VENTSPILS AUGSTSKOLA INFORMĀCIJAS TEHNOLOĢIJU FAKULTĀTE

KURSA DARBS

"IS vai pētniecisko projektu izstrāde"

Plašā skata radio-astronomijas novērojumu attēlu analīze

Autors	Ventspils Augstskolas
	Informācijas tehnoloģiju fakultātes
	maģistra studiju programmas "Datorzinātnes"
	2. kursa students
	Kristaps Siksna
	Matr.nr. 20080007
	Paraksts:
Darba vadītājs	Mg.sc.comp. Jānis Šteinbergs
	Paraksts:

SATURS

IE	EVADS.		3
1.	LIE	LO DATU ASTRONOMIJA	4
	1.1.	Lielo datu astronomija	4
	1.2.	Astronomijā lietotie datu formāti	4
	1.2.1	. FITS faila saturs	5
	1.3.	Plašā skata radio-astronomijas attēli	7
2.	RA	DIO-ASTRONOMIJAS NOVĒROJUMU ATTĒLU ANALĪZE	9
	2.1.	Attēlu apstrāde ar DS9 rīku	9
	2.2.	Trokšņu noņemšana un signāla izgriešana	10
	2.3. Ob	jektu lokalizācija attēlā	13
	2.3.1	. Šablonu salīdzināšanas algoritms un šablonu izveidošana	13
	2.3.2	. Šablonu salīdzināšanas algoritma rezultāti	14
S]	ECINĀJ	UMI UN PRIEKŠLIKUMI	17
ΙZ	ZMANT	OTĀS LITERATŪRAS UN AVOTU SARAKSTS	18
G	ALVOJ	UMS	20

IEVADS

Pēdējā desmitgadē datu apjoms no jaunās paaudzes radio teleskopiem, kā, piemēram, LOFAR ir palielinājies par aptuveni divām reizēm. Jaunās paaudzes teleskopi nodrošina nepieredzētas novērojumu iespējas, pateicoties augstajai izšķirtspējas jaudai, plašam skatu leņķim un augstai spektrālai un laika izšķirtspējai. Tādēļ aizvien plašāk novērojumu rezultātā iegūto datu apstrādē tiek izmantotas mašīnmācīšanās metodes. Šī darba ietvaros tiks apskatīti radio novērojumu attēli ar plašu skata leņķi, un viens no objektu atrašanas algoritmiem — Paraugu salīdzināšanas algoritms.

Iegūtie dati no novērojumiem ir ļoti liela izmēra — ikdienā esam pieraduši pie Gigabaitiem, dažiem Terabaitiem. Cietā diska apjoms parasti svārstās no 256 GB — 2TB, bet CD disks ir tikai 4,7 GB liels. Astronomijā šie apjomi ir robežās no vairākiem Terabaitiem līdz pat Petabaitiem (1 Petabaits = 1000 TB). Piemēram, teleskops "VISTA" (angļu vaļ. Visible and Infrared Survey Telescope for Astronomy) gada laikā iegūst 150 TB datus, bet LSST (angļu val. Large Synoptic Survey Telescope) pat 500 TB mēnesī. [1] Nākotnes teleskops būs SKA (angļu val. Square Kilometre Array). Vienā sekundē tiks iegūti no 0,5 — 1 TB datu. Novērojumi varēs ilgt pat vairākas stundas. Rezultātā, tiks iegūts datu kubs, kas varēs saturēt pat 2¹⁵ pikseļu uz vienas telpiskās ass un 2¹⁶ frekvenču kanālus. SKA teleskops radīs apmēram 300 PB (petabaitu) datu gadā. [2]

Darba mērķis: Iepazīties ar radio astronomisko novērojumu attēlu apstrādi. Demonstrēt vienu no objektu lokalizācijas metodēm.

Darba uzdevumi:

- Iepazīties ar radio astronomisko novērojumu rezultātā iegūto attēlu apstrādi.
- Demonstrēt objektu lokalizāciju plaša skata radio astronomijas novērojumu rezultātā iegūtajos attēlos.

1. LIELO DATU ASTRONOMIJA

1.1. Lielo datu astronomija

Astronomisko novērojumu datu apjoms ir audzis, jo teleskopi iegūst datus, izmantojot augstu izšķirtspēju, plašu skatu leņķi un augstu spektrālo un laika izšķirtspēju. [3] Attīstoties dažādu veidu astronomiskajiem novērojumiem — infrasarkanā, ultravioletā, optiskā spektra, pieaug arī iegūto datu izmērs. Tas var sasniegt pat vairākus petabaitus. [4]

Novērojumu rezultātā var iegūt spektru - Furjē transformāciju no signāla laikā. Spektru iespējams iegūt katram attēla pikselim, izveidojot datu kubu, vai attēlu kubu. Pēc tam ir iespējams izveidot katalogu - sarakstu ar objektiem un to īpašībām.

1.2. Astronomijā lietotie datu formāti

Astronomijā tiek lietoti dažādi datu formāti. Tie iedalās datu kubos (angļu val. Data cubes), spektrālajos datos (angļu val. Spectral data), katalogu datos (angļu val. Catalog data) un attēla datos (angļu val. Image data).

- Spektrs parāda kā objekta izstarotā gaisma sadalās dažādos viļņu garumos. Tie satur informāciju par tādiem fiziskajiem raksturlielumiem kā temperatūra un ķīmiskais sastāvs. [5] Uzņemot daudz šāda veida attēlus ir iespējams izveidot 2D galaktiku īpašību kartes.
- Attēla dati satur attēla pikseļu intensitātes vērtības 2D matricu formā, kur katra vērtība apzīmē spilgtumu. Parasti šie attēli tiek saglabāti FITS formātā (angļu val. Flexible Image Transport System). Attēla faila paplašinājums ir ".fits".
- Datu kubi ir divu dimensiju attēli, kur katrs attēla pikselis nesatur spilgtuma
 vērtību, bet gan visu spektru. Tā kā spektrs ir 1D, bet attēls ir 2D, tad rodas ilūzija, ka tiek iegūts
 3D objekts. Tāpēc to arī sauc par datu kubu.
- Kataloga dati satur informāciju par dažādiem astronomijas objektiem. Izmantojot kataloga datus, var iegūt papildus informāciju par objektu un salīdzināt novērojumā iegūto informāciju ar citos novērojumos iegūto informāciju. Piemēram, ja tiek novērota zvaigzne, var iegūt atrašanās vietas koordinātes, kustības trajektoriju, temperatūru, spilgtumu dažādos viļņu garumos. [1] [5]

Šī darba ietvaros tiks strādāts ar Attēla datu tipu FITS formātā.

1.2.1. FITS faila saturs

FITS failu formāts satur astronomisko novērojumu datus matricu formā. Šo, nu jau standartu, ieviesa 20. gadsimta beigās. Fails sastāv no divām daļām — galvenes (angļu val. header) un datu daļas.

Galvenes daļa satur informāciju par pašu failu jeb faila metadatus cilvēkam saprotamā veidā. Piemēram, kad uzņemts attēls un ar ko tas uzņemts. Šīs informācijas attēlošanai tiek izmantota notācija atslēgas vārds = vērtība / komentārs (angļu val. keyword = value/comment). Galvene sākotnēji tika veidota no perfokartēm (angļu val. punch cards), katrai kartei aizņemot 80 baitus. Mūsdienās šis process notiek digitāli. Atslēgas vārds sastāv no 8 simbolu garas ASCII virknes. Komentārs paskaidro atslēgas vārdu saprotamā veidā. HDU (angļu val. Header Data Unit) parasti satur aprakstīto informāciju (tabula 1.1.). [6] [7] [8] Failam ir iespējamas arī vairākas galvenes, taču ir tikai viena galvenā. Informāciju ir iespējams papildināt, dzēst un mainīt. [9]

1.1. tabula Galvenie galvenes faila parametri un to skaidrojumi

Parametrs	Skaidrojums
SIMPLE	Apraksta FITS formāta standartu. Katrs
	galvenes fails sākas ar šo parametru.
BITPIX	Parāda, cik biti ir katrā pikselī.
BUNIT	Mērvienība, piemēram, Jy. (Jansky per
	Beam).
NAXIS	Asu skaits.
NAXIS1	Parāda elementu skaitu uz 1. ass.
NAXIS2	Parāda elementu skaitu uz 2. ass.
NAXIS3	Parāda elementu skaitu uz 3. ass.
NAXIS4	Parāda elementu skaitu uz 4. ass.
EXTEND	Boolean vērtība, kas nosaka, vai failu var
	paplašināt.
TELESCOP	Norāda, ar kādu teleskopu uzņemts attēls.
INSTRUME	Norāda, ar kādu instrumentu uzņemts attēls.
OBSERVER	Norāda, kurš uzņēma attēlu.

DATE-OBS vai DATE	Datums, kad uzņemts attēls.
OBJECT	Objekts, kas tika uzņemts.
СТҮРЕ	Parāda, kas atrodas uz katras no asīm.
CTYPEn	Parāda, kādas vērtības ir uz katras no asīm.
	CTYPE1 atbilst NAXIS1 vērtībai, CTYPE2
	atbilst NAXIS2 vērtībai utt.

Datu daļa satur novērojuma datus matricu formā. Tie satur intensitātes vērtības — jo lielāks skaitlis, jo spožāka šī vieta ir attēlā. Intensitāte tiek mērīta mērvienībās Jy/Beam (angļu val. Jansky per Beam), kas nosaka, cik daudz jaudas nonāk no objekta līdz teleskopam. Šo saīsinājumu Jy var izteikt kā Vati uz kvadrātmetru hercu (formula 1.1). [10]

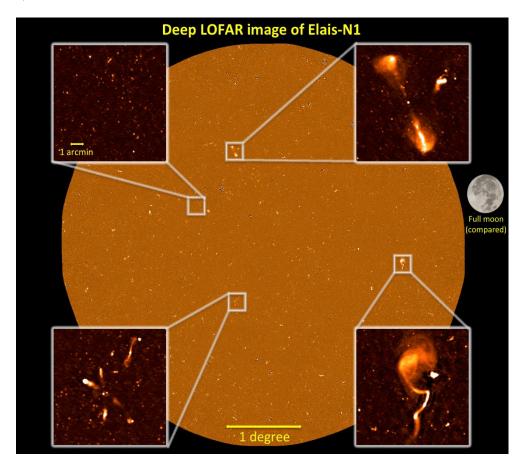
$$1jansky = 1Jy \equiv 10^{-26}Wm^{-2}Hz^{-1}$$
 (1.1)

 Wm^{-2} – Vati uz kvadrātmetru.

 Hz^{-1} – apzīmē hercus (svārstību skaitu 1 sekundē).

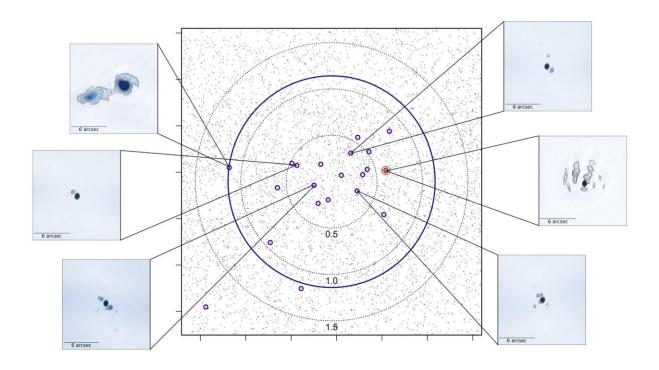
1.3. Plašā skata radio-astronomijas attēli

Plaša skata novērojums nozīmē, ka novērojuma laikā tiek aptverta liela teritorija. No antenām raidītais signāls tiek apkopots, rezultātā izveidojot novērojumu attēlu (attēls. 1.1., attēls. 1.2.). [11]



1.1. att. LOFAR novērojums "Elais-N1" ar dažiem objektiem palielinājumā. [13]

Attēlā 1.1. redzams novērojums "Elais-N1", kas tika uzņemts ar LOFAR teleskopu. Novērojums ilga 164 stundas. Tas attēlo galaktikas. Daži no objektiem ir palielināti, uz ko norāda kvadrāts ap tiem.



1.2. att. LOFAR novērojums. [14]

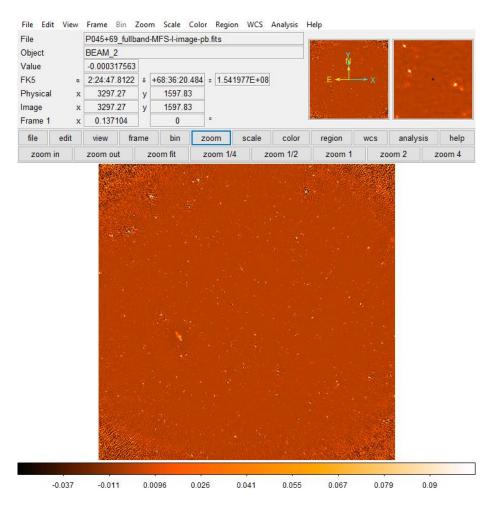
Attēlā 1.2. ir redzams viens no piemēriem plaša skata attēliem. Tajā ir redzams P205+55 lauks. [12] Līdzīgi tiks meklēti objekti arī darba praktiskajā daļā, izmantojot Paraugu salīdzināšanas metodi.

2. RADIO-ASTRONOMIJAS NOVĒROJUMU ATTĒLU ANALĪZE

2.1. Attēlu apstrāde ar DS9 rīku

DS9 ir rīks, ar kura palīdzību var apskatīt FITS formāta attēlu. Šis rīks dod iespēju apskatīt un modificēt dažādus raksturlielumus, piemēram, histogrammu, mainīt apskatāmo intervālu attēlā un krāsu diapazonu, kādā attēlot attēlu (angļu val. Colormap), izgriezt un apskatīt attēla konkrētu daļu (attēls 2.1.). [13] Ir pieejamas līdzīgas vides DS9, piemēram, CARTA (angļu val. Cube Analysis and Rendering Tool for Astronomy) [14], taču šī darba ietvaros tiks izmantots DS9 rīks.

Ar šī rīka palīdzību tiks veidoti paraugi, kas tālāk tiks izmantoti paraugu salīdzināšanas algoritmā.



2.1. att. Attēls ielasīts DS9 vidē.

Attēlā 2.1. redzams viens no novērojumu attēliem ar noņemtu troksni un nomainītu krāsu, lai labāk izceltu objektus. Augšējā daļā redzama informācija par attēlu — nosaukums, kā arī pikseļu koordinātes x un y, kas mainās, bīdot peles kursoru pa attēlu. WCS (angļu val. World Coordinate System) apzīmē astronomisko koordināšu sistēmu, parādot koordinātas stundās, minūtēs, sekundēs.

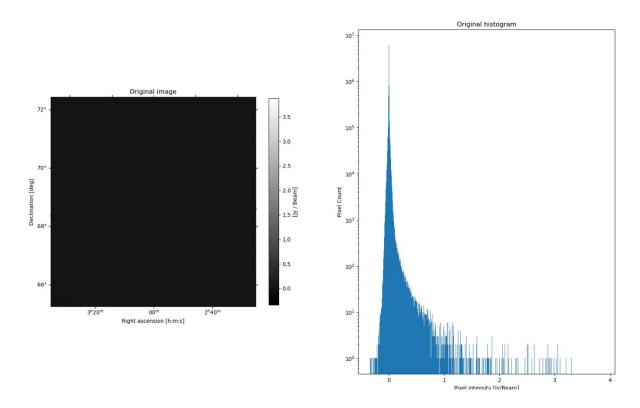
2.2. Trokšņu noņemšana un signāla izgriešana

Turpmāk attēlu apstrāde tiks veikta, izmantojot programmēšanas valodu Python, kas mūsdienās ir ļoti plaši izmantota. Tā savas sintakses dēļ ir piemērota tieši zinātnē, jo ļauj ērti apstrādāt datus, izmantojot iebūvētās bibliotēkas. Nav jāsatraucas par mainīgā tipu, tos definējot.

Galvenās no izmantotajām bibliotēkām:

- matplotlib.pyplot lai attēlotu attēlus un grafikus [15]
- numpy satur matemātiskās operācijas (min, max, std....) [16]
- astropy.io -> FITS satur funkcijas FITS failu ielasīšanai un apstrādei [17]
- skimage.feature -> match_template viena no galvenajām funkcijām, kas tika
 izmantota paraugu meklēšanai [18]

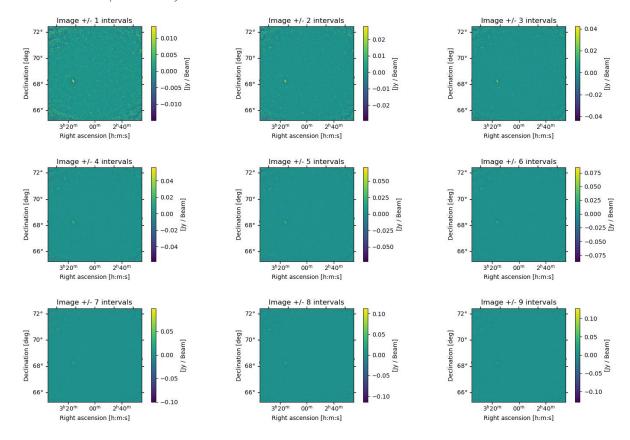
Lai noteiktu troksni, parasti attēlam taisa histogrammu, kura attēlo pikseļu intensitātes vērtības *(attēls 2.2.)*. Histogrammā tika meklēts intensitātes apgabals ar visvairāk pikseļiem.



2.2. att. Oriģināls attēls un tā histogramma.

Kā redzams 2.2. attēlā, pārsvarā visas vērtības attēlā ir ap 0, par ko liecina melnā krāsa skalā, kas atbilst 0 vērtībai. Pēc histogrammas var secināt, ka tik tiešām intensitātes apgabals, kurā ir visvairāk pikseļu ir ap 0. Histogramma un attēls ir attēloti logaritmiskajā skalā.

Lai atrastu optimālu apgabalu un izgrieztu nevajadzīgo troksni, tika izveidoti 9 intervāli, sākot no maksimālās vērtības attēlā un virzoties no tās uz abām pusēm, par soli ņemot standartnovirzi. (attēls 2.3.).



2.3. att. Novērojuma attēls dažādos intervālos.

Attēla 2.3. rezultāti tika salīdzināti ar DS9 vidi, lai pārliecinātos, ka intervāli iegūti pareizi. Vislabākais ir attēls, kas atrodas kreisajā apakšējā malā ar intervālu [-0.1005555, 0.09916625], jo tajā vislabāk attēlojas objektu struktūra. Šis intervāls tika iegūts no 7 standartnovirzēm.

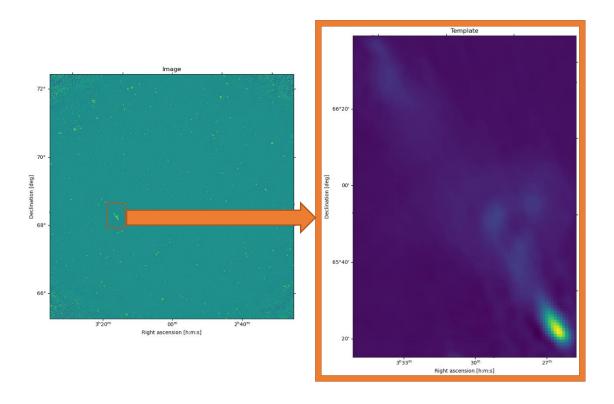
2.3. Objektu lokalizācija attēlā

2.3.1. Šablonu salīdzināšanas algoritms un šablonu izveidošana

Šablonu (paraugu) salīdzināšanas algoritms (angļu val. Template Matching) ir algoritms, ko izmanto attēlu apstrādē, lai detektētu objektus attēlā. Galvenais princips: izmantojot paraugu (attēla daļu vai objektu), atrast objektus, ar ko šis paraugs sakrīt attēlā. Rezultātā, tiek parādītas vietas, parasti ar kvadrātu vai apli, kurās tika atrasti objekti. Šim algoritmam ir vairākas metodes, pēc kurām var meklēt paraugus attēlos. Piemēram, korelācija, kross – korelācija, noteikšana pēc attēla un parauga stūriem u.c. [19]

Praktiskās daļas ietvaros, tā kā tiek strādāts ar Python programmēšanas valodu, šī metode tiks veikta, izmantojot iebūvēto funkciju *match_template* no *skimage* bibliotēkas. [18]

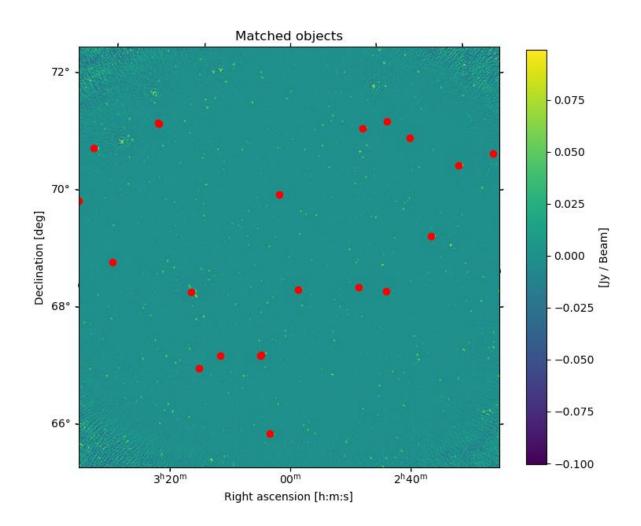
Lai būtu iespējams izmantot šo metodi, ir nepieciešams izveidot paraugus (attēls 2.4.). Paraugi būs veidoti DS9 vide. Tajā tiks ielasīts attēls un no attēla tiks izgriezti atsevišķi objekti – paraugi, izmantojot Edit -> Cut funkcionalitāti. Pēc tam šos paraugus ielasīs kā attēlus Python kodā un veiks paraugu salīdzināšanas metodi.



2.4. att. Kreisajā pusē attēls, labajā pusē palielināts paraugs.

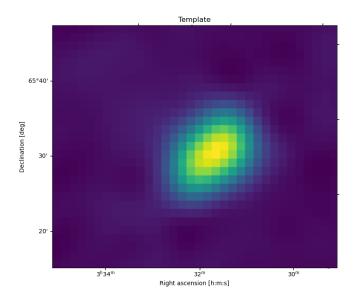
2.3.2. Šablonu salīdzināšanas algoritma rezultāti

Saite uz kodu Github: Praktiskais_Kristaps_Siksna



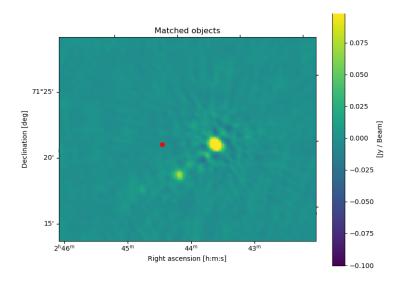
2.5. att. Paraugu salīdzināšanas algoritma rezultāti.

Secinot pēc 2.5. attēla, paraugu salīdzināšanas algoritms nav efektīvs, jo dotajā attēlā ir parādīti atrastie objekti ar soli 0.2. Lai rezultāti būtu ticami, solim ir jābūt vismaz 0.7. Šim par iemeslu varbūt vairāki faktori. Viens no tiem - neprecīzi izveidoti paraugi. Izmantotais paraugs rezultātam redzams 2.6. attēlā. Kopumā tika izmantoti 3 dažādi paraugi. Daži no tiem vairāk pietuvināti, daži attālināti, kā, piemēram, 2.6. attēlā redzamais paraugs.



2.6. att. Izmantotais paraugs.

Pietuvinot vienu no sarkanajiem punktiem, var secināt, ka paraugs nav atrasts tieši tur, kur tam jābūt, bet gan blakus tam *(attēls 2.7.)*.



2.7. att. Viens no atrastajiem paraugiem neprecīzā vietā.

Turpinot šo tēmu maģistra darbā tiks veikts:

- Objektu atrašana attēlos, izmantojot neironu tīklus
- Atrasto objektu parametru noteikšana (piemēram, koordināšu noteikšana astronomiskajā sistēmā)
- Optimāla intervāla noteikšana. Kā viens no attēla kvalitātes rādītājiem tiks aprēķināts RMS (angļu val. Root Mean Square)
- Objektu salīdzināšana ar katalogiem, lai iegūtu papildus informāciju par tiem
- Skripta izveide citā programmēšanas valodā (Java vai C++). Sintakses un datu apstrādes grūtības pakāpes salīdzināšana
- Aprakstīt attēlu izveidošanas procesu radio interferometriju.

SECINĀJUMI UN PRIEKŠLIKUMI

- 1) Pēc kursa darbā paveiktā var secināt, ka paraugu salīdzināšanas algoritms nav ļoti efektīvs, jo objekti tiek atrasti neefektīvi, izmantojot mazu 0.2 slieksni. Ar normālu sliekšņa vērtību, piemēram, 0.7 netika atrasts neviens objekts.
- 2) Atrasto objektu skaits ir atkarīgs arī no izmantotajiem paraugiem un to kvalitātes. Paraugu neprecīza izveide, piemēram, pārāk tāls vai tuvs (pikseļots) objekts vai objekts ar troksni.
- 3) Ir iespējas optimālā intervāla noteikšanā, vizuālo inspekciju aizstāt ar matemātisku pieeju, piemēram, minimizēt RMS attēlā.
- 4) Plaša skata novērojumi ļauj iegūt novērojumus ar vairākiem objektiem. Tas nozīmē, ka pēc tam ir iespēja detektēt un noteikt parametrus uzreiz vairākiem objektiem.
- 5) Sekmīgākai paraugu noteikšanai viena no iespējām būtu apmācīt neironu tīklus, kas apmācību rezultātā spēju atrast objektus precīzāk.

IZMANTOTĀS LITERATŪRAS UN AVOTU SARAKSTS

- [1] M. Pössel, "A Beginner's Guide to Working with Astronomical Data," *arXiv preprint arXiv:1905.13189*, 2019, Available online: https://arxiv.org/pdf/1905.13189.pdf.
- [2] A.M.M. Scaife, "Big telescope, big data: towards exascale with the Square Kilometre Array," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 378, Mar. 2020, p. 20190060, Available online: https://royalsocietypublishing.org/doi/10.1098/rsta.2019.0060, (Accessed: January 8, 2022).
- [3] G.B. Berriman and S.L. Groom, "How Will Astronomy Archives Survive the Data Tsunami? Astronomers are collecting more data than ever. What practices can keep them ahead of the flood?," *Queue*, vol. 9, 2011, pp. 20–27, Available online: https://dl.acm.org/doi/pdf/10.1145/2039359.2047483.
- [4] Y. Zhang and Y. Zhao, "Astronomy in the Big Data Era," *Data Science Journal*, vol. 14, May. 2015, p. 11, Available online: http://datascience.codata.org/articles/10.5334/dsj-2015-011/, (Accessed: December 29, 2021).
- [5] J.P. Long and R.S. de Souza, "Statistical methods in astronomy," *arXiv:1707.05834* [astro-ph, physics:physics, stat], Oct. 2017, Available online: https://arxiv.org/pdf/1707.05834.pdf, (Accessed: January 1, 2022).
- [6] D.C. Wells and E.W. Greisen, "FITS-a flexible image transport system," 1979, p. 445, Available online: https://adsabs.harvard.edu/pdf/1979ipia.coll..445W.
- [7] W.D. Pence, L. Chiappetti, C.G. Page, R.A. Shaw, and E. Stobie, "Definition of the flexible image transport system (fits), version 3.0," *Astronomy & Astrophysics*, vol. 524, 2010, p. A42, Available online: https://www.aanda.org/articles/aa/pdf/2010/16/aa15362-10.pdf.
- [8] J. Ponz, R. Thompson, and J. Munoz, "The FITS image extension," *Astronomy and Astrophysics Supplement Series*, vol. 105, 1994, pp. 53–55, Available online: https://adsabs.harvard.edu/pdf/1994A%26AS..105...53P.
- [9] M. Scroggins, B. Boscoe, "Once FITS, always FITS" Available online: https://arxiv.org/ftp/arxiv/papers/1809/1809.09224.pdf, (Accessed: December 6, 2021).
- [10] James J. Condon and Scott M. Ransom, "2 Radiation Fundamentals." Essential Radio Astronomy" Available online: https://www.cv.nrao.edu/~sransom/web/Ch2.html#S1, (Accessed: December 6, 2021).
- [11] Sabater, J. et al, "LoTSS Deep fields" Available online: https://www.lofar-surveys.org/deepfields_press.html, (Accessed: January 9, 2022).
- [12] Morabito, L. K. et al, "LOFAR Long-baseline Surveys" Available online: https://www.lofar-surveys.org/longbaselines.html, (Accessed: January 9, 2022).
- [13] "Introduction to the ds9 Interface DS9" Available online: http://ds9.si.edu/doc/user/gui/index.html, (Accessed: January 16, 2022).
- [14] CARTA team, "CARTA documentation" Available online: https://carta.readthedocs.io/_/downloads/en/1.4/pdf/, (Accessed: January 20, 2022).
- [15] John Hunter, Darren Dale, Eric Firing, "matplotlib.pyplot Matplotlib 3.5.1 documentation" Available online: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.html, (Accessed: January 21, 2022).
- [16] "NumPy" Available online: https://numpy.org/, (Accessed: January 21, 2022).

- [17] The Astropy Developers, "FITS File Handling (astropy.io.fits) Astropy v5.0" Available online: https://docs.astropy.org/en/stable/io/fits/index.html, (Accessed: January 21, 2022).
- [18] scikit-image development team, "Template Matching skimage v0.20.0.dev0 docs" Available online: https://scikit-image.org/docs/dev/auto_examples/features_detection/plot_template.html, (Accessed: January 16, 2022).
- [19] Adaptive Vision, "Template Matching" Available online: https://docs.adaptive-vision.com/4.7/studio/machine_vision_guide/TemplateMatching.html, (Accessed: January 16, 2022).

GALVOJUMS

Ar šo es, Kristaps Siksna, galvoju, ka šis kursa darbs kursā "IS vai pētniecisko projektu
izstrāde" ir manis paša patstāvīgi izpildīts oriģināls darbs. Visi informācijas avoti, kā arī no tiem
ņemtie dati un definējumi ir norādīti darbā. Šis darbs tādā vai citādā veidā nav iesniegts neviena
citai pārbaudījumu komisijai un nav nekur publicēts.

Esmu informēts (-a), ka mans kursa darbs tiks ievietots un apstrādāts Vienotajā datorizētajā plaģiāta kontroles sistēmā plaģiāta kontroles nolūkos.

2022. gada 18. janvārī

Es, *Kristaps Siksna*, atļauju Ventspils Augstskolai savu kursa darbu bez atlīdzības ievietot un uzglabāt Latvijas Nacionālās bibliotēkas pārvaldītā datortīklā Academia (www.academia.lndb.lv), kurā tie ir pieejami gan bibliotēkas lietotājiem, gan globālajā tīmeklī tādā veidā, ka ikviens tiem var piekļūt individuāli izraudzītā laikā, individuāli izraudzītā vietā.

Piekrītu_____

Nepiekrītu