

University of Groningen

Efficient morphological tools for astronomical image processing

Moschini, Ugo

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Moschini, U. (2016). *Efficient morphological tools for astronomical image processing*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Samenvatting

In dit proefschrift worden nieuwe methoden en software tools uit het gebied van de mathematische morfologie, geschikt voor beeldanalyse in astronomische en satelliet data sets, verkend en ontwikkeld. Gegeven het enorme aantal, en de hoge bit-diepte van de met moderne instrumenten vergaarde data, is automatische segmentatie van de belangrijke objecten vereist.

In de mathematische morfologie zijn z.g. *connected filters* in staat beelden te bewerken door pad-samenhangende groepen pixel met dezelfde intensiteit, de z.g. *samenhangende componenten*, te behouden, of te verwijderen. *Connected filters* gebaseerd op vorm of grootte, die worden gebruikt voor beeldfiltering of segmentatie zijn onvoldoende in staat om met de grote heterogeniteit van astronomische objecten om te gaan. Bovendien zijn veel van de structuren die van belang zijn voor astronomen om de evolutie en morfologie van melkwegstelsel te doorgronden bijzonder zwak en diffuus, en moeilijk te onderscheiden van ruis.

Zelfs menselijke waarnemers kunnen niet altijd gemakkelijk deze structuren zien. In dit proefschrift wordt een nieuw *connected filter* voorgesteld, gebaseerd op de statistische verdeling van de ruis in astronomische datasets. De filtering is geïmplementeerd in een *max-tree data structuur*. Deze structuur is een hiërarchische representatie van de samenhangende componenten in het beeld, en is geschikt om op efficiënte manier beelden te filteren en te segmenteren. Clusterende connectiviteits maskers, uit de theorie van de z.g. *tweede-generatie connectiviteit*, zijn onderzocht om te zien of zo zwakkere, diffuse delen van melkwegstelsels behouden konden worden.

Om met data sets met groottes in de orde van gigapixels of gigavoxels om te gaan zijn parallelle algoritmes voor het bouwen van *max-trees* ontwikkeld in recente jaren. Een nieuw algoritme wordt voorgesteld, dat data met extreem hoog dynamisch bereik efficiënt kan verwerken, iets wat met de bestaande technieken onmogelijk was. Verwerkingsnelheden zijn 20 keer sneller dan met het snelste bestaande sequentiële algoritme, wat de bruikbaarheid van de morfologische filter

technieken vergroot.

Er zijn verscheidene experimenten uitgevoerd betreffende het klassificeren van astronomische objecten. Het doel was samensmeltende of botsende melkwegstelsels te onderscheiden van objecten die toevallig overlaptten. Patroonspectra gebaseerd op connected filters werden voor alle objecten berekend en gebruikt als data vectoren gebruikt in een classificatie algoritme, waarbij een nauwkeurigheid van 80% werd gehaald.

Tot slot is een nieuw algoritme gebaseerd op z.g. hyperconnectiviteit ontwikkeld om beelden te filteren op "hypersamenhangende" componenten die behoren bij een specifieke vorm van hyperconnectiviteit, met als doel het z.g. "lekkage" probleem van connected filters aan te reduceren.

In alle hoofdstukken wordt de theorie en uitleg over de nieuwe algoritmes vergezeld van stukken pseudo code. Waar mogelijk worden experimentele resultaten en vergelijkingen met andere methoden gepresenteerd. Voordelen en beperkingen van de nieuwe methoden worden uitgelegd zodat de lezer een goed beeld kan vormen van de bruikbaarheid en toepasbaarheid in andere vakgebieden.

In Hoofdstuk 2 wordt een nieuwe vorm van tweede generatie connectiviteit geïntroduceerd en een methode ontwikkeld om zwakke structuren in astronomische objecten te behouden. Het doel is om zwakke structuren die opgebroken zijn door ruis samen te voegen. Daarnaast proberen we compacte, heldere objecten die vlak bij elkaar staan *niet* samen te voegen tot één object. Experimenten laten zien dat het mogelijk is om met z.g. connectiviteits-maskers gebaseerd op *visceuze* morfologische dilatie en erosie de gewenste selectieve clustering tot stand kunnen brengen. Diverse parameter instellingen zijn getest, en de mate van correcte en incorrecte clustering werd bepaald. In deze toepassing werkte visceuze filters beter dan normale structurele dilatie en erosie.

Echter, zelfs met een verbeterde clustering, is object segmentatie verre van triviaal. Een max-tree gebaseerde segmentatie methode wordt beschreven in Hoofdstuk 3. Dit combineert connected filters van beeld componenten, met standaard connectiviteit, met statistische kennis van het type ruis in astronomische beelden. De nieuwe methode gebruikt een statistische test gebaseerd op het z.g. *power* attribuut, dat evenredig is met de integraal van het kwadraat van de beeldintensiteit binnen een beeldcomponent, voor iedere node in de max-tree. Groepen componenten die niet met een zekere waarschijnlijkheid, gegeven de statistische verdeling van het power attribuut door het ruismodel kunnen worden verklaard, worden behouden als objecten. Deze aanpak is een verbetering ten opzichte van de vaste drempelwaarde die gebruikt wordt in het programma Source Extractor, een state-of-the-art tool voor astronomische beeldbewerking. Source Extractor faalt vaak in de detectie van uitgebreide, zwakke objecten. De nieuwe methode heeft ook een voordeel in een betere schatting van de achtergrond, die minder afwijking vertoont in aanwezigheid van objecten in vergelijking met de methode in Source Extractor.

De methoden zijn vergeleken op data uit de Sloan Digital Sky Survey Data Release 7.

In Hoofdstuk 4, wordt getoond dat er een grote noodzaak is voor een parallel algoritme voor het bouwen van max-trees, gegeven de hoge resolutie en het hoge dynamische bereik van astronomische en satelliet data. Bestaande parallelle methoden zijn getest en het werd aangetoond dat deze alleen effectief zijn tot ongeveer 16 bits per pixel. Een nieuw parallel algoritme geschikt voor z.g. shared-memory parallelle computers werd ontwikkeld, door het combineren van top-down en bottom-up algoritmes in een twee-staps process. Het algoritme is uitgebreid getest op gesimuleerde en echte beelden, zowel in 2D als in 3D.

Het parallelle algoritme en de statistische max-tree gebaseerde segmentatie methode worden samen gebruikt in Hoofdstuk 5. Experimenten met het segmenteren van hoge-resolutie 3D radio volumes van de emissie van melkwegstelsels, zoals waargenomen door de WSRT (Westerbork Synthesis Radio Telescoop) zijn uitgevoerd. De resultaten zijn nog beperkt maar veelbelovend, en moedigen verder onderzoek naar het gebruik van max-trees en statistische filters op soort data aan.

Object klassificatie is ook onderzocht. Het doel in dit geval is het onderscheiden van melkwegstelsels die inderdaad met elkaar interacteren (botsingen) van melkwegstelsel die toevallig in dezelfde kijkrichting staan aan de hemel (projecties). Het gebruik van meerdere lokale 2D patroonspectra als data vectoren voor het klassificeren van deze astronomische objecten lijkt veelbelovend. Patroonspectra zijn univariate of multivariate histogrammen van beeldstructuren, waarbij iedere categorie in het histogram aangeeft hoeveel beeldinhoud binnen een bepaalde vorm- of grootteklasse ligt.

Patroonspectra en een methode om deze parallel te berekenen worden geïntroduceerd in Hoofdstuk 6, in een toepassing op satelliet beelden. Reken tijden liepen terug van aan paar minuten tot 8-15 seconden op de gebruikte gigapixel beelden. In plaats van de patroonspectra op gehele beelden te berekenen introduceren we de mogelijkheid om *lokale* patroonspectra te berekenen in Hoofdstuk 7. In dit geval worden patroonspectra berekend voor ieder gesegmenteerd object, in plaats van op het gehele beeld. De parallelle methode uit Hoofdstuk 6 is aangepast voor 2D lokale patroonspectra. Een 2D patroon spectrum van oppervlak versus invariante momenten wordt gegenereerd voor ieder melkwegstelsel en de resulterende data vectoren dienen als invoer voor een classificatie programma. Bij gebruik van de C4.5 beslissingsboom methode met z.g. bagging, werd in 80% van de objecten de correcte classificatie gevonden.

Een nieuw algoritme voor het berekenen van z.g. attribuutfilters op basis van een *visceuze* hyperconnectie wordt beschreven in Hoofdstuk 8, als uitbreiding van het nog beperkte arsenaal filters gebaseerd op hyperconnectiviteit. Het geeft een mogelijke oplossing voor het z.g. lekkage probleem van connected filters. Vaak kunnen ruispixels of andere ongewenste structuren objecten aan elkaar plakken,

door paden van verbonden pixels te creëren die een verbinding vormen tussen objecten die eigenlijk gescheiden moeten worden. De voorgestelde visceuze hyperconnectiviteits klasse staat overlap van objecten to zonder ze samen te voegen, wat het lekkage effect stopt, zolang het pad tussen de structuren smaller is dan een bepaalde waarde. Het nieuwe algoritme berekend attribuut filter gebaseerd op visceus-hypersamenhangende componenten. Eerst wordt een max-tree van een geërodeerde versie van het beeld berekend met een standaard algoritme. De attribuut-waarden in iedere knoop van deze bomen worden aangepast met behulp van de pixelwaarden die verschillend zijn in het geërodeerde beeld en een dilatatie ervan. Dit laatste onderdeel is nog niet op astronomische beelden toegepast, maar bied wel een nieuwe manier van beelden filteren. Ook andere toepassingen moeten nog worden onderzocht.

Het werk gepresenteerd in dit proefschrift draagt bij aan een dieper inzicht in hoe methoden en algoritmen uit de mathematische morfologie kunnen worden toegepast om problemen op het gebied van beeldsegmentatie en classificatie, met name in astronomische beelden. Uit veel van de onderzochte technieken en nieuwe methoden blijkt inderdaad goed de toepasbaarheid van de theorie in de praktijk.