UDC 520.88
IRSTI 41.51.41

https://doi.org/10.55452/1998-6688-2026-23-3-461-472

[1,2]Gluchshenko A.,
MSc, ORCID ID: 0000-0002-0738-7725,
e-mail: gluchshenko@fai.kz
[1]Izmailova I.,
MSc, ORCID ID: 0000-0001-9878-0989,
e-mail: izmailova@fai.kz
[1,3]Umirbayeva A.,
MSc, ORCID ID: 0000-0001-9339-4990,
e-mail: umirbayeva@fai.kz
[1,2]Kuvatova D.,
MSc, ORCID ID: 0000-0002-5937-4985,
e-mail: kuvatova@fai.kz
[1]Yurin D.,
PhD, ORCID ID: 0000-0002-5604-9757,
e-mail: yurin@fai.kz

[1]Fesenkov Astrophysical Institute, Almaty, Kazakhstan
[2]Farabi University, Almaty, Kazakhstan
[3]Nazarbayev University, Astana, Kazakhstan

AUTOMATED WAVELENGTH CALIBRATION OF ASTRONOMICAL SPECTRA WITH DTW-RANSAC MATCHING AND CNN-BASED PREPROCESSING

Abstract

This paper presents an automated pipeline for wavelength calibration of slit-based astronomical spectra, combining robust peak-matching techniques with a lightweight preprocessing step. The method is based on initial peak alignment using Dynamic Time Warping (DTW), followed by refinement of the dispersion solution with the Random Sample Consensus (RANSAC) algorithm. To ensure consistency between observed and reference spectra, a compact one-dimensional convolutional neural network (CNN) is employed at the preprocessing stage to identify the spectral range and corresponding diffraction grating. This step enables the selection of an appropriate reference spectrum but does not directly affect the calibration procedure. The DTW-based approach provides reliable initial correspondences between spectral peaks under nonlinear distortions and noise, while RANSAC filtering removes incorrect matches and yields a robust polynomial dispersion solution. The method is validated on real observational data from calibration lamps and demonstrates stable performance, high matching accuracy, and full automation capability. Compared to traditional manual or semi-automatic approaches, the proposed pipeline significantly reduces processing time while maintaining calibration precision. The developed solution is planned to be deployed as a service in the virtual observatory infrastructure and is applicable for processing numerous spectral data.

Keywords: astronomical data processing, slit-based spectral calibration, dynamic time warping, RANSAC, machine learning.

Received May 15, 2026; revised August 27, 2026; accepted August 28, 2026.

Introduction

Spectral studies play a fundamental role in modern astrophysics, as it is spectroscopy that enables us to determine the chemical composition, physical parameters, kinematics, and evolutionary state of astronomical objects. With the development of virtual observatories, there is a growing need for

automated methods of processing spectroscopic data. The Fesenkov Astrophysical Institute (FAI) holds a significant volume of raw observational data, including both contemporary and archived material. This makes the development of methods ensuring consistent and reliable spectral reduction particularly relevant. In particular, the automation of wavelength calibration represents an important task, linked both to improving the efficiency of processing new observations and to the possibility of re-analysing previously accumulated data.

Despite its importance, the practical implementation of these tasks faces significant operational hurdles. Traditional reduction methods (e.g., within the IRAF environment) create a critical bottleneck: during a typical monthly observing cycle at the AZT-20 telescope, which yields approximately 40 raw frames (with about 10 requiring precise calibration), manual processing takes roughly 15 minutes per frame. Cumulatively, data preparation for even a single instrument consumes substantial expert time that could otherwise be dedicated to physical interpretation. Given the extensive data archives and the planned expansion of the telescope network at the Academician T.B. Omarov Assy-Turgen Observatory (ATO), the total data volume is projected to increase severalfold. Such scaling renders manual intervention fundamentally unscalable.

Wavelength calibration ensures the conversion of detector pixel coordinates into a physical wavelength scale, typically based on identifying emission lines from calibration lamps and approximating a dispersion solution. However, manual line matching remains a labour-intensive and error-prone process, especially in the presence of noise, incomplete line sets, or nonlinear distortions of the wavelength scale. To automate calibration, template matching and cross-correlation methods are widely used in modern spectroscopic surveys [1, 2]. Nevertheless, fully automated calibration remains challenging, particularly under realistic observational conditions where distortions and noise can significantly affect the matching of spectral features. For example, the IRAF package [3] provides automated modes, but initial setup and quality control still frequently require observer intervention, particularly in cases of low signal-to-noise ratio or complex dispersive distortions. While machine learning has been explored for spectral classification and parameter estimation [4–10], the core calibration procedure still largely relies on establishing manual correspondences to ensure a stable solution.

Recent work has demonstrated the applicability of DTW to automated spectroscopic wavelength calibration. Davenport et al. [11] used DTW for first-pass alignment between observed and reference spectra, followed by refinement using conventional emission-line fitting. In contrast, in the present work DTW is used to generate candidate peak correspondences based on local spectral similarity, which are subsequently filtered using RANSAC to obtain a robust polynomial dispersion solution.

In this work, we develop an automated pipeline for wavelength calibration that addresses these limitations. The method is based on a convolutional neural network (CNN) [12], which is used at the preprocessing stage to determine the spectral range, followed by two sequential methods: initial alignment using the Dynamic Time Warping (DTW) algorithm [11, 13] and subsequent refinement using the Random Sample Consensus (RANSAC) algorithm [14].

Materials and methods

Data Description and Automatic Determination of Spectral Range

The observational data used to develop and test our algorithm (https://github.com/fai-kz/ransac_wv_calib) were primarily acquired with the AZT-20 telescope at FAI's ATO, which provides the main source of slit spectroscopic observations. The data were obtained using a He-Ne-Ar calibration lamp in the H_α, H_β, and combined H_α/H_β spectral ranges. The proposed pipeline uses a user-provided, previously wavelength-calibrated spectrum as the reference. In the present implementation, the reference spectrum used for testing was calibrated with IRAF. Depending on the wavelength coverage, diffraction gratings with line densities of 360 lines/mm for the combined H_α/H_β range, 1800 lines/mm for H_α, and 2400 lines/mm for H_β were used. To improve the robustness of the method and extend the coverage of input conditions, the dataset was supplemented with synthetic spectra.

The first step in processing a spectral frame is the automatic identification of the type of diffraction grating used to obtain the spectrum. This information is required to correctly define the operating wavelength range and select the parameters for further calibration. To automate this procedure, a neural model has been developed in the form of a one-dimensional CNN. Given the frequent absence of wavelength range in FITS headers, CNN-based approach reconstructs the observational context directly from the raw data, automatically identifying the grating configuration through spectral morphology to select the appropriate calibration template.

The network architecture consists of three convolutional layers with decreasing kernel sizes (50, 30 and 14) and an increasing number of channels (8, 16 and 32), allowing both global and local spectral features to be extracted. Batch normalisation is applied after each convolution, and non-linearity is introduced using the ReLU activation function [15]. The extracted features are then passed to a fully connected layer that outputs class probabilities for the corresponding grating type.

The model was trained using synthetic spectra generated from real observational data. Synthetic samples were produced by applying inverse polynomial calibration to pre-calibrated spectra, with random variation of the calibration coefficients under the constraint of a positive derivative over the calibration interval. Gaussian noise was added to improve robustness. Approximately 5,000 synthetically generated spectra across the three classes (360, 1800, and 2400 lines/mm) were used for training. The synthetic spectra for each class were generated using one representative observational frame obtained with the corresponding grating configuration; these observational frames were not included in the training dataset. The dataset was divided into training, validation, and test subsets in proportions of 70%, 15%, and 15%, respectively. Training and data generation were implemented using PyTorch, with parallelisation of synthetic data generation via mpi4py [16], enabling efficient preparation of large training sets. The model was trained for 100 epochs with a batch size of 32 using the Adam optimiser [17] and cross-entropy loss. The initial learning rate was set to 0.001 and varied during training. Dropout regularisation with a rate of 0.3 was applied to reduce overfitting [18]. The input to the network consisted of a single-channel spectrum of 1024 samples. The CNN comprised three convolutional layers with 8, 16, and 32 output channels, kernel sizes of 50, 30, and 14, and padding values of 25, 15, and 7, respectively. Each convolutional layer was followed by batch normalisation and a ReLU activation. The final linear layer mapped the resulting (32x128) features to the three output classes.

The CNN is used exclusively at the preprocessing stage to determine the spectral range, after which the calibration proceeds using peak matching and model fitting methods. As shown in Figure 1, the network achieves high classification accuracy in distinguishing among the three grating configurations.

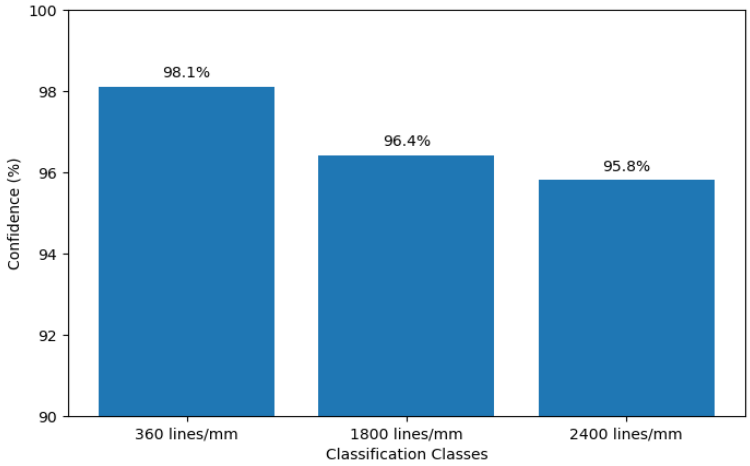

Figure 1 – Neural network classification accuracy across 3 classes

Initial Peak Matching Using DTW-RANSAC

Spectral calibration aims to establish a relationship between the coordinates of the detector pixels and the physical wavelength values. To achieve this, the spectrum of the calibration lamp is compared with the reference spectrum of the corresponding source from a spectral atlas.

The alignment is performed by comparing the recorded spectrum of the lamp with a reference set of lines for that lamp (the so-called spectral atlas (https://noirlab.edu/science/data-services/other/spectral-atlas)). The traditional method is based on the use of reference lines – characteristic emission maxima that are clearly distinguishable in the spectrum. For each such line, a correspondence is determined between its position on the detector (in pixels) and the known wavelength from the atlas. The resulting set of corresponding points is used to construct a dispersion solution, which is approximated by a polynomial function:

$$\lambda(p) = \sum_{n=0}^{k} a^n p_n, \quad (1)$$

where λ is the wavelength, a^n is the coefficient of the n^{th} power, p_n is the pixel number, and k is the order of the polynomial used for the approximation. In automated data processing, the correct matching of spectral lines presents a non-trivial challenge. The algorithm must remain robust against various observational artefacts, including noise, missing real lines, the appearance of false maxima, significant non-linear distortions of the wavelength scale, and other effects. Nevertheless, when successfully implemented, automatic calibration significantly speeds up data processing and reduces the likelihood of human-induced errors.

In manual mode, aligning a single spectrum takes on average around fifteen minutes. In the same amount of time, an automated system is capable of aligning approximately 4 to 20 frames, depending on computational resources and the level of noise in the data. The efficiency of the method in terms of time can be assessed using Figure 2.

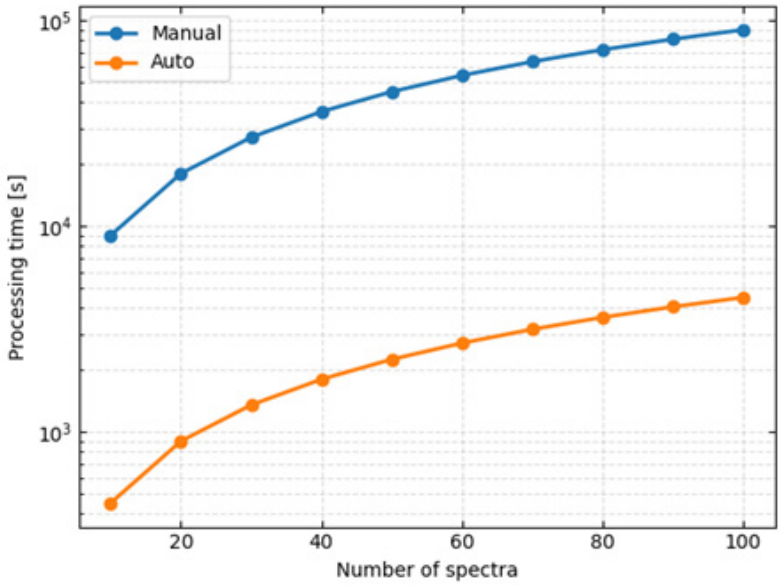

Figure 2 – Comparison of the time spent on processing spectra in the case of the traditional method and the proposed method

In the course of seeking solutions to improve alignment accuracy, it was found that the use of standard spectral atlases for calibration lamps is not always effective. The operational observation ranges are generally limited (e.g., ~6200-7000 Å for H_α, ~4600-5200 Å for H_β and ~3000-7700 Å for H_α/H_β), whereas atlases cover a much wider range (~3000-10600 Å). Furthermore, observational data differ significantly in terms of spectral resolution: their resolution is approximately 0.5-4.5 Å/pixel, depending on the grating used, whereas the atlases are presented at a high resolution of around ~0.02 Å.

Additional discrepancies arise due to observational effects – noise, overexposure and underexposure – as well as local non-linearities in the dispersion solution (compression and stretching of the wavelength scale). These factors exacerbate the discrepancy between the observed spectra and the atlas data. Consequently, the use of real, pre-calibrated observations proved to be a more effective basis for developing and testing the algorithm.

Let us introduce the following notation:

S is the spectrum to be calibrated, represented as a pixel-intensity function;

R is the spectrum used to calibrate spectrum S, represented as a "wavelength – intensity" relationship and treated as a reference spectrum;

P_S^m, P_R^m, where $n = 0, 1, \ldots, N$, $m = 0, 1, \ldots, M$ – peaks detected in spectra S and R, respectively;

Model is a polynomial of the form (1) that defines the dispersion solution and determines the calibration of spectrum S.

Before the algorithm begins, the spectra S and R are normalized to the dimensionless range [0,1] along both the abscissa (pixel/wavelength) and ordinate (intensity) axes. This normalization reduces the dynamic range of the data, improves the numerical stability of the algorithm, and decreases sensitivity to noise and variations in observational conditions.

Spectral peaks in the normalized spectra S and R are detected using `scipy.signal.find_peaks`. Peak detection is performed with a prominence threshold of 0.01 in normalized-intensity units, while the default SciPy values are used for the minimum peak height, width, separation, and other parameters.

Having the sets P_S^n, P_R^m, the next step is to establish correct correspondences between the peaks of spectra S and R, i.e., $\Pi^{nm} = (P_S^n, P_R^m)$, ensuring the accuracy of the subsequent calibration. For each peak P_S^n, the most probable match among the peaks P_R^m is determined by comparing their local neighborhoods. The similarity between the neighborhoods is evaluated using the DTW method applied to local windows centered at the corresponding peak positions. Each local window contains 45 pixels, and the DTW distance between the intensity sequences is calculated using the absolute-difference (L1) metric. The similarity between neighborhoods is evaluated using the DTW method on short windows centered at peak position, which returns a similarity score. The peak P_R^m, that maximizes this score is selected as the corresponding match. Thus, pairs $\Pi^{nm} = (P_S^n, P_R^m)$ are formed. Based on the identified correspondences, a polynomial approximation is constructed, defining the calibration function for spectrum S.

It should be noted that in the presence of closely spaced spectral lines, incorrect matches may occur, leading to a degradation in the accuracy of the dispersion solution (as shown in Figure 3. A related work [13] has demonstrated a calibration scheme in which DTW is used for first-pass spectral alignment, followed by refinement using traditional emission-line fitting.

In our case, DTW is used only as a similarity measure for generating candidate correspondences, which makes the resulting set of matches inherently noisy and prone to outliers. Therefore, not all correspondence pairs Π^{nm} are reliable, and an additional filtering stage is required to identify reliable matches. To address this task, the RANSAC algorithm is employed – an iterative probabilistic method for estimating model parameters in the presence of outliers in the data.

In this study, the algorithm is applied as follows:

1. A subset of (K=5) distinct correspondence pairs Π^{nm} is randomly selected without replacement from the DTW-generated candidate-pair list, and these correspondences are assumed to be correct;

2. Based on the selected pairs, a calibration model is constructed, which is then applied to the spectrum S;

3. The accuracy of the obtained model is then evaluated. As an error metric, the deviation between the wavelengths predicted by the model for the peaks of spectrum S and the wavelengths of the corresponding peaks in spectrum R is used:

$$\varepsilon_{nm} = |\lambda_{p^n{}_S} - \lambda_{p^m{}_R}|; \quad (2)$$

4. For each pair, the error tolerance condition is checked: if $\varepsilon_{nm} < \varepsilon_{max}$, the correspondence is considered valid;

5. If the number of valid correspondence pairs Π^{nm} at the current iteration exceeds that of the previous best result, the model is considered optimal and is retained.

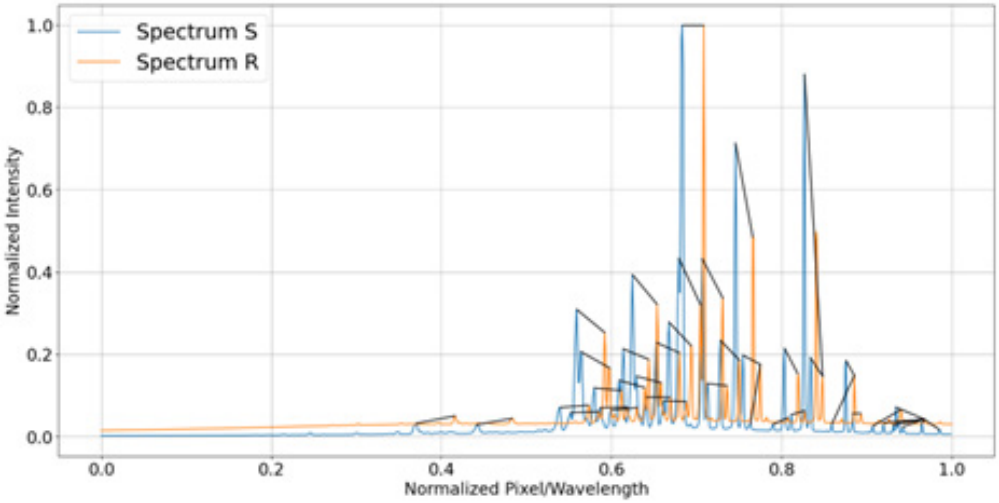

Figure 3 – Result of peak matching between
spectra R and S using the standalone DTW algorithm

The automated RANSAC procedure uses a fixed stopping criterion of 3,000 iterations. This fixed value was introduced to avoid the subjective manual determination of the point at which the number of inliers ceases to change substantially.

The use of RANSAC makes it possible to filter out incorrect pairs and retain only those peak correspondences that minimize the chosen error function and provide the highest calibration accuracy. Since the algorithm is stochastic, the optimal solution may be found either at early iterations or only after a large number of iterations. It is at the RANSAC stage that the final set of consistent correspondences between spectral peaks is formed.

Unlike legacy tools (e.g., the `IRAF reidentify` task) that necessitate a priori knowledge of the reference spectrum and manual initialization, the proposed DTW and RANSAC-based framework ensures complete calibration autonomy while exhibiting high robustness to non-linear distortions and stochastic noise.

As a final step, an inverse transformation from normalized coordinates to physical quantities – wavelengths – is performed.

Results and discussion

We performed a series of tests of the developed algorithm using calibration lamp data (He-Ne-Ar). Peak detection was first applied to both the measured spectrum and the reference spectrum, producing two sets of spectral features: image peaks and reference peaks.

For wavelength calibration, we modeled the relation between pixel position and wavelength using a polynomial of order k=3 A third-order polynomial was chosen because it provides sufficient flexibility to describe nonlinear variations in the pixel-to-wavelength relation while retaining a relatively small number of free parameters, which is important for robust fitting based on a limited

number of peak correspondences. Higher-order polynomials were avoided to limit model complexity and sensitivity to mismatched peaks. This choice aligns with established automated calibration approaches (e.g., [19]). A polynomial of this order requires a minimum of four points to be defined; however, to improve robustness against noise and mismatches, we used random subsets of 5 peak pairs when estimating model parameters.

All spectra contain 1,024 pixels along the dispersion axis. Therefore, a fixed tolerance of (ε_{nm} =1/1024) in normalized wavelength units corresponds approximately to one detector pixel. This value is used exclusively as the RANSAC inlier threshold: a peak correspondence is considered consistent with a candidate polynomial solution if its residual is smaller than this threshold. It should therefore not be interpreted as the final wavelength calibration accuracy. Since the equivalent tolerance in physical wavelength units depends on the spectral range and dispersion of the particular grating configuration, the value of approximately 4.64 Å applies specifically to the 360 lines/mm grating and is not used as a fixed physical tolerance for the other configurations. The final wavelength calibration is obtained only after RANSAC identifies the consistent peak correspondences and the third-order polynomial wavelength solution is determined from the retained matches.

Due to the stochastic nature of the RANSAC approach, the same number of iterations under identical conditions can yield different numbers of inliers. To estimate the algorithm's effectiveness and stability, we measured how the mean number of inliers grows with an increasing number of iterations. The results are shown in Figure 4.

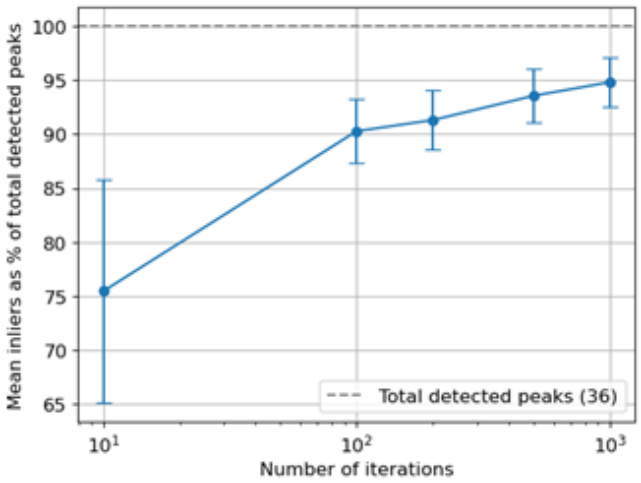

Figure 4 – Dependence of the mean number of detected model inliers
on the maximum number of RANSAC iterations.
Error bars represent the standard deviation

In the test spectrum, with peak-finding prominence set to 10^{-2}, 36 peaks were detected. The analysis shows that a few hundred iterations are sufficient to achieve successful calibration of up to ~90% of the detected peaks, while further increasing the number of iterations leads to gradual saturation of the solution quality. This behavior is consistent with the probabilistic nature of RANSAC, where additional iterations primarily increase the likelihood of finding near-optimal peak correspondences rather than dramatically improving the solution. This proportion of successfully matched pairs allows for a robust polynomial fit of the pixel–wavelength relation of the chosen order, as illustrated in Figure 5.

It should be noted that RANSAC can be applied as a standalone method for peak matching, even without preliminary conditioning using DTW or other techniques. However, its effectiveness strongly depends on the constraints imposed on the candidate correspondences. In its basic form, RANSAC operates solely on positional information and does not take into account the shape of spectral features or their local neighborhoods.

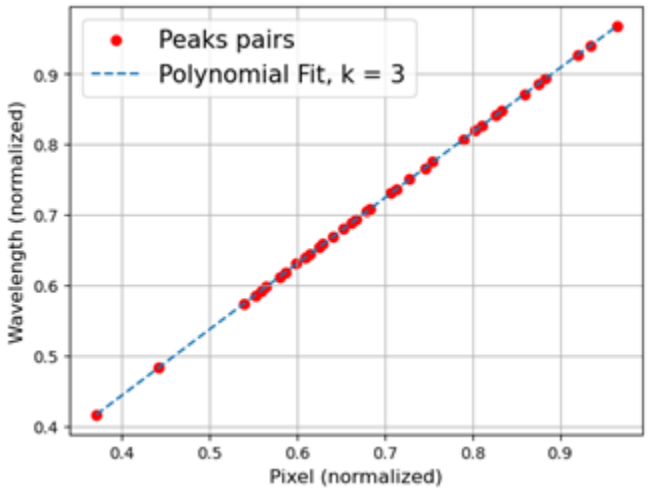

Figure 5 – Polynomial fit to the identified peak correspondences

For reliable peak matching, it is therefore important that the input spectra (both target and reference) are broadly consistent, i.e., they correspond to the same spectral range, lamp type, and overall spectral structure. In practice, this consistency can be ensured using preliminary classification methods or similarity-based approaches such as DTW.

Even in the absence of DTW, effective matching can be achieved by introducing additional constraints into the pair selection process. A naive construction of candidate pairs scales as $N_{pairs} \sim N_R \cdot N_S$, which leads to a large number of incorrect combinations. To reduce this search space, candidate correspondences can be restricted to peaks that are close in normalized coordinate space. At the same time, this limitation also justifies the use of DTW at the preconditioning stage. DTW is able to compare local spectral patterns in a way that is robust to non-linear distortions along the dispersion axis. Unlike direct positional matching, DTW aligns local intensity profiles while allowing elastic shifts between sequences, making it well suited for spectra affected by stretching, compression, or small calibration offsets. As a result, DTW provides more reliable initial correspondences in cases where simple proximity-based criteria may fail.

The resulting polynomial can then be applied to the entire spectrum to perform wavelength calibration, as shown in Figure 6.

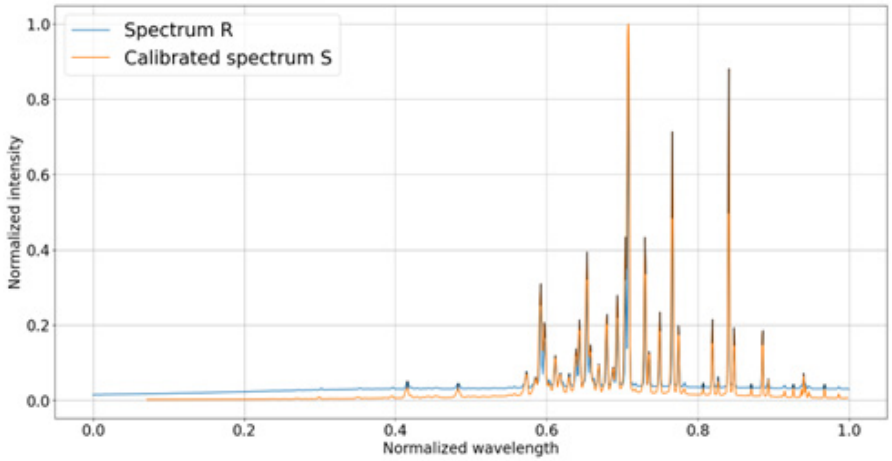

Figure 6 – Spectra R and calibrated S,
with black lines indicating the matched peaks

The error between peaks in spectrum R and their corresponding matched peaks in spectrum S remains within the acceptable tolerance defined by the RANSAC filtering and stays approximately constant across iterations. This indicates that the proposed approach allows an increase in the number of matched peak pairs without a loss of calibration accuracy.

The subsequent application of RANSAC made it possible to eliminate false correspondences and significantly improve the calibration accuracy. Figure 7 shows the final result of the described algorithms: the calibrated spectrum of the calibration lamp and the spectrum of the observed object – the standard star HD 205027 – mapped using the derived calibration.

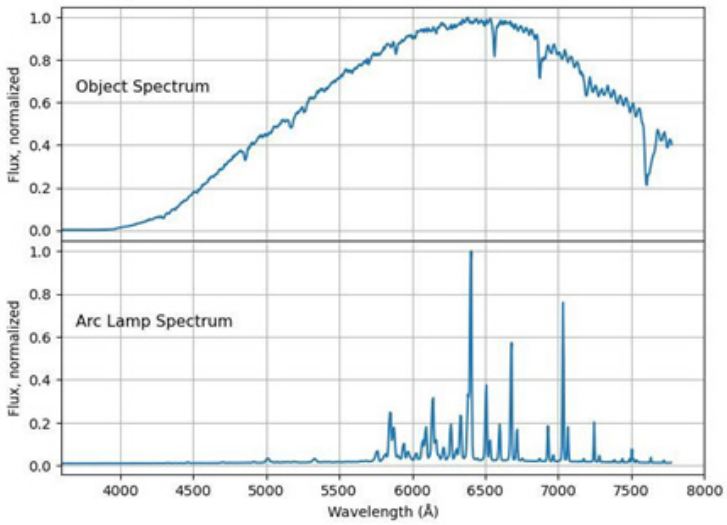

Figure 7 – Calibrated spectrum of the object (top)
based on the calibrated lamp spectrum (bottom)

Overall, the proposed approach demonstrates that combining DTW-based initial matching with subsequent RANSAC filtering provides a robust and efficient solution to the spectral calibration problem. The DTW stage enables the identification of candidate correspondences based on local spectral structure, while RANSAC effectively eliminates outliers and refines the calibration model using only geometrically consistent peak pairs.

The results show that the method achieves a high proportion of correctly matched peaks and produces a stable dispersion solution, even in the presence of noise, missing lines, and closely spaced spectral features. The observed saturation in the number of inliers with increasing iterations confirms the expected probabilistic behavior of RANSAC and indicates that a relatively small number of iterations is sufficient for practical applications.

The final calibrated spectra demonstrate good agreement between the reference and measured data, confirming the validity of the obtained polynomial model. The method is computationally efficient and can be readily applied in automated data reduction pipelines, significantly reducing processing time compared to manual calibration while maintaining high accuracy. The automation and computational efficiency of the proposed approach suggest its potential applicability to space-based observational systems, where data processing may require minimal operator involvement and operate under computational constraints. However, its application to such systems has not yet been tested and will require dedicated validation. Prospective orbital telescope systems, including the proposed cislunar telescope (https://fai.kz/projects/cislunar_phase_1), therefore represent a potential direction for future application of the method.

Conclusion

In this work, we developed and tested an automated pipeline for wavelength calibration of astronomical spectra that combines robust signal matching techniques. The present implementation

was developed and validated using observational data obtained with the AZT-20 telescope, a He-Ne-Ar calibration lamp, and diffraction gratings with line densities of 360, 1800, and 2400 lines/mm. The proposed approach successfully addresses key challenges of spectral calibration, including noise sensitivity, nonlinear distortions, and incorrect peak identification. The DTW algorithm proved effective for initial peak matching, while the subsequent application of RANSAC significantly improved the robustness of the calibration by eliminating false peak matches.

A key contribution of this work lies in the integration of DTW-based local similarity analysis with probabilistic RANSAC filtering into a unified calibration pipeline, enabling reliable correspondence identification without manual intervention. This combination allows the method to operate effectively under realistic observational conditions, where noise, missing lines, and line blending are present.

The results show that the developed method enables fully automated processing of spectral data and substantially increases efficiency compared to traditional manual calibration procedures. This makes it particularly valuable for large datasets, such as those stored in distributed astronomical data centers and made available through Virtual Observatory networks. In addition, the computational efficiency and robustness of the method make it well suited for application in resource-constrained observational systems, including future space-based instruments.

Future work will focus on extending the proposed method to a wider variety of spectral instruments and on its systematic validation and refinement using a substantially larger and more diverse set of real observational data. While the current classification model was primarily trained using synthetically generated spectra and evaluated on the available observational data, further testing on spectra obtained under different observational conditions and instrumental configurations will be essential for assessing and improving its robustness and generalization capability. As additional observational data become available, the pipeline will be progressively adapted and optimized to improve the reliability and accuracy of the wavelength calibration procedure. Furthermore, the developed pipeline has already been integrated into the Kazakhstan Virtual Observatory (KazVO) infrastructure, providing both an interactive web-based interface (https://vo.fai.kz/calibration/) for users and a scalable service (https://vo.fai.kz/obs_data.php?path=/misc/wvl_calib/comp/external) compliant with the IVOA protocol.

Acknowledgments

This research received funding from the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24992807 and Grant No. BR24992759).

REFERENCES

1 Stoughton, C., et al. Sloan Digital Sky Survey: Early Data Release. The Astronomical Journal, 123 (1), 485–548 (2002). https://doi.org/10.1086/324741

2 Law, D.R., et al. The Data Reduction Pipeline for the SDSS-IV MaNGA IFU Survey. The Astronomical Journal, 152, 83 (2016). https://doi.org/10.3847/0004-6256/152/4/83

3 Tody, D. The IRAF Data Reduction and Analysis System. Instrumentation in Astronomy VI, Proc. SPIE, 627, 733–748 (1986). https://doi.org/10.1117/12.968154

4 Odewahn, S.C., et al. Automated Star/Galaxy Discrimination with Neural Networks. In: Digitised Optical Sky Surveys, 215–224 (1992). https://doi.org/10.1086/116063

5 Weir, N., et al. The SKICAT System for Processing and Analyzing Digital Imaging Sky Surveys. Publications of the Astronomical Society of the Pacific, 107 (718), 1243–1254 (1995). https://doi.org/10.1086/133683

6 Collister, A.A., and Lahav, O. ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. Publications of the Astronomical Society of the Pacific, 116 (818), 345–351 (2004). https://doi.org/10.1086/383254

7 Dieleman, S., Willett, K.W., and Dambre, J. Rotation-Invariant Convolutional Neural Networks for Galaxy Morphology Prediction. Monthly Notices of the Royal Astronomical Society, 450 (2), 1441–1459 (2015). https://doi.org/10.1093/mnras/stv632

8 Smith, M.J., and Geach, J.E. Astronomia ex Machina: A History, Primer and Outlook on Neural Networks in Astronomy. Royal Society Open Science, 10 (5) (2023). https://doi.org/10.1098/rsos.221454

9 Fabbro, S., et al. An Application of Deep Learning in the Analysis of Stellar Spectra. Monthly Notices of the Royal Astronomical Society, 475 (3), 2978–2993 (2018). https://doi.org/10.1093/mnras/stx3298

10 Zhang, T., et al. Application of Machine Learning in Astronomical Spectral Data Mining. Astronomical Techniques and Instruments, 2 (2), 73–87 (2025). https://doi.org/10.61977/ati2024057

11 Davenport, J.R.A., et al. Automated Spectroscopic Wavelength Calibration Using Dynamic Time Warping. https://doi.org/10.48550/arXiv.2508.05862

12 LeCun, Y., et al. Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, 86 (11), 2278–2324 (1998). https://doi.org/10.1109/5.726791

13 Sakoe, H., and Chiba, S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing, 26 (1), 43–49 (1978). https://doi.org/10.1109/TASSP.1978.1163055

14 Fischler, M.A., and Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM, 24 (6), 381–395 (1981). https://doi.org/10.1145/358669.358692

15 Nair, V., and Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In: Proceedings of the 27th International Conference on Machine Learning, 807–814 (2010). http://www.icml2010.org/papers/432.pdf

16 Rogowski, M., et al. mpi4py.futures: MPI-Based Asynchronous Task Execution for Python. IEEE Transactions on Parallel and Distributed Systems, 34 (2), 611–622 (2022). https://doi.org/10.1109/TPDS.2022.3225481

17 Kingma, D.P., and Ba, J. Adam: A Method for Stochastic Optimization. (2015). https://doi.org/10.48550/arXiv.1412.6980

18 Srivastava, N., et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15 (1), 1929–1958 (2014). https://www.cs.toronto.edu/~hinton/absps/JMLRdropout.pdf

19 Mirapeix, J., et al. In-Process Automatic Wavelength Calibration for CCD-Spectrometers. In: Optical Sensors 2008, Proc. SPIE, 7003, 70031T (2008). https://doi.org/10.1117/12.781040

[1]Глущенко А.,
магистр, ORCID ID: 0000-0002-0738-7725, e-mail: gluchshenko@fai.kz
[1]Измайлова И.,
магистр, ORCID ID: 0000-0001-9878-0989, e-mail: izmailova@fai.kz
[1,3]Умирбаева А.,
магистр, ORCID ID: 0000-0001-9339-4990, e-mail: umirbayeva@fai.kz
[1,2]Куватова Д.,
магистр, ORCID ID: 0000-0002-5937-4985, e-mail: kuvatova@fai.kz
[1]Юрин Д.,
PhD, ORCID ID: 0000-0002-5604-9757,e-mail: yurin@fai.kz

[1]В.Г. Фесенков атындағы астрофизикалық институты, Алматы қ., Қазақстан
[2]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан
[3]Назарбаев университеті, Астана қ., Қазақстан

АСТРОНОМИЯЛЫҚ СПЕКТРЛЕРДІҢ ТОЛҚЫН ҰЗЫНДЫҒЫН АВТОМАТТЫ ТҮРДЕ КАЛИБРЛЕУ: DTW–RANSAC СӘЙКЕСТЕНДІРУ ЖӘНЕ CNN НЕГІЗІНДЕГІ АЛДЫН АЛА ӨҢДЕУ

Андатпа

Мақалада машиналық оқыту әдістері мен сенімді сәйкестендіру алгоритмдерін біріктіруге негізделген астрономиялық саңылаулы спектрлердің толқын ұзындығын калибрлеуге арналған автоматтандырылған алгоритм ұсынылады. Ұсынылған тәсіл спектрлік жолақты автоматты түрде анықтауға арналған бірөлшемді конволюциялық нейрондық желіні (CNN), одан кейін динамикалық уақытты ығыстыру (DTW) әдісі арқылы

бастапқы шыңдарды сәйкестендіруді және кездейсоқ үлгілер консенсусы (RANSAC) алгоритмі негізінде кейінгі нақтылауды қамтиды. CNN моделі синтетикалық және нақты спектрлік деректерде оқытылып, әртүрлі дифракциялық торлар үшін жоғары жіктеу дәлдігін көрсетеді. DTW әдісі сызықтық емес бұрмаланулар мен шу болған жағдайда да бақыланатын және эталондық спектрлердің сенімді бастапқы тураласуын қамтамасыз етеді. RANSAC алгоритмін қолдану қате шың сәйкестіктерін сүзу және дисперсиялық шешімді оңтайландыру арқылы калибрлеу дәлдігін одан әрі арттырады. Әдіс калибрлеу шамдарынан алынған нақты бақылау деректерінде сыналып, жоғары сенімділік, тиімділік және толық автоматтандыру мүмкіндігін көрсетеді. Дәстүрлі қолмен немесе жартылай автоматтандырылған тәсілдермен салыстырғанда, ұсынылған өңдеу құбыржолы жоғары калибрлеу дәлдігін сақтай отырып, өңдеу уақытын едәуір қысқартады. Әзірленген шешімді виртуалды обсерватория инфрақұрылымында қызмет ретінде орналастыру жоспарлануда және көптеген спектрлік деректерді өңдеу үшін қолданылады.

Түйін сөздер: астрономиялық деректерді өңдеу, щелелі спектрлерді калибрлеу, dynamic time warping, RANSAC, машиналық оқыту.

[1]Глущенко А.,
магистр, ORCID ID: 0000-0002-0738-7725, e-mail: gluchshenko@fai.kz
[1]Измайлова И.,
магистр, ORCID ID: 0000-0001-9878-0989, e-mail: izmailova@fai.kz
[1,3]Умирбаева А.,
магистр, ORCID ID: 0000-0001-9339-4990, e-mail: umirbayeva@fai.kz
[1,2]Куватова Д.,
магистр, ORCID ID: 0000-0002-5937-4985, e-mail: kuvatova@fai.kz
[1]Юрин Д.,
PhD, ORCID ID: 0000-0002-5604-9757,e-mail: yurin@fai.kz

[1]Астрофизический институт им. В.Г. Фесенкова, г. Алматы, Казахстан
[2]Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан
[3]Назарбаев Университет, г. Астана, Казахстан

АВТОМАТИЧЕСКАЯ КАЛИБРОВКА ДЛИНЫ ВОЛНЫ АСТРОНОМИЧЕСКИХ СПЕКТРОВ С ИСПОЛЬЗОВАНИЕМ DTW–RANSAC СОПОСТАВЛЕНИЯ И ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ НА ОСНОВЕ CNN

Аннотация

В данной статье представлен автоматизированный алгоритм калибровки длины волны щелевых астрономических спектров, основанный на сочетании методов машинного обучения и надежных алгоритмов сопоставления. Предлагаемый подход включает в себя одномерную сверточную нейронную сеть (CNN) для автоматической идентификации спектрального диапазона, за которой следует первоначальное сопоставление пиков с использованием метода динамического временного сдвига (DTW) и последующее уточнение с помощью алгоритма консенсуса случайных выборок (RANSAC). Модель CNN обучена как на синтетических, так и на реальных спектральных данных и демонстрирует высокую точность классификации для различных дифракционных решеток. Метод DTW обеспечивает надежное начальное совмещение наблюдаемых и эталонных спектров даже при наличии нелинейных искажений и шума. Применение RANSAC дополнительно повышает точность калибровки за счет отсеивания некорректных соответствий пиков и оптимизации решения дисперсии. Метод был проверен на реальных данных наблюдений от калибровочных ламп и демонстрирует высокую устойчивость, эффективность и возможность полной автоматизации. По сравнению с традиционными ручными или полуавтоматическими подходами, предложенный конвейер значительно сокращает время обработки, сохраняя при этом высокую точность калибровки. Разработанное решение планируется развернуть в виде сервиса в инфраструктуре виртуальных обсерваторий и применять для обработки многочисленных спектральных данных.

Ключевые слова: обработка астрономических данных, калибровка щелевых спектров, dynamic time warping, RANSAC, машинное обучение.