Problems of Economics, vol. 184, issue 10, pp. 389-396, 2016.

[30] W. P. McCray, The biggest data of all: Making and sharing a digital universe, Osiris, vol. 32, issue 1, pp. 243-263, 2017. doi: 10.1086/693912.

[31] I. Vavilova, et al., Surveys, catalogues, databases, and archives of astronomical data, Knowledge Discovery in Big Data from Astronomy and Earth Observation: Elsevier, pp. 57-102, 2020. doi: 10.1016/B978-0-12-819154-5.00015-1.

[32] M. Bellanger, Digital Signal Processing: Theory and Practice. Wiley, 2024.

[33] V. Akhmetov, et al., Astrometric reduction of the wide-field images, Advances in Intelligent Systems and Computing, vol. 1080, pp. 896–909, 2020. doi: 10.1007/978-3-030-33695-0_58.

[34] M. Lösler, C. Eschelbach, and S. Riepl, A modified approach for automated reference point determination of SLR and VLBI telescopes: First investigations at Satellite Observing System Wettzell, Technisches Messen, vol. 85, pp. 616–626, 2018. doi: 10.1515/teme-2018-0053.

[35] S. Minaee, Y. Y. Boykov, F. Porikli, et al., Image Segmentation Using Deep Learning: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, issue 7, pp. 3523-3542, 2021. doi: 10.1109/TPAMI.2021.3059968.

[36] M. Dadkhah, V. V. Lyashenko, Z. V. Deineko, et al., Methodology of wavelet analysis in research of dynamics of phishing attacks. International Journal of Advanced Intelligence Paradigms, vol. 12, issue 3-4, pp. 220-238, 2019. doi: 10.1504/IJAIP.2019.098561.

[37] L. Kirichenko, A.S.A. Alghawli, T. Radivilova, Generalized approach to analysis of multifractal properties from short time series, International Journal of Advanced Computer Science and Applications, vol. 11, issue 5, pp. 183–198, 2020. doi: 10.14569/IJACSA.2020.0110527.

[38] K. Hampson, et al., High precision automated alignment procedure for two-mirror telescopes. Applied Optics, vol. 58, pp. 7388–7391, 2019. doi: 10.1364/AO.58.007388.

[39] I. Kudzej, et al., CoLiTecVS – A new tool for the automated reduction of photometric observations, Astronomische Nachrichten, vol. 340, pp. 68–70, 2019. doi: 10.1002/asna.201913562.

[40] S. Khlamov, et al., Development of computational method for matched filtration with analytic profile of the blurred digital image, Eastern-European Journal of Enterprise Technologies, vol. 5, issue 4-119, pp. 24–32, 2022. doi: 10.15587/1729-4061.2022.265309.

[41] W. Burger, and M. Burge, Digital image processing: an algorithmic introduction, Springer Nature, 945 p., 2022. doi: 10.1007/978-3-031-05744-1.

[42] Y. Bodyanskiy, S. Popov, F. Brodetskyi, O. Chala, Adaptive Least-Squares Support Vector Machine and its Combined Learning-Selflearning in Image Recognition Task, IEEE 17th International Scientific and Technical Conference on Computer Sciences and Information Technologies, pp. 48–51, 2022. doi: 10.1109/CSIT56902.2022.10000518.

[43] V. Savanevych, et al., Mathematical methods for an accurate navigation of the robotic telescopes, Mathematics, vol. 11, issue 10, 2246, 2023. doi: 10.3390/math11102246.

[44] A. Tantsiura, et al., Evaluation of the potential accuracy of correlation extreme navigation systems of low-altitude mobile robots, International Journal of Advanced Trends in Computer Science and Engineering, vol. 8, issue 5, pp. 2161–2166, 2019. doi: 10.30534/ijatcse/2019/47852019.

[45] N. Yeromina, V. Tarshyn, S. Petrov, et al., Method of reference image selection to provide high-speed aircraft navigation under conditions of rapid change of flight trajectory, International Journal of Advanced Technology and Engineering Exploration, vol. 8, issue 85, pp. 1621–1638, 2021. doi: 10.19101/IJATEE.2021.874814.

[46] V. Akhmetov, et al., Cloud computing analysis of Indian ASAT test on March 27, 2019, in: IEEE PIC S and T, pp. 315–318, 2019. doi: 10.1109/picst47496.2019.9061243.

[47] T. Li, D. DePoy, J. Marshall, et al., Monitoring the atmospheric throughput at Cerro Tololo Inter-American Observatory with aTmCam. Ground-based and Airborne Instrumentation for Astronomy V, vol. 9147, pp. 2194–2205, 2014. doi: 10.1117/12.2055167.

[48] Minor Planet Center, List Of Observatory Codes. Available at: https://www.minorplanetcenter.net/iau/lists/ObsCodesF.html.

[49] Lemur software, CoLiTec project. Available at: https://colitec.space.

[50] S. Khlamov, V. Savanevych, O. Briukhovetskyi, and A. Pohorelov, CoLiTec software - Detection of the near-zero apparent motion, Proceedings of the International Astronomical Union, vol. 12, issue S325, pp. 349-352, 2016. doi: 10.1017/S1743921316012539.