



Universiteit
Leiden
The Netherlands

Pattern Recognition in High-Throughput Zebrafish Imaging

Nezhinsky, A.E.

Citation

Nezhinsky, A. E. (2013, November 21). *Pattern Recognition in High-Throughput Zebrafish Imaging*. Retrieved from <https://hdl.handle.net/1887/22286>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/22286>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/22286> holds various files of this Leiden University dissertation

Author: Nezhinsky, A.E.

Title: Pattern recognition in high-throughput zebrafish imaging

Issue Date: 2013-11-21

Dutch Summary

In de levenswetenschappen wordt gebruik gemaakt van zogenaamde High Throughput methoden; deze bestaan uit experimenten waarbij een grote hoeveelheid data wordt gegenereerd om zo bepaalde eigenschappen te kunnen vaststellen. De uitvoering van deze methoden verschilt per toepassing. In dit proefschrift wordt met name aandacht besteed aan de High Throughput methoden waarin een effect wordt gemeten aan de hand van beelden van objecten, high-throughput betekent dat er grote hoeveelheden beelden worden gegenereerd en er dus veel objecten moeten worden herkend om uiteindelijk een uitspraak te kunnen doen over eigenschappen van deze objecten.

Een voorbeeld van een High Throughput proces is afgebeeld in Figuur 1.1. Het specifieke doel van het onderzoek beschreven in dit proefschrift is het vaststellen van robuuste methoden voor patroon herkenning binnen een High Throughput analyse, deze methoden te onderzoeken alsmede te valideren. Binnen het onderzoek beschreven in dit proefschrift wordt Patroon Herkenning gezien in zowel de context van object herkenning als ook in de context van data mining en beide aspecten worden in dit proefschrift aangehaald.

We hebben reeds aangegeven dat de High-Throughput context in dit onderzoek vooral veel beelden gegenereerd, deze beelden, microscoop beelden in het bijzonder, zijn dan de input voor analyse. Het is zaak in deze beelden de objecten kunnen herkennen en analyseren. In het veld van de beeldanalyse wordt daarvoor segmentatie gebruikt; dit is de verdeling van een beeld in interessante objecten en de achtergrond. Wij als mensen segmenteren doorlopend, dat wil zeggen scheiden van voor- en achtergrond, voor een computer is dat echter veel lastiger. Dit moet worden geleerd. Segmentatie technieken kunnen worden gescheiden in twee hoofdgroepen: bottom-up segmentatie versus top-down segmentatie. Bottom up segmentatie is gebaseerd op het vinden van gelijkwaardige eigenschappen van de objecten of de achtergrond om vervolgens deze van elkaar te scheiden. Echter, vaak is een scheiding van een beeld in de voor-en achtergrond regio's niet voldoende om het voorwerp van specifieke interesse te vinden. Daarom is daarnaast een aanpak vereist die gebaseerd is op patroon herkenning, een zogenaamde top-down benadering.

In hoofdstuk 2 presenteren we twee benaderingen voor de High-Throughput context die is bestudeerd in het kader van dit proefschrift: i.e. het segmenteren en annoteren van de objecten die onderwerp van het experiment zijn. Voor toepassing wordt gebruik gemaakt van de zebravis als model systeem, en daarom leggen we de nadruk op beelden die zebravis embryos bevatten. We onderscheiden twee stappen: een voorbereidingstap waarbij de achtergrond wordt gescheiden van de voorgrond (bottom-up) en de Template-Matching stap, dat bestaat uit het zoeken naar een vervormbaar sjabloon in het voorbereide beeld (top-down), een zogenaamde Deformable Template Matching

methode. Voor deze Deformable Template Matching stap beschrijven we twee methoden die gebaseerd zijn op het silhouet van een gemiddeld object als een sjabloon. Vervolgens nemen we aan dat het beeld een of meerdere objecten bevat die een misvormde of verschoven afbeelding zijn van het sjabloon en we beschrijven een functie waarmee wordt aangegeven in hoeverre een geïdentificeerde vorm in het beeld overeenkomt met het vervormde sjabloon. Het zoeken naar een correcte deformatie is een complex proces, daarom passen we verschillende vormen van optimalisatie toe. Tevens bestuderen we de mogelijkheden om het objecten die als resultaat gevonden worden te normaliseren om zo een beter vergelijkbaar resultaat te verkrijgen. Het algoritme is geëvalueerd met zowel synthetische beelden als met beelden van zebra-embryos.

In hoofdstuk 3 presenteren we een methode van Deformable Template Matching die op een andere wijze verder is uitgewerkt. Het object waar we naar zoeken wordt nu, in plaats van een silhouet, gerepresenteerd door middel van zogenaamde ankerpunten. Dit is robuster en meer natuurgetrouw, omdat de vorm die gezocht wordt niet meer een van te voren vastgestelde grootte dient te hebben, maar als een schaal onafhankelijke beschrijving wordt gerepresenteerd. De voorgestelde methode kan volledig worden geautomatiseerd, en blijkt sneller en nauwkeuriger dan de methodes die in hoofdstuk twee zijn gepresenteerd. We hebben het algoritme geëvalueerd met beelden van zebra-embryos en xenopus-embryos.

In hoofdstuk 4 beschrijven we hoe de beschreven algoritmen zijn geïntegreerd in een software-pakket, i.e. ZFA. Dit pakket wordt gebruikt voor de analyse van High-Throughput data. We presenteren een interface die aansluit bij de workflow van de gebruikers. De gebruikers zijn in dit geval onderzoekers in de levenswetenschappen. De verschillende processen die vanuit de workflow ontstaan zijn, zijn in de software opgenomen als componenten: de voorbewerking, de interface, de segmentatie en de gegevensverwerking. Het High-Throughput toepassingsgebied waar de software wordt gebruikt is het onderzoek naar het doorgronden van infectie-mechanismen die ten grondslag liggen aan Tuberculose. Tuberculose is een ernstige ziekte en een aanzienlijk deel van de wereldbevolking is geïnfecteerd. Het is echter nog steeds moeilijk om een effectieve behandeling te vinden vanwege de resistentie van bacteriën. Tuberculose wordt veroorzaakt door de bacterie - *Mycobacterium tuberculosis*. Voor het onderzoek naar infectie-mechanismen wordt echter gebruik gemaakt van een nauw verwante soort - *Mycobacterium marinum*. Door gebruik te maken van een model-organisme, de zebra-embryo, kan het infectieproces worden bestudeerd in een groot aantal individuen. De zebra-embryos moeten dan in beelden worden geïdentificeerd en vervolgens verder worden geanalyseerd. Onze software doet dat.

In hoofdstuk 5 worden de resultaten vergeleken die verkregen zijn uit het toepassen van ZFA in een lopend onderzoek betreffende de infectie van zebra-embryos met *Mycobacterium marinum*. Een dergelijke infectie wordt gekenmerkt door de aanwezigheid van zogenaamde granulomen. Granulomen zijn clusters van immuuncellen en bacteriën die wijzen op succesvolle infectie van de bacterie. In het onderzoek worden de bacteriën gevisualiseerd met fluorescent label. Om meer inzicht te krijgen op de progressie van de infectie is het nodig om de spreiding ervan te analyseren in de gastheer, de zebra-embryo, gedurende een bepaalde periode. De resultaten worden onder meer gebruikt om vast te

stellen welke genen verantwoordelijk zijn voor infectie. Eerst wordt, via het algoritme uit hoofdstuk 3, de vorm van de zebra-vis embryo's gelokaliseerd. Vervolgens worden verschillende kenmerken van de infectie geanalyseerd door het deel van de software dat verantwoordelijk is voor de gegevensverwerking. Onze aanpak maakte het mogelijk om langs geautomatiseerde weg het gedrag van de infectie per zebra-vis te analyseren. Om te bepalen welke genen van *Mycobacterium marinum* zijn betrokken bij vorming van granulomen zijn 1000 willekeurige mutanten van deze bacteriën geanalyseerd en daarbij zijn de mutanten geselecteerd die zebra-vis larven niet ziek maakten. Op deze manier zijn 30 mutanten geïdentificeerd. We hebben onderzocht of er een patroon aanwezig is in de spreiding of grootte van granuloma clusters in bepaalde weefsels en hoe het gedrag verschilt per mutant ten opzichte van larven besmet met de wild-type bacteriën.

Tenslotte worden in hoofdstuk 6 resultaten van de voorgaande hoofdstukken samengevat. We beginnen met de segmentatie stap (Pattern Recognition in Images). We hebben drie verschillende methoden beschreven. We vergelijken de prestaties van de drie methoden. Alle methoden werden getest op dezelfde dataset. Resultaten laten een verbetering in de prestaties zien. Het Ankerpunt gebaseerde (hoofdstuk 3) methode gaf de beste resultaten. Dit algoritme zal daarom worden gebruikt voor de analyse en is deel van de software.

