WORDT
NIET UITGELEEND

Grayscale Feature Extraction

toegepast op nummerbordherkenning

Rijksuniversiteit Groningen
Bibliotheek Wiskunde & Informatica
Postbus 800
9700 AV Groningen
Tel. 050 - 363 40 01

Afstudeeropdracht
Eelco Hoekstra

WORDT
NIET UITGELEEND

Gegevens

Student	Eelco Hoekstra Hofstraat 17a 9712 JA Groningen eelco@dikkeprima.com 06-19234713
Studie	Rijksuniversiteit Groningen Informatica
Specialisatie	Computional Intelligence and Intelligent Systems
Afstudeerdocent	Dr. Ir. J.A.G. Nijhuis
Afstudeerdatum	28 november 2003

Rijksuniversiteit Groningen
Bibliotheek Wiskunde & Informatica
Postbus 800
9700 AV Groningen
Tel. 050 - 363 40 01

Inhoudsopgave

FIGUREN 4
TABELLEN 5
ABSTRACT 6
ENGLISH 6
NEDERLANDS 6
INLEIDING 7

1 NUMMERBORDHERKENNING 9
1.1 CLPR: CAR LICENSE PLATE RECOGNITION 9
1.2 DOMEINAFBAKENING 10
2 ONDERZOEKSVRAGEN 13
2.1 KARAKTERHERKENNING 13
2.2 CENTRALE VRAAG 14
2.3 DEELVRAGEN 14
3 THRESHOLDING 16
3.1 LEESBAARHEID 16
3.2 METHODE 16
3.2.1 Thresholdmethoden 16
3.2.2 Klassen 18
3.2.3 Beoordeling 19
3.3 RESULTATEN 20
3.4 CONCLUSIES DEELVRAAG 1 21
3.5 CONTEXT: NUMMERBORDHERKENNING 22
4 GRAYSCALE KARAKTER HERKENNING 24
4.1 FEATURE EXTRACTORS 24
4.1.1 Verkenning 25
4.1.2 Vergelijking en selectie 28
4.1.3 Tussenbeschouwing 29
4.2 IMPLEMENTATIE 30
4.2.1 Inleiding 30
4.2.2 Topografische features 30
4.2.3 Pseudo Zernike momenten 38
4.2.4 Overzicht van features 40
4.2.5 Principal Component Analysis 41
4.2.6 Tussenbeschouwing 42
4.3 NEURALE CLASSIFICATIE 43
4.3.1 Inleiding 43
4.3.2 Opzet 43
4.3.3 Strategie 44
4.3.4 Methode 45
4.3.5 Resultaten 46
4.3.6 Constateringen 51
4.3.7 Tussenbeschouwing 52
4.4 CONCLUSIES DEELVRAAG 2 53
4.5 CONTEXT: NUMMERBORDHERKENNING 54
5 EVALUATIE 56
5.1 RELEVANTE ASPECTEN 57
5.2 CONCLUSIES HOOFDVRAAG 58
5.3 CONTEXT: NUMMERBORDHERKENNING 60

REFERENTIES 61
APPENDICES 62
APPENDIX A: RESULTATEN VAN CLASSIFICATIE NAAR THRESHOLD 62
APPENDIX B: NUMMERBORDEN NA THRESHOLDING EN TOPOLABELLING 64
APPENDIX C: CODE VOOR PSEUDO ZERNIKE MOMENTEN 65
APPENDIX D: TESTRESULTATEN WINNENDE NEURALE NETWERKEN 67
APPENDIX E: VERGELIJKING BINARY EN GRAYSCALE FEATURES 71

Figuren

Figuur 1: Het nummerbord herkenningproces.. 9
Figuur 2: Karakterherkenningsproces van CLPR. 10
Figuur 3: Een niet herkend nummerbord met zes goed geïsoleerde karakters. 11
Figuur 4: Het karakterherkenningsproces. 13
Figuur 5: Beoordeling van karakters met behulp van drie thresholdmethoden. 17
Figuur 6: Het effect van postprocessing na thresholding. 18
Figuur 7: Klassen voor de beoordeling van een karakter op leesbaarheid 18
Figuur 8: Voorbeelden voor beoordeling naar thresholdklassen. 19
Figuur 9: Vier criteria ter toetsing van feature extraction methoden 24
Figuur 10: Een diagonale Gabor filter gebruikt bij convolutie. 25
Figuur 11: De letter S met verschoven grijswaardenbereik. 27
Figuur 12: Het toepassingstraject van topografische features. 30
Figuur 13: Alle mogelijke topografische labels 31
Figuur 14: Intuïtieve versus Lee en Kim labelling. 32
Figuur 15: De gebruikte labels bij topografische pixel labelling. 32
Figuur 16: Twee oriëntaties bij pixel-labelling 33
Figuur 17: Labelling in twee verschillende richtingen. 33
Figuur 18: Het effect van filteren van een beeld 34
Figuur 19: Het effect van thresholding op labelling. 35
Figuur 20: Postprocessing met behulp van union find 35
Figuur 21: De vorm van de achtergrond. 36
Figuur 22: Verdeling in zones levert een feature vector op 37
Figuur 23: Horizontale projectie van voorgrondpixels. 37
Figuur 24: Implementatie traject voor pseudo Zernike momenten. 38
Figuur 25: Verticale en Horizontale Sobel-kernels 38
Figuur 26: Illuminatie invaratie middels de richting van gradienten van het beeld. 39
Figuur 27: Het complete beeld wordt geschaald op de eenheidscirkel. 40
Figuur 28: Trainingprogress van respectievelijk Zoning en Zernike 49
Figuur 29: Trainingprogress Zoning en Zernike na heropdeling datasets. 49
Figuur 30: RDOplots van de complete testset op vier soorten netwerken. 50

Tabellen

Tabel 1: Uitkomsten klassificatie van leesbaarheid na thresholding per karakter. 20
Tabel 2: Contextuele verdeling naar herkenning 22
Tabel 3: Uitkomsten classificatie per nummerbord in aantallen en verhoudingen. 23
Tabel 4: Samenvatting vergelijking van mogelijkheden. 28
Tabel 5: Overzicht van features 40
Tabel 6: Overzicht van features en de verkregen principal components 41
Tabel 7: Prestatie van verschillende trainmethoden. 46
Tabel 8: Vergelijking tussen prestaties topografische features met en zonder achtergrondinformatie. 47
Tabel 9: Resultaten van pseudo Zernike features met verschillende aantallen hidden neuronen en trainmethoden. 47
Tabel 10: Prestaties van topografische features met verschillende trainmethoden en hidden neuronen 48
Tabel 11: Tabel met testresultaten van de complete testset bij verschillende criteria. 51
Tabel 12: Verschil tussen prestaties binary en grayscale feature extraction. 55
Tabel 13: Aantal herkende, verworpen en foutief herkende nummerborden bij gebruikmaking van zowel binary als grayscale feature extraction. 55

Abstract

English

This research examines the effects of grayscale feature extraction on, primary, the recognition of characters and, secondary, the recognition of license plates. Examination of the influence of thresholding on the recognition of characters displays the possible advantage of grayscale feature extraction. Two methods are selected to do this job, pseudo Zernike moments and topographic labeling of pixels. Features from zoning topographic labels show the best result in use with neural networks. The robustness of topographic features is observed when compared to classification with binary features. In the context of license plate recognition this means incorrect classification can be reduced while preserving correct classifications.

Nederlands

In deze scriptie wordt onderzoek gedaan naar de gevolgen van toepassing van grayscale feature extraction op primair de herkenning van karakters en secundair de herkenning van nummerborden. Uit onderzoek naar de invloed van thresholding op de herkenning van karakters blijkt dat het toepassen van grayscale features daadwerkelijk nut heeft. Vervolgens worden twee methoden gebruikt om grayscale features te extraheren van karakterbeelden. Dit betreft pseudo Zernike momenten en topografische pixellabeling. Features verkregen middels zoning van topografisch gelabelde beelden, leveren de beste herkenning bij classificatie met neurale netwerken. In vergelijking met classificatie door features van binaire beelden blijkt, dat topografische features een robuustere methode is om na thresholding verminderd leesbare karakters correct te classificeren. Dit betekent voor de herkenning van nummerborden dat het het aantal misclassificaties verminderd kan worden terwijl het aantal correcte classificaties gelijk blijft.

Inleiding

Het begrijpen en herkennen van informatie op beelden, bijvoorbeeld foto's, of van documenten kent vele praktische toepassingen. Het is dan ook al tientallen jaren onderwerp van onderzoek. [1] Zodoende kunnen tegenwoordig postcodes op brieven worden herkend, handgeschreven tekst en ingescande documenten kunnen worden gelezen.

Een andere toepassing is het herkennen van nummerborden op foto's van voertuigen. [2] Het kenteken van bijvoorbeeld een auto of een motor kan zodoende automatisch worden herkend. Een kenteken bestaat uit een combinatie van verschillende letters en cijfers. Om een kenteken correct te herkennen zullen alle individuele karakters op een nummerbord moeten worden herkend.

Een karakter op een beeld kan worden onderscheiden doordat het karakter een andere kleur of structuur heeft dan de achtergrond waarop het zich bevindt. Deze tekst is leesbaar omdat de zwarte karakters een witte achtergrond hebben. Bij de herkenning van karakters wordt van deze bilevel eigenschap gebruik gemaakt, zowel door mens als machine.

Vanwege deze bilevel eigenschap van een karakter wordt binairisatie van het beeld toegepast. Dit houdt in dat het originele beeld van een karakter wordt getransformeerd naar een beeld met maar twee kleuren, zwart en wit. Dit proces staat bekend als thresholden. Hierbij is het de bedoeling dat na thresholding de zwarte pixels de voorgrond, het karakter, en de witte pixels de achtergrond representeren. Van dit zwart-wit beeld worden karakteristieken (features) ontleend (extraction). Hiervoor is een heel scala aan methoden beschikbaar, de één succesvoller dan de ander. [3] Met behulp van deze features worden de beelden vervolgens geclassificeerd.

Een foto kan echter onduidelijk zijn of ruis bevatten. Hierdoor is het mogelijk dat een beeld na thresholding minder herkenbaar wordt, zowel voor mens als machine. De invloed hiervan op de herkenning van karakters is niet direct meetbaar, omdat de herkenning mede afhangt van de robuustheid van de feature extractors. De invloed van thresholding op de leesbaarheid van karakters is echter wel te bepalen. Door nu onderzoek te doen naar de invloed van thresholding op de representatie van karakters wordt inzicht verschaft in het herkenningsproces.

Het originele beeld is opgebouwd uit pixels. Een pixel van dat beeld heeft een grijswaarde in een bepaald bereik (grayscale). Naast features van binaire beelden, kunnen ook features van het originele beeld worden gebruikt bij classificatie. Dat brengt ons bij het doel van dit onderzoek. Het doel van deze scriptie is te onderzoeken in hoeverre grayscale feature extraction een toegevoegde waarde heeft bij het herkennen van karakters en nummerborden.

De hoofdvraag in dit onderzoek volgt dan uit de doelstelling: "Wat is het gevolg van het gebruik van het originele beeld bij feature extraction op de herkenning van karakters?" De beantwoording van de centrale vraag vindt plaats middels de beantwoording van twee deelvragen. In de eerste plaats: "Wat is de invloed van thresholding op de representatie van karakters?" In de tweede plaats: "Wat is het resultaat van het gebruik van grayscale features op de herkenning van karakters?" Dit onderzoek zal in vier etappes, verdeeld over vier hoofdstukken plaatsvinden.

In hoofdstuk 1 zal het domein van nummerbordherkenning nader worden onderzocht. Slechts enkele stappen uit het proces hebben zijn relevant voor de daadwerkelijke herkenning van de karakters van de nummerborden. Deze stappen zullen worden besproken. Dit leidt uiteindelijk tot een afgebakend domein van bruikbare data.

Hoofdstuk 2 omvat de vragen die in dit onderzoek beantwoord gaan worden. De hoofdvraag zal hier worden besproken. Deze centrale vraag wordt opgedeeld in twee deelvragen. Hierbij zullen aannames worden gemaakt en verantwoord. De deelvragen worden afzonderlijk in twee verschillende hoofdstukken beantwoord, alvorens de centrale vraag zijn beantwoording zal vinden.

De invloed van thresholding op de representatie van karakters staat centraal in hoofdstuk 3. Hierin wordt mede aandacht besteed aan de definitie van leesbaarheid van een karakter. Vervolgens wordt dit deelonderzoek uitgevoerd. Aan de verkregen resultaten zullen conclusies worden verbonden, zowel voor karakter- als voor nummerbordherkenning. Dit hoofdstuk beantwoordt de eerste deelvraag.

De prestaties van grayscale feature extraction worden onderzocht in hoofdstuk 4. Dit deelonderzoek is opgedeeld in 3 fasen. Eerst zullen methoden worden gezocht die effectief features van grayscale beelden kunnen ontwaren. Deze feature extractors worden vervolgens toegespitst op het domein door specifieke implementatiekeuzes te maken. Als laatste zullen de prestaties van deze feature extractors worden getest. Deze tests vinden plaats middels neurale classificatie. Aan de resultaten van de test worden conclusies verbonden. Dit hoofdstuk beantwoordt de tweede deelvraag.

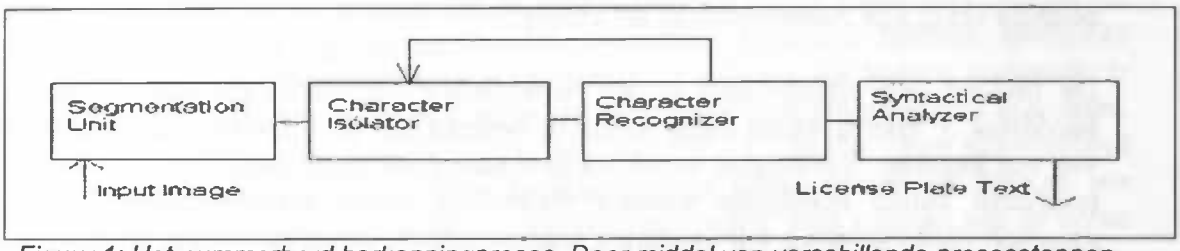
Nadat alle deelonderzoeken zijn gedaan kan het hele onderzoek worden geëvalueerd. Dat gebeurt in hoofdstuk 5. De relevante aspecten zullen op een rijtje worden gezet en de fundamentele keuzes zullen nogmaals worden aangestipt. Tezamen met de resultaten van het onderzoek kan de centrale onderzoeksvraag weloverwogen worden beantwoord.

1 Nummerbordherkenning

Het werkdomein wordt aangeleverd in de vorm van een reeds werkzaam nummerbordherkenningssysteem. Dit proces kent verscheidene stappen die uiteindelijk moeten leiden tot een correct herkend nummerbord. Deze stappen worden kort besproken. Vervolgens wordt het domein ingeperkt tot een handbare omgeving waarin uitspraken over vooral karakterherkenning het meest tot hun recht komen.

1.1 CLPR: Car License Plate Recognition

Het uitgangspunt van het nummerbordherkenningssysteem is een foto met daarop een voertuig met een Nederlands kenteken. Het doel is om het nummerbord van het gefotografeerde voertuig te herkennen. Het nummerbord beslaat maar enkele procenten van het totale beeld. Om tot een correcte herkenning te komen is het proces opgedeeld in verschillende stappen. [2] (Zie figuur 1)

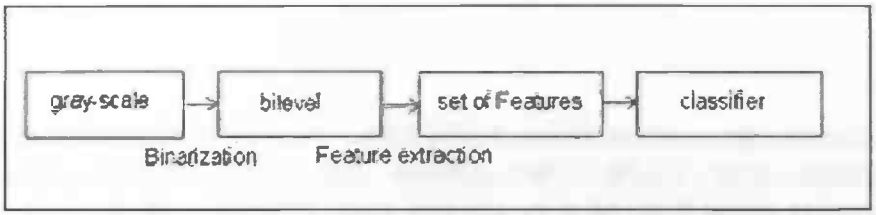


Figuur 1: Het nummerbord herkenningproces. Door middel van verschillende processtappen wordt van een foto van een voertuig het kenteken gevonden.

Segmentatie
Voordat verdere stappen ondernomen kunnen worden zal eerst de locatie van het nummerbord moeten worden gevonden. Door middel van verschillende morfologische operaties wordt het nummerbord gesegmenteerd.

Karakter isolatie
Hier wordt getracht de zes karakters van het nummerbord te isoleren.

Karakter herkenning
Elk geïsoleerd karakter wordt individueel herkend. Hiertoe worden verschillende methoden aangewend om die herkenning te bereiken. Elke methode volgt in ieder geval het stramien uit figuur 2.



Figuur 2: Karakterherkenningsproces van CLPR. Van elk grayscale beeld worden na thresholding features ter classificatie aangeboden.

In dit stramien wordt het grayscale beeld eerst binair gemaakt middels thresholding. Vervolgens worden van dit binaire beeld features ontwaard. Hierin onderscheiden zich de verschillende methoden. Deze features worden aangeboden aan een getraind neuraal netwerk. Mocht de herkenning van een karakter niet naar tevredenheid zijn verlopen dan bestaat de mogelijkheid om een karakter te relocaliseren.

Syntactische Analyse
Hierin worden de zes karakters en hun herkenningsresultaten weer bij elkaar gevoegd. De syntax van een (Nederlands) nummerbord is a-priori kennis. Het bestaat altijd uit drie combinaties van twee letters of cijfers. Deze kennis heeft twee gevolgen. Ten eerste kunnen nummerborden niet alleen worden afgewezen vanwege hun lage herkenningsscore, maar als extra controle kunnen ze ook op hun syntax worden gecontroleerd. Ten tweede ontstaat zo de mogelijkheid om in de herkenning van karakters rekening te houden met de ambivalente combinatie van vier verschillende cijfer-letter paren. De ambivalentie komt tot uitdrukking in de gelijkenis van de karakters. Dat zijn *8-B, 2-Z, 0-D, 5-S.* Afhankelijk van de herkenning van het andere karakter uit een koppel kan dan de juiste waarde worden toegewezen. Daartoe zal één van de karakters uit een koppel sowieso moeten worden geclassificeerd naar een volledige output, zodat de keuze van het andere karakter daadwerkelijk gemaakt kan worden.

1.2 Domeinafbakening

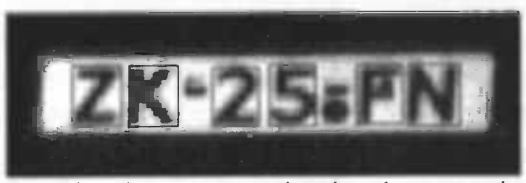
De afbakening van het domein gebeurt op twee niveaus. Ten eerste vindt een afbakening plaats op het niveau van CLPR. Niet alle stappen worden in overweging genomen. Ten tweede vindt een afbakening plaats op het niveau van de data. Niet alle nummerborden worden geselecteerd voor gebruik. Door deze keuzes ontstaat een context waarin uitspraken over karakterherkenning het meest tot hun recht komen, maar waarin de gevolgen voor nummerbordherkenning duidelijk omschreven moeten worden.

Omdat dit onderzoek zich richt zich op de herkenning van karakters, is de manier van segmentatie van het nummerbord en de manier van isolatie van de karakters irrelevant. Als uitgangspunt wordt dan ook gekozen voor nummerborden waarvan de karakters correct zijn geïsoleerd. De karakterherkenner zal in hoofdstuk 2 verder worden doorgenomen. Van de syntactische analyse wordt in dit onderzoek geen gebruik gemaakt. Alleen het gegeven dat ambivalente karakters onder voorwaarden naar 1 klasse kunnen worden geclassificeerd, wordt in dit onderzoek meegenomen.

Als onderzoeksdata worden alleen Nederlandse nummerborden aangewend die niet zijn herkend door CLPR. Dat gebeurt om twee redenen. Resultaten van nummerborden die al herkend zijn, zijn niet meer relevant voor dit onderzoek, omdat hierbij het thresholden inherent geen invloed heeft gehad op mislukte herkenning. Daarnaast zijn zonder de reeds herkende nummerborden de verbeteringen van nieuwe aan te dragen methoden direct meetbaar.

Aan deze nummerborden worden vervolgens nog 3 eisen gesteld. Ten eerste wordt de eis gesteld dat de nummerborden voor het menselijk oog leesbaar zijn. Hiervoor wordt gekozen omdat het automatisch herkennen van karakters ver achter blijft bij de prestaties van de mens op dat gebied. Vooralsnog kan niet worden verondersteld dat een huidig systeem dat zal overtreffen. Ten tweede wordt uitgegaan van een nummerbord waarvan de 6 karakters goed zijn geïsoleerd. Dit houdt in dat de isolatie het karakter nauw omvat. Hier is voor gekozen omdat verbeteringen omtrent karakterisolatie geen onderwerp van onderzoek zijn en omdat correcte herkenning van 1 goede isolatiepoging volstaat bij herkenning van een nummerbord. Voor een voorbeeld van goed geïsoleerde karakters wordt verwezen naar figuur 3.

Als laatste wordt de eis gesteld dat de nummerborden alleen karakters bevatten die frequent voorkomen op nummerborden. Sommige karakters hebben specifieke doeleinden of zijn gemakkelijk te verwarren met anderen. Ze komen zodoende te weinig of niet op nummerborden voor om er in dit onderzoek rekening mee te houden. Het betreft de letters 'A', 'C', 'E' , 'I', 'M', 'O', 'Q', 'U' en 'W'. Een nummerbord bevat zelden meer dan 1 van deze karakters. Deze laatste eis wordt daarom iets versoepeld. De karakters van deze nummerborden worden, mits voldaan aan de eerste twee eisen, wel meegenomen bij onderzoek naar thresholding en herkenning. Hiervoor is gekozen om de dataset niet te veel te laten slinken. Echter, er kan geen rekening worden gehouden met deze nummerborden bij de herkenning. Voor frequenties van karakters in de dataset wordt verwezen naar appendix A.



Figuur 3: Een niet herkend leesbaar nummerbord met zes goed geïsoleerde karakters.

Deze domeinafbakening leidt tot de volgende context. Er wordt onderzoek gedaan naar karakterherkenning in het domein van nummerbordherkenning. De eisen die aan nummerborden worden gesteld versmallen de selectie van de 960 beschikbaar gestelde, niet herkende nummerborden naar 556. Deze selectie is handmatig uitgevoerd.

De geselecteerde 556 kentekens bevatten 556 x 6 = 3336 karakters. Het zijn deze karakters waarmee het onderzoek zal plaatsvinden. Deze set van karakters is echter niet representatief voor de set van alle karakters van nummerborden vanwege het feit dat de nummerborden niet zijn herkend. Als de resultaten van het onderzoek aan het eind van het verslag in context van CLPR worden geplaatst zal deze informatie daarin meegenomen moeten worden.

Daarnaast is de mate van individuele herkenning van een karakter door CLPR onbekend. De resultaten omtrent karakterherkenning die in dit onderzoek vrijkomen zijn daarom moeilijk vergelijkbaar. Toch kunnen verbeteringen makkelijk worden getoetst doordat het domein alleen bestaat uit niet herkende nummerborden.

Samenvattend kan worden gesteld dat eerst de context van nummerbordherkenning zal worden losgelaten om onderzoek te doen naar de individuele karakterherkenning, om vervolgens de verkregen resultaten weer in context te plaatsen. Daarom wordt elk hoofdstuk vanaf nu afgesloten met een paragraaf waarin de gevolgen van het onderzoek tot dan toe op de herkenning van nummerborden wordt omschreven. Hierbij wordt de kanttekening gemaakt dat alle weergegeven visualisaties nummerborden zullen zijn, al dan niet met de karakters individueel geïsoleerd.

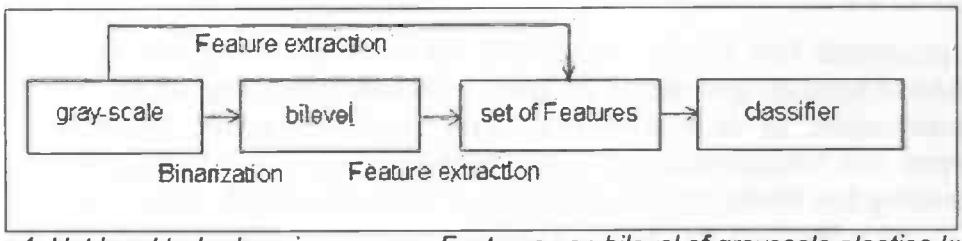
Recapitulerend kan worden vastgesteld dat er in dit onderzoek goed geïsoleerde karakters worden gebruikt van niet herkende, doch menselijk goed leesbare nummerborden. In de context van CLPR houdt dat in dat wordt onderzocht in hoeverre een vereenvoudiging in het huidige systeem van karakterherkenning een verbetering kan betekenen in het herkennen van niet herkende nummerborden.

2 Onderzoeksvragen

Nu het domein vaststaat kunnen de onderzoeksvragen nader worden gespecificeerd. Daartoe zal eerst het proces van karakterherkenning verder onder de loep worden genomen.

2.1 Karakterherkenning

De karakterherkenner neemt als input een geïsoleerd karakter en heeft als doel dit karakter correct te classificeren. Dit proces kent enkele stappen. Uitgangspunt is een geïsoleerd grijswaardenplaatje. Een plaatje van een karakter is bilevel van opzet, namelijk een donkere voorgrond op een lichtere achtergrond. Vanwege deze eigenschap wordt thresholding toegepast. Er zijn verschillende methoden om dat te bewerkstelligen. [4] Dat levert als resultaat een binair plaatje. Van een binair plaatje worden vervolgens features ontleend. Hiervoor zijn verschillende methoden voorhanden. Zie [3] voor enkele voorbeelden. Deze features kunnen op verschillende manieren worden geclassificeerd.



Figuur 4: Het karakterherkenningsproces. Features van bilevel of grayscale plaatjes kunnen worden gebruikt bij karakterclassificatie.

Feature extraction is het proces waarbij uit data relevante karakteristieken worden geselecteerd. Deze features worden gepresenteerd als een cijferreeks. Om feature extraction succesvol te gebruiken bij classificatie zullen features van data uit gelijke klassen (ongeveer) gelijke features moeten opleveren.

In plaats van binaire beelden als basis voor features extraction te gebruiken, kunnen ook grayscale beelden als basis hiervoor dienen. (zie figuur 4) Dit zou een vereenvoudiging in het aantal stappen van het herkenningsproces inhouden. Aangezien de prestaties van deze aanpak onbekend zijn in dit domein, komen we bij de centrale vraag in dit onderzoek.

2.2 Centrale vraag

De mogelijkheid van het gebruik van features van grayscale beelden brengt ons bij de centrale vraag van dit onderzoek. Zoals reeds eerder gezegd zal dit worden toegepast op de herkenning van karakters. De gevolgen op nummerbordherkenning zijn afhankelijk van de gevolgen op de karakterherkenning. Na elk hoofdstuk zal worden vermeld in hoeverre de centrale vraag is beantwoord. Daarnaast worden dan ook de gevolgen voor nummerbordherkenning beschreven.

Centrale vraag:

Wat is het gevolg van het gebruik van het originele beeld bij feature extraction op de herkenning van karakters?

2.3 Deelvragen

Om een bevredigend antwoord op de centrale vraag te geven zullen twee deelvragen worden beantwoord. In de eerste plaats zal een onderzoek gedaan worden naar de effecten van thresholding op het herkenningsproces van karakters. Hierbij zullen enkele aannames worden gemaakt en verantwoord. In de tweede plaats bestaat er interesse in de daadwerkelijke resultaten van toepassing van grayscale features op de herkenning van karakters.

Om het nut van het toepassen van grayscale features te onderzoeken is kennis omtrent de invloed van thresholding op de herkenning van karakters noodzakelijk. Deze invloed is echter niet direct inzichtelijk, omdat de herkenning na thresholden afhangt van de gekozen thresholdmethode en afhankelijk is van de robuustheid van de gebruikte feature extractor. Over de individuele herkenningspercentages van karakters is bovendien geen data beschikbaar.

Daarom wordt onderzoek gedaan naar de leesbaarheid van karakters voor en na thresholden. Van de grayscale karakters is bekend dat ze leesbaar zijn, omdat dit werd geëist in het bepalen van het werkdomein. Een definitie van leesbaarheid na thresholding wordt gegeven in het begin van hoofdstuk 3.

Leesbaarheid vereist een participerend object. Karakters worden gelezen door iets of iemand. Deze tekst wordt gelezen door u, de data in dit onderzoek door het herkenningssysteem. De leesbaarheid van karakters na thresholding wordt vastgesteld door een human expert. Wil dit in het deelonderzoek betreffende leesbaarheid nut hebben, dan moet er een verband bestaan tussen menselijke en machinale leesbaarheid van karakters. Dit verband bestaat.

Vooropgesteld kan worden dat mensen beter karakters kunnen herkennen dan machines. Dat blijkt alleen al uit het feit dat alle nummerborden uit de niet herkende dataset herkenbaar zijn voor elke alfabeet. Als een karakter menselijk onleesbaar is, kan van een machine derhalve niet beter worden verwacht. Van het grijze gebied hier tussenin kan worden gesteld dat een vermindering in menselijke leesbaarheid resulteert in een vermindering in machinale leesbaarheid. De aanwezigheid van na thresholding menselijk minder leesbare karakters zal dit verband alleen maar bekrachtigen, omdat ze afkomstig zijn van machinaal minder leesbare nummerborden.

Het verband tussen menselijke en machinale leesbaarheid impliceert dat karakters die na thresholding menselijk minder leesbaar, ook moeilijker classificeerbaar zijn. Door dus onderzoek te doen naar de representatie van karkaters na thresholding wordt inzicht verkregen in het nut van het toepassen van grayscale features. Dit wordt onderzocht door middel van deelvraag 1.

Deelvraag 1:

Wat is de invloed van thresholding op de representatie van karakters?

Deelvraag 2:

Wat is het resultaat van het gebruik van grayscale features op karakterherkenning?

Uit de invloed van thresholding op de representatie van karakters, blijkt het nut van toepassing van grayscale features. Deelvraag 2 omvat de daadwerkelijke resultaten omtrent grayscale karakterherkenning. De deelvragen 1 en 2 worden respectievelijk in hoofdstuk 3 en 4 beantwoord.

3 Thresholding

De karakters waarmee in het onderzoek wordt gewerkt zijn goed geïsoleerd en afkomstig van menselijk leesbare nummerborden. Daarnaast is het proces van karakterherkenning besproken, met daarin thresholding als één van de processtappen. Het effect van thresholding op de karakterrepresentatie zal nu worden onderzocht. Dat beantwoordt deelvraag 1. Maar eerst zal de term leesbaarheid worden gedefinieerd. Aan de resultaten zullen de nodige conclusies worden verbonden die vervolgens in de context van nummerbordherkenning zullen worden geplaatst.

3.1 Leesbaarheid

Het is lastig om de directe invloed van thresholding te meten, omdat het één stap is uit een langer proces. Daarbij is de wijze van classificatie van binaire karakters niet direct inzichtelijk. Centraal staat daarom de beoordeling op leesbaarheid na thresholding. De term leesbaarheid moet dan wel duidelijk gedefinieerd worden. Leesbaarheid na thresholding wordt gedefinieerd als een representatief binair beeld van de voorgrond, het karakter, van het originele beeld. Het wegvallen van karakterdelen of het verschijnen van binaire voorgrond op plaatsen waar geen karakter is, verminderd de leesbaarheid.

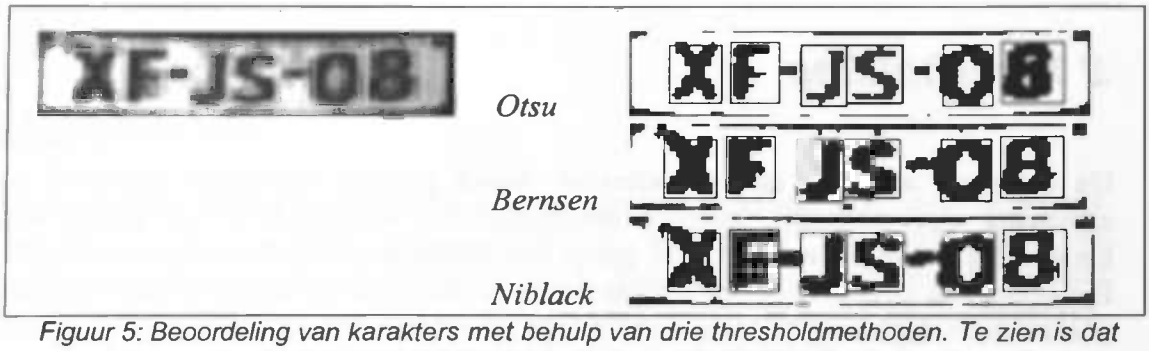
3.2 Methode

De beoordeling op leesbaarheid na thresholding is visueel van aard en vindt plaats door middel van een menselijke expert. Uit de beoogde beoordeling volgen de klassen die zullen worden onderscheiden. De beoordeling zal plaatsvinden per karakter. Hierbij zullen meerdere thresholdmethoden worden aangewend. De beste uitkomst van alle methoden bepaalt de klasse waarin een karakter wordt ingedeeld. Door verschillende thresholdmethoden per karakter te gebruiken, worden de probleemgevallen geïsoleerd.

3.2.1 Thresholdmethoden

Dat gebeurt door het gebruik van de beste thresholdmethoden voor handen. Trier, Jain en Taxt hebben verschillende methoden gerangschikt naar prestaties. [4, 5] Hierbij is onderscheid gemaakt tussen globale en locale thresholdmethoden.

Niblack's en Bernsen's thresholdmethoden zullen worden toegepast omdat ze tot de top van de beste locale methoden behoren. Deze methoden worden afgesloten met de postprocessing stap van Yanowitz. Locale methoden presteren over het algemeen beter dan globale thresholdmethoden. Toch wordt de methode van Otsu ook toegepast. Deze methode presteert het best van de globale thresholders. Het kan namelijk voorkomen dat een karakter na globale thresholding beter leesbaar is dan na locale thresholding. (Zie figuur 5)



Figuur 5: Beoordeling van karakters met behulp van drie thresholdmethoden. Te zien is dat de globale methode niet per defenitie slechtere prestaties levert dan de locale methoden.

De methode van Otsu selecteert een globale threshold aan de hand van een minimalisatie van de variatie in twee groepen van pixels, gescheiden door die threshold.[6]

De methode van Niblack gebruikt voor een pixel *(x,y)* in een $r \times r$ window de volgende threshold.

$$T(x,y) = m(x,y) + k \cdot s(x,y)$$

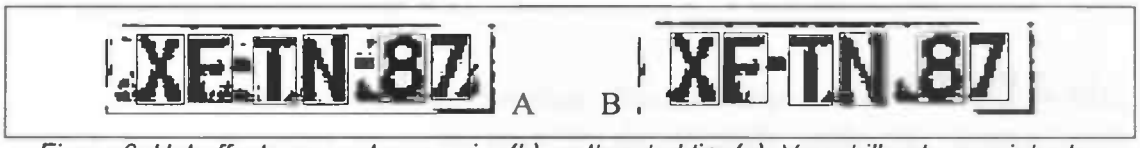
Hierbij is m de gemiddelde grijswaarde in het window, s de standaarddeviatie in het window en k het gewicht van s in de threshold. Als k klein wordt gekozen zal er minder achtergrond boven de threshold vallen. Na enig experimenteren heb ik $k = -1$ gekozen, omdat over meerdere nummerborden deze waarde de beste leesbaarheid tot gevolg heeft. Het window over het pixel moet groter dan een karakterstreep en kleiner dan het karakter zijn om locale variatie te handhaven. Ik heb gekozen voor een window van *10x10* pixels, zodat deze locale variatie kan worden benut.

De methode van Bernsen gebruikt voor elk pixel *(x,y)* in een $r \times r$ window de threshold

$$T(x,y) = \frac{Z_{low} + Z_{high}}{2}$$, met als contrast maat $C(x,y) = (Z_{high} - Z_{low}) < l$

De twee waarden van Z zijn de locale minimale en maximale grijswaarden. Mochten deze twee waarden nu dicht bij elkaar liggen dan wordt het pixel ingedeeld bij de achtergrond. Deze contrast maat wordt na experimenteren bepaalt op *15*, omdat over meerdere nummerborden deze waarde de beste leesbaarheid tot gevolg heeft. De windowgrootte wordt ook hier weer *10x10* pixels gekozen vanwege de locale variatie.

Na het thresholden wordt door middel van post-processing, ontwikkeld door Yanowitz en Bruckstein, het beeld verbeterd door kleine objecten te verwijderen. Eerst wordt het input gefilterd met een 3x3 mean-filter. Vervolgens worden alle objecten verwijderd waarvan het gemiddelde van de gradienten van de randen van het object beneden een zekere threshold *Tp* liggen. Bij de classificatie is hier een kleine threshold voor gekozen, omdat het gewenst is de essentiele informatie te behouden. Experimenteel is deze threshold vastgesteld op *Tp* = 10. Deze waarde is groot genoeg om kleine objecten, veelal geen karakteronderdelen, weg te filteren en grote objecten, veelal een karakter representerend, behouden blijven. (Zie figuur 6)



Figuur 6: Het effect van postprocessing(b) na thresholding(a). Verschillende smerigheden worden weggefilterd.

3.2.2 Klassen

Uit de beoogde beoordeling op succesvolle thresholding volgen de klassen die zullen worden onderscheiden. Met succesvol thresholden wordt bedoeld: het proces dat een leesbaar binair karakter oplevert. Er is één klasse voor de karakters die na thresholding leesbaar zijn zijn en één klasse voor de karakters die dat niet zijn. Het kan gebeuren dat het te thresholden karakter vergezeld wordt door andere objecten in het beeld die het succesvol thresholden hinderen. Bijvoorbeeld door schroeven of stickers. Deze karakters worden ingedeeld in een aparte klasse. Zie figuur 7 voor de definities van de klassen.

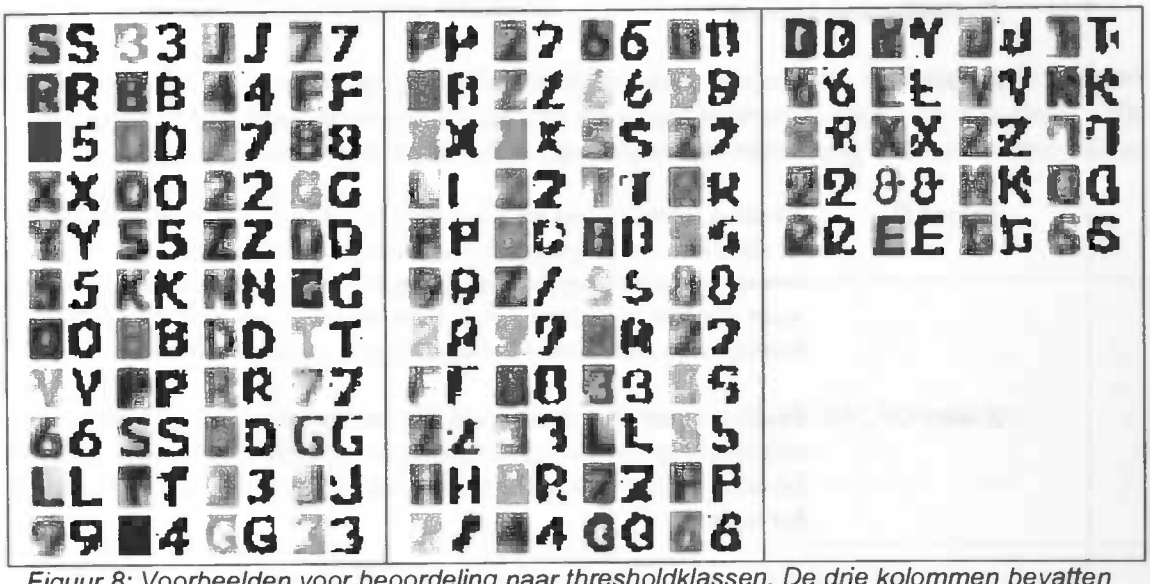
Definities van klassen

Klasse A	*Beelden waarvan, na thresholding, de zwarte pixels de data van het origineel representeren, die tot de voorgrond behoort. Een leesbaar origineel transfomeert hierbij naar een leesbaar binair beeld.*
Klasse B	*Beelden waarvan, na thresholding, de zwarte pixels niet exclusief de data van het origineel representeren, die tot de voorgrond behoort. In deze beelden kan voorgrond worden gevonden op plekken waar die niet is en vica versa. Een leesbaar orgineel transformeert hierbij naar een minder leesbaar binair beeld.*
Klasse C	*Beelden waarvan, na thresholding, de zwarte pixels de data van het origineel representeren, die tot de voorgrond behoort. Deze beelden bevatten tevens ongewilde voorgrond, die niet gerelateerd is met het karkater.*

Figuur 7: Drie klassen voor de beoordeling van een karakter op leesbaarheid na thresholding.

3.2.3 Beoordeling

De indeling naar klassen vindt plaats middels een human expert. Dat houdt in dat alle karakters visueel worden geïnspecteerd en beoordeeld op leesbaarheid na thresholding. Dit betreft een waardeoordeel. Het denkproces dat de classificatie bewerkstelligd zal onder woorden worden gebracht en geïllustreerd met voorbeelden. Dit proces zal tevens de indeling in de drie klassen verduidelijken.
Een geïsoleerd karakter met hinderde objecten is duidelijk vast te stellen. Objecten als stickers, trekhaken en schroeven vallen onmiddellijk op. Deze karakters worden ingedeeld in klasse C. Het kan echter voorkomen dat objecten in de voorgrond samenvallen met het karakter. Bijvoorbeeld een schroef op een karakter. Door het feit dat het niet of minder opvalt wordt in deze gevallen het thresholden niet verhinderd. De volgende denkstap wordt dan ondernomen. Het gethresholde karakter wordt beoordeeld op leesbaarheid. Deze representatie wordt door de onderzoeker beoordeeld. Het wegvallen van karakterdelen of het verschijnen van binaire voorgrond op plaatsen waar geen karakter is, is reden deze in te delen in klasse B. Deze beoordeling vind plaats voor alle drie de methoden. Als één ervan het origineel goed representeert, wordt deze in klasse A geplaatst. Zie figuur 8 voor vele voorbeelden bij de drie klassen.



Figuur 8: Voorbeelden voor beoordeling naar thresholdklassen. De drie kolommen bevatten van links naar rechts karakters uit respectievelijk de klassen A, B en C.

3.3 Resultaten

In totaal zijn 556 nummerborden bekeken. Daaruit volgt dat er 556 x 6 = 3336 karakters geclassificeerd zijn. Dat leidt tot de volgende verdeling. Gedetailleerde uitkomsten staan in tabel 1. Appendix A bevat tevens een procentuele verdeling van karakters over de klassen.

Verdeling:

Klasse A	*2660*	*(79,7 %)*
Klasse B	*479*	*(14,4 %)*
Klasse C	*197*	*(5,9 %)*

	kA	kB	kC	Tot.		kA	kB	kC	Tot.
A	1	0	0	1	T	128	21	17	166
B	96	60	3	159	U	0	0	0	0
C	4	0	0	4	V	176	5	9	190
D	130	7	5	142	W	0	0	0	0
E	0	0	0	0	X	123	13	3	139
F	118	21	17	156	Y	44	3	4	51
G	101	16	8	125	Z	116	4	5	141
H	100	14	2	116					
I	0	0	0	0					
J	86	7	13	106	0	79	5	5	89
K	30	6	13	49	1	86	17	19	122
L	84	9	17	110	2	137	29	8	174
M	0	0	0	0	3	95	10	2	107
N	125	9	0	134	4	92	10	1	103
O	1	0	0	1	5	78	21	5	104
P	115	28	9	152	6	71	22	3	96
Q	0	0	0	0	7	76	9	14	99
R	97	24	6	127	8	83	48	2	133
S	121	16	4	141	9	67	29	3	99

Tabel 1: Uitkomsten klassificatie van leesbaarheid na thresholding per karakter.

Over de uitkomsten kunnen de volgende constateringen worden vermeld. Van alle karakters kan 14.4% als probleemgeval voor thresholding worden bestempeld en blijkt 5.9% moeilijk hanteerbaar. Om onduidelijke redenen blijken cijfers moeilijker te thresholden dan letters. Het thresholden van de 8 en de B verloopt moeizamer dan bij andere karakters.

3.4 Conclusies deelvraag 1

In hoofdstuk 2 werd geconstateerd dat minder leesbare karakters minder makkelijk classificeerbaar zijn. Vanwege het feit dat de invloed van thresholding niet direct meetbaar was, werd derhalve besloten de representatie van karakters na thresholding te onderzoeken. In het onderzoek werden de probleemgevallen geïsoleerd door gebruik te maken van drie verschillende thresholdmethoden.

Deelvraag 1:

Wat is de invloed van thresholding op de representatie van karakters?

Het antwoord op deelvraag 1 kan dan gegeven worden. De invloed van thresholding op de representatie van karakters is dat 14.4% van alle karakters na thresholding verminderd leesbaar bleek. Vanwege het verband tussen menselijk leesbaarheid en machinale leesbaarheid, in combinatie met het gegeven dat alleen de probleemgevallen zijn gevonden, kan dan worden gesteld dat thresholding van invloed is op de herkenning van karakters.

Deze uitkomst legitimeert een onderzoek naar mogelijke verbeteringen. Deze verbeteringen kunnen gevonden worden, door gebruik te maken van het menselijk leesbare originele grayscale beeld. Het onderzoek dat deelvraag 2 beantwoord zal in hoofdstuk 4 behandeld worden.

Deelvraag 2:

Wat is het resultaat van het gebruik van grayscale features op karakterherkenning?

3.5 Context: nummerbordherkenning

Het nummerbordherkenningsproces, dat is beschreven in hoofdstuk 1, herkent niet alle aangeboden nummerborden. Van de 960 niet herkende nummerborden die beschikbaar zijn gesteld voor onderzoek bleken 556 te voldoen aan de eisen van leesbaarheid en tevredenstellende isolatie van de individuele karakters.

Van alle 6 x 556 = 3336 karakters bleken 14.4% verminderd leesbaar na thresholding. Dit betrof alleen de probleemgevallen. In de context van nummerbordherkenning zal dit percentage hoger zijn, omdat er per nummerbord 1 thresholdmethode wordt toegepast, en niet 3 thresholdmethoden.

Er is geconcludeerd dat er mogelijk verbeteringen zijn aan te brengen in het herkennen van karakters door het thresholden van beelden niet meer te gebruiken in het herkenningsproces. Verbeteringen omtrent nummerbordherkenning zijn minder gemakkelijk inzichtelijk, omdat de prestaties van de karakterherkenner onbekend zijn en omdat er 6 karakters op een nummerbord staan. Om dat in te zien wordt er eerst gekeken naar de verdeling van herkende en niet herkende karakters uit de 3 klassen. Vervolgens wordt gekeken naar de samenstelling van karakters uit de 3 klassen over de nummerborden.

	Herkend	*Niet herkend*
Klasse A	*p*	*q*
Klasse B	*x*	*y*
Klasse C	*k*	*l*

Tabel 2: De drie verschillende klassen zullen verschillende verhoudingen herkend-niet herkende percentages hebben bij de karakterherkenning in CLPR.

Een vaststaand feit is dat geen enkel nummerborden in dit onderzoek correct is herkend. Dat betekent dat minstens één karakter van het nummerbord onherkenbaar bleek. Het bestaande systeem zal karakters uit de verschillende klassen wel of niet herkennen. Herkenning van een karakter treedt op als classificatie na gebruik van één van de drie thesholdmethoden succesvol blijkt. In tabel 2 is dat simpel weergegeven. De exacte aantallen zijn onbekend, omdat individuele karakter herkenningspercentages niet beschikbaar zijn. Toch kunnen op deze plek reeds verwachtingen hierover worden uitgesproken.

79.9% procent van alle karkaters bleek na thresholding menselijk leesbaar. Vanwege deze leesbaarheid bestaat de verwachting dat p veel groter zal zijn dan q. Oftewel vrijwel alle karakters uit klasse A zullen correct worden herkend bij classificatie na gebruik van één van de drie thresholdmethoden. Bij klasse C is dat precies omgekeerd. Er is geen reden om aan te nemen dat het systeem troebele karakters correct herkent. Wellicht dat het zo en nu en dan toch gebeurt. Daarom wordt verwacht dat k veel kleiner zal zijn dan l. *42.8%* van alle nummerborden bevatten minstens een karakter uit klasse B. Over de verhouding x/y kan alleen gezegd worden dat y niet laag zal zijn in vergelijking met x, omdat de aanwezigheid van nummerborden met karakters uit klasse B wel kleiner zou zijn geweest.

In tabel 3 staat vermeld in welke samenstelling van klassen de nummerborden zijn onderverdeeld. De aanwezigheid van karakters uit klasse C op een nummerbord is gereduceerd tot één kolom. Gedetailleerde gegevens staan vermeld in appendix A. Hieruit valt op te maken dat het verband tussen menselijke en machinale leesbaarheid wordt bevestigd doordat 71.0 % van de nummerborden verminderd leesbare karakters bevat. Daarnaast bevat 42.8% van alle nummerborden minimaal een karakter uit klasse B.

	Sequences							
	6A 0B	5A 1B	4A 2B	3A 3B	2A 4B	1A 5B	0A 6B	met C
Aantal	161	136	56	26	16	2	2	157
Verhouding	29.0%	24.5%	10.1%	4.7%	2.9%	0.4%	0.4%	28.2%

Tabel 3: Uitkomsten classificatie gepresenteerd in aantal voorkomens van klasse B op een een nummberbord.

Alhoewel precieze uitspraken over verbeteringen omtrent karakterherkenning niet gedaan kunnen worden zijn verbeteringen aan nummerbordherkenning meetbaar vanwege het feit dat het domein alleen bestaat uit nummerborden die niet zijn herkend.

4 Grayscale karakter herkenning

Deelvraag 1 is beantwoord. Hieruit kon worden geconcludeerd dat thresholden bijdraagt aan het mislukken van zowel karakter- als nummbordherkenning. De vraag die nu rijst is in hoeverre daar verbeteringen in kunnen worden aangebracht door de binairisatiestap over te slaan. Deelvraag 2 vindt zijn beantwoording in dit hoofdstuk. Dat zal plaatsvinden in drie fasen. Eerst zullen grayscale feature extractors moeten worden gevonden. Daarvoor zal de literatuur worden geraadpleegd. Mogelijke kandidaten zullen worden getoetst aan opgestelde criteria en daaruit zal een selectie vloeien. Vervolgens wordt in fase 2 deze selectie toegespitst op het domein ter hand. Dat gebeurt door specifieke implementatiekeuzes te maken. Uiteindelijk resulteert dat in een set van karakter feature extractors. In de laatste fase worden deze feature extractors getoetst en vergeleken door hun prestaties met behulp van neurale classificatie te controleren.

4.1 Feature extractors

Om een goede vergelijking tussen de verschillende mogelijkheden op het gebied van grayscale feature extraction te bewerkstelligen, is het noodzakelijk om voor dit onderzoek relevante criteria te formuleren. Verschillende aspecten spelen een rol. Voor een geschikte toepassing op het domein zal rekening moeten worden gehouden met de eigenschappen van de karakters.

Een karakter is klein van omvang, gemiddeld *12x15* pixels, en heeft een uiteenlopend grijswaardenbereik. Bij toepassing van features van grayscale beelden zal derhalve de illuminatie van het beeld in overweging moeten worden genomen. Invariatie omtrent belichting is dan noodzakelijk. Daarnaast zal bekeken worden of de feature kandidaten om kunnen gaan met de lage resolutie. Dat zal worden getoetst door onder andere vast te stellen of de methode wordt toegepast op een vergelijkbaar domein. Naast vergelijkbare toepassing is belangrijk dat toepassing van een methode leidt tot positieve resultaten in vergelijking met features van binaire beelden. In combinatie met een vergelijkbaar domein lijkt winst in herkenning dan haalbaar. Het is een simpel feit dat goed voorbeeld goed doet volgen. De mate waarin een toepassing daarom navolging krijgt in de literatuur is een maatstaf voor het succes ervan. Gevonden kandidaten worden getoetst op de geselecteerde criteria, die in figuur 9 samengevat zijn weergegeven.

Illuminatie invariatie
Vergelijkbare toepassing
Navolging
Resultaten

Figuur 9: Vier criteria ter toetsing van feature extraction methoden

In principe zijn er meerdere criteria denkbaar, bijvoorbeeld rekenintensiteit, aantoonbare robuustheid, ruisgevoeligheid en meer. Toch zijn bovenstaande vier criteria voldoende om een goede toetsing mee uit te voeren, omdat zowel rekening wordt gehouden met het werkdomein als met de prestaties van mogelijk kandidaten.

De literatuur voorziet in een hele reeks mogelijkheden om features van grijswaardenbeelden te bepalen. Ze zullen niet allemaal behandeld worden. Alleen de meest veelbelovende zijn met elkaar vergeleken.

4.1.1 Verkenning

4.1.1.1 Gabor filters

Gabor Jets bestaan uit een set van outputs van gabor filters. Een gabor jet is een representatie van de locale grijswaardendistributie. Het ontstaat na een convolutie met een gabor golf. Een gabor golf is een lopende golf, begrensd door de functie van Gauss. Gabor jets zijn complex van aard. Het complexe deel is relatief ongevoelig voor onregelmatigheden, en kan worden gebruikt bij classificaties.

Figuur 10: Een diagonale Gabor filter gebruikt bij convolutie.

Aan gabor jets kunnen karakteristieken worden ontleend. Na eventuele normalisatie van het beeld worden vier jets geconstrueerd. Eén in verticale, één in horizontale en twee in de diagonale richtingen. In dit proces worden de horizontale en verticale Gauss' cut-off frequenties voor de golf, en de amplitude van de golf experimenteel bepaald. Figuur 10 laat een diagonale kernel zien. De vier jets worden opgedeeld in elkaar overlappende gedeelten, bijvoorbeeld in 4x4 subblocks. Van elk van deze subblocks worden weer vier jets gecreëerd. Dit leidt tot 256 features, die zich lenen voor neurale classificatie. [7]

Deze methode wordt door Yoshimura en anderen toegepast in zeer onregelmatige omgeving en behalen daarbij een herkenning van ongeveer 90 procent. Ter vergelijking wordt met dezelfde data neurale classificatie verricht na thresholding. Hierbij wordt de thresholdmethode van Otsu gebruikt. Het blijkt dat de gabor jets hebben geleid tot 3 procent winst.

Gabor jets zijn minder gevoelig voor lage resoluties en onregelmatigheden in de input. Dat is vergelijkbaar met de situatie in dit onderzoek. Het is echter bekend dat de methode van Otsu niet de best presterende thresholdmethode is. Het toepassen van jets levert maar een minimale winst op in vergelijking met classificatie na thresholding.

4.1.1.2 Momenten

Numerieke karakteristieken van een gedistribueerde dataset worden momenten genoemd. Ze zijn een representatie van de structuur van een plaatje en kunnen worden toegepast op zowel binaire als grijswaarden beelden. De keuze van momenten die als features kunnen dienen is bewerkelijk. Daarbij moet gelet worden op invariatie, of juist variatie, in rotatie, translatie, locatie.

Bij karakterherkenning wordt gebruik gemaakt van invariante momenten van binaire beelden. Dit levert een herkenning van ongeveer 90 procent. Er wordt opgemerkt dat de fouten die het systeem maakt veelal misclassificatie zijn. Dit feit maakt momenten niet geschikt als alleenstaande feature extractor, maar maakt het wel geschikt om in combinatie met andere feature extractors te gebruiken. De momenten die hierbij worden gebruikt zijn geïnitieerd door Hu en verder uitgewerkt.

Zernike-momenten kunnen worden toegepast op grijswaardenplaatjes en blijken daar ook geschikt voor te zijn.[3] Voor grijswaardenbeelden blijken complexe momenten beter te functioneren dan de invariante momenten van Hu.[8] Van pseudo-Zernike momenten is bekend dat ze het best presteren van de complexe momenten. Bovendien is aangetoond dat ze het minst gevoelig zijn voor ruis in de data. Pseudo Zernike momenten zijn daarom het meest geschikt voor eventuele toepassing. [9]

Ook bij pseudo-Zernike momenten zullen invarianties moeten worden ingebouwd. Omdat de oriëntatie van de karakters in dit onderzoek vaststaat is rotatie-invariantie niet noodzakelijk. Translatie-invariantie kan worden ingebouwd door de beelden te schalen naar een vaste grootte. Beter is het om translatie-invariantie in te bouwen in de constructie van de pseudo-Zernike momenten. [10] Belangrijker is de illuminatie invariantie. Om momenten van grijswaardenbeelden succesvol toe te passen, zal daarvoor een oplossing moeten worden aangedragen. Momenten zijn inherent hun opmaak niet illuminatie invariant. Om dit wel te bereieken zijn verschillende mogelijkheden beschikbaar. Invariatie in illuminatie kan worden bereikt door de afgeleiden van beelden te gebruiken of door de grijswaarden naar een vastgelegd bereik te schalen.

In de literatuur krijgen peudo Zernike momenten redelijk wat navolging. Wat daarbij opvalt is dat vrijwel elke toepassing van deze momenten gebaseerd is op toepassing op binaire beelden. Van geskeletoneerde, alleen cijfers representerende, handgeschreven karakters levert dit een herkenningspercentage van 97,3% met behulp van een multilayer perceptron. [10]

4.1.1.3 Triangulatie

Een grayscale beeld kan worden opgevat als een oppervlak. Hierbij zijn de pixelwaarden de hoogte boven het vlak van de grid. Een oppervlak kan worden benaderd met driehoeken. Dit wordt vooral gedaan om datacompressie te bereiken voor hoogtevelden van grote omvang. Er blijken echter ook features aan te ontlenen.

Een oppervlak kan op verschillende manieren worden getrianguleerd. Om het oppervlak naar tevredenheid te benaderen kan de data-dependent Delauney triangulatie worden aangewend. [11, 12] Hierbij worden in een initiële triangulatie telkens die punten toegevoegd, die de grootste afstand tot het vlak van de omvattende driehoek hebben. Een triangulatie is niet illuminatie invariant. De gemiddelde deviatie van het grijswaardengemiddelde van het beeld bepaalt bijvoorbeeld de stijgingscoëfficiënten van lijnsegmenten van een driehoek. De gemiddelde stijgingscoëfficiënt in een vastgestelde richting per zone van een karakter, kunnen worden gebruikt als features.

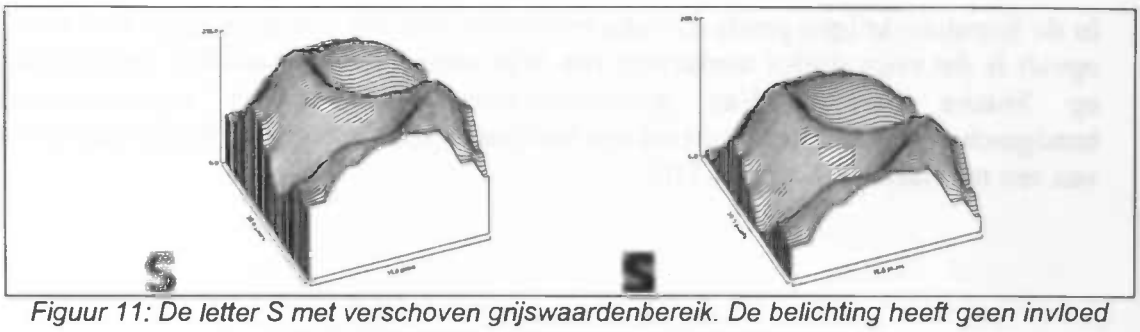
Van deze triangulatie kan door het opsporen van die ribben van driehoeken die de grootste stijgingscoëfficiënten hebben, een boundary worden gevonden van het karakter.[13] Deze boundary kan worden aangewend voor classificatie, maar is niets anders dan een alternatief edge detection systeem.

Daarnaast is de data niet vergelijkbaar met het domein, want van hoge resolutie. Hoge resolutie data levert een stabielere traingulatie op. Resultaten zijn niet beschikbaar.

4.1.1.4 Topografische Features

Oppervlakken hebben eigenschappen als heuvels, pieken, dalen, zadels, randen en vlakke oppervlakken. Door pixels te labellen met de eigenschap die ze op het oppervlak bezitten, ontstaat een basis waaruit karakteristieken kunnen worden ontleend. [14, 15]

De eigenschap van een pixel op het oppervlak is onafhankelijk van de grijswaarden in het beeld. Oftewel, de globale belichting van een karakter kan worden veranderd zonder dat dit effect heeft op de vorm van het oppervlak. In het voorbeeld in figuur 11 zou dat simpelweg een verticale verschuiving betekenen.



Figuur 11: De letter S met verschoven grijswaardenbereik. De belichting heeft geen invloed op de topografische eigenschappen van het oppervlak.

Topografische eigenschappen van een oppervlak leiden tot segmentatie van voor- en achtergrondinformatie. Het proces heeft overeenkomsten met local adaptive thresholding. In een bepaald window worden pixel vergelijkende operaties uitgevoerd, die tot juiste pixelclassificatie moeten leiden. Local adaptive thresholden mislukt vaak bij gebrek aan locale variatie. Topografische labelling vindt ook plaats binnen een window. De winst die topografische labelling kan brengen komt van een kleiner window rond een pixel. Het feit dat de gebruikte karakters menselijk leesbaar zijn, betekent dat er locale variatie aanwezig moet zijn. Topografische labelling benut dat optimaal. Uit de voorbeelden uit de literatuur [14, 15] blijkt bovendien dat het vergelijkbaar is met de data die in deze opdracht wordt gebruikt.

Bijkomend voordeel is dat de achtergrond ook wordt gesegmenteerd. De vorm van de achtergrond vertelt iets over de vorm van de voorgrond. Deze informatie kan worden toegepast bij herkenning.

Het enige gevonden tegengeluid betreffende topografische labelling komt van Trier, Taxt en Jain.[16] Topografische labelling op hun domein van weerkaarten geeft geen grote verbetering Bovendien verwerpen ze de methode om praktische redenen (rekentijd). Deze twee argumenten kunnen terzijde worden gelegd, omdat ze geen rol spelen binnen dit probleem. Enerzijds omdat het de gebruikte domeinen verschillen en anderzijds omdat rekentijd geen rol speelt in dit onderzoek.

4.1.2 Vergelijking en selectie

Verschillende methoden zijn besproken en getoetst aan meerdere criteria. In tabel 4 wordt een samenvattende vergelijking gepresenteerd.

Topografische features blijken bij vergelijkbare data positieve resultaten te behalen. Bovendien zijn ze invariant onder variabele belichting. Deze methode zal worden geïmplementeerd. Pseudo Zernike momenten zullen ook worden geïmplementeerd. In illuminatie invariantie zal dan moeten worden voorzien. Daarvoor zijn verschillende technieken beschikbaar. De andere twee methoden zijn gezien hun prestaties minder geschikt voor implementatie.

	Illuminantie invariantie	Navolging	Vergelijkbare toepassing	Resulaten	Totaal
pZ Momenten	-	+	-	+	+
Gabor Filters	-	-	+	-	-
Triangulaties	-	-	-	onbekend	-
Topografica	+	+	+	+	+

Tabel 4: Samenvatting vergelijking van mogelijkheden. Pseudo Zernike momenten kent naast goede resultaten op binaire beelden ook navolging in de literatuur. Topografische pixellabelling kent op elk criterium een positieve uitkomst. Deze twee methoden worden geselecteerd.

4.1.3 Tussenbeschouwing

De beantwoording van deelvraag 2 heeft nu de eerste fase gepasseerd. In deze fase was het de bedoeling om methoden te vinden die in aanmerking komen om features van grayscale beelden te vinden. Deze selectie heeft geleid tot de keuze voor topografische pixellabelling en pseudo Zernike momenten.

In fase 2, de volgende paragraaf, zullen deze methoden worden toegespitst op het domein. Hierin zullen afwegingen worden gemaakt omtrent keuzes die moeten leiden tot een robuuste set features van grayscale beelden.

Fase 3 omvat de daadwerkelijke toepassing van de hierboven gevonden methoden. Dat wordt besproken in paragraaf 4.3. Pas hierna zal de effectiviteit van de gekozen kandidaten blijken.

4.2 Implementatie

4.2.1 Inleiding

In dit hoofdstuk zullen de geselecteerde grayscale feature extractors worden toegespitst op het domein van onderzoek. Daartoe worden eerst in detail de implementatie van de topografische features en pseudo Zernike momenten doorgenomen. Verschillende afwegingen en keuzes worden daarbij toegelicht. Dat leidt uiteindelijk tot een set van features, waarvan ook onderzocht zal worden in hoeverre deze redundante informatie bevat. Na de implementatie volgt het testen.

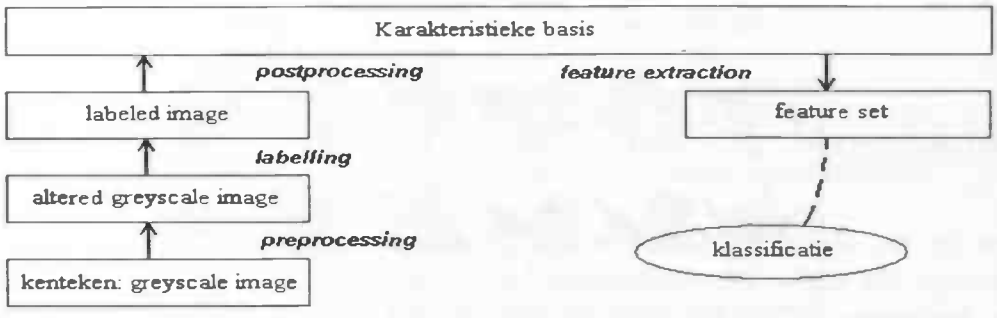
4.2.2 Topografische features

Het traject dat leidt van theorie tot toepassing, passeert enkele stations. Dat kan worden opgesplitst in twee gedeelten. (zie figuur 12) Eerst wordt richting een optimale karakteristieke topografische basis gewerkt. Uit deze karakteristieke basis worden vervolgens features geselecteerd die worden aangewend bij neurale classificatie. De classificatie zelf wordt besproken in de volgende paragraaf.

In een optimale karakteristieke basis wordt het originele kenteken zo goed mogelijk gerepresenteerd. Het doel van de implementatie van karakteristieke basis is dan ook een zo goed mogelijk *menselijk leesbare* basis te verkrijgen. Het verband tussen menselijke en machinale leesbaarheid is reeds aangetoond. Gedurende de implementatiefase zullen steeds waardeoordelen gegeven worden omtrent de *leesbaarheid* van de verschillende ter beschikking staande opties. Deze keuzes zullen zoveel mogelijk worden ondersteund met figuren.

Daar wordt naar toe gewerkt door eerst het originele beeld te bewerken voordat het wordt gelabeld. Na labelling kan een gelabeld image verbeterd worden door kleine onregelmatigheden uit het beeld te filteren. Uit de basis die hierdoor ontstaat zullen features worden geselecteerd voor neurale classificatie.

Al deze facetten worden hieronder doorgenomen, te beginnen bij de labelling. Er volgt nu echter eerst een bespreking van een eigenschap van topografische labelling en de gevolgen daarvan op het proces.



Figuur 12: Het toepassingstraject van topografische features. Na eventuele voorbewerking wordt het greyscale beeld gelabeld. Daarna kan het worden nabewerkt. Van de basis die dit oplevert kunnen features worden gevonden die ter klassificatie zullen worden aangeboden.

4.2.2.1 Voorgrond en Achtergrond

Een eigenschap van topografische labelling is dat naast segmentatie van de voorgrond, het karakter zelf, ook de achtergrond wordt gesegmenteerd. Omdat de vorm van de achtergrond in relatie staat tot de vorm van de voorgrond, kan de achtergrond wellicht van toepassing komen bij classificatie. Alhoewel niet volledig analoog, bestaat er een verband tussen het morfologisch skelet van de binaire voorgrond van een karakter en de gesegmenteerde achtergrond.

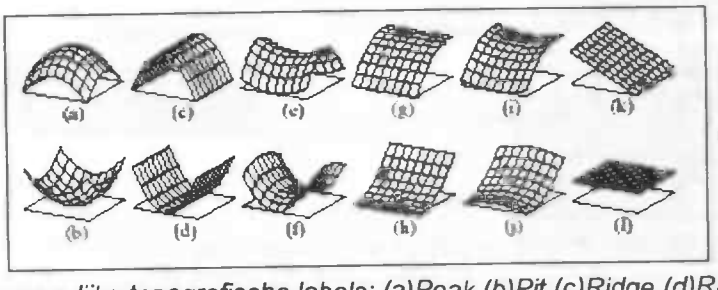
De voor- en achtergrond van een nummerbord zullen verschillend worden gesegmenteerd, omdat er verschillende belangen zijn. Omtrent de voorgrond bestaat er interesse in maximale representatie van pixels die tot de voorgrond behoren. Bij de achtergrond van een karakter staat de vorm centraal. Verderop zal dat worden toegelicht met voorbeelden. Vanwege deze tweedeling wordt het traject tot de karakteristieke basis verschillend doorlopen: er zullen verschillende keuzes worden gemaakt voor de voor- en achtergrond.

Alle bewerkingen en de effecten daarvan worden hieronder doorgenomen. Als dat bekend is kunnen voor de twee trajecten de daadwerkelijke keuzes worden gemaakt, die zullen leiden tot een optimale karakteristieke basis.

4.2.2.2 Labels

Elk pixel wordt voorzien van een label die de eigenschap van dat pixel op het oppervlak representeert. Er moeten keuzes worden gemaakt omtrent welke labels zullen worden gebruikt en met welk algoritme deze worden toegewezen.

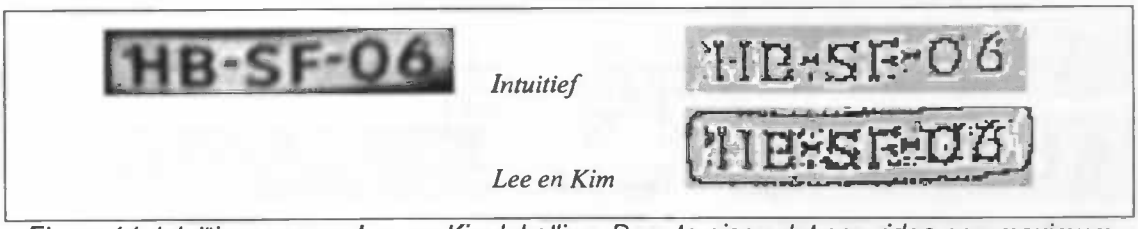
Er zijn verschillende pixel-label schema's denkbaar. (zie figuur 13) In ieder geval worden de zeven basis labels gebruikt. Dat zijn *Peak, Pit, Saddle, Ridge, Ravine, Flat* en *Hill*. [14, 15] Met deze zeven labels ontstaat een basis voor het bepalen van karakteristieken. De *peak*s en *ridge*s representeren de voorgrond, de *pit*s en *ravine*s de achtergrond. Het mogelijk onderscheid te maken tussen verschillende *saddle*'s en *hill*'s.[16]



Figuur 13: Alle mogelijke topografische labels: (a)Peak (b)Pit (c)Ridge (d)Ravine (e)Ridge saddle (f)Ravine saddle (g)Convex hill (h)Concave hill (i)Convex saddle hill (j)Concave Saddle hill (k)Slope hill (l)Flat

Dit zou nodig zijn voor een correcte segmentatie tussen voor en achtergrond. Een *convex hill* behoort bijvoorbeeld nog bij de voorgrond en een *concave hill* niet. Dit onderscheid zal hier niet worden gebruikt om twee redenen. Een overgang van voor naar achtergrond gaat gepaard met tussenliggende hills. Het is niet nodig daar een onderscheid in aan te brengen als er alleen interesse bestaat voor de structuur van het oppervlak. Bovendien is de gebruikte data klein van resolutie en kan de overgang van voor naar achtergrond dus radicaal zijn.

In nummerborden worden de karakters omgeven door achtergrond. Deze eigenschap leidt tot een andere teowijzing van labels dan gebruikt door Lee en Kim. Door te eisen dat een *ridge* een locaal maximum bevat, worden randen van een plateau niet als *ridge* bestempeld. In figuur 14 is te zien hoe zich dat manifesteert.



Figuur 14: Intuïtieve versus Lee en Kim labelling. Door te eisen dat een ridge een maximum heeft onstaat een leesbaarderkenteken.

De labels die gebruikt worden vormen de minimale set nodig voor labelling. Elk van hen is intuïtief gedefinieerd. (Figuur 15) Het is denkbaar een threshold in te bouwen bij pixelvergelijkende operaties. Dit zal verderop worden besproken. Bij het bepalen van een locaal maximum of minimum in een richting wordt een window rond het pixel gelegd. In tegenstelling tot local thresholding wordt dit window zo klein mogelijk gekozen *(3x3)* om de locale variatie optimaal te benutten.

Hill	Standaard
Flat	Omliggende pixels liggen even hoog
Ravine	Een minimum in één richting
Ridge	Een maximum in één richting
Pit	Een minimum in twee richtingen
Peak	Een maximum in twee richtingen
Saddle	Een minimum in één richting en een maximum in de andere richting

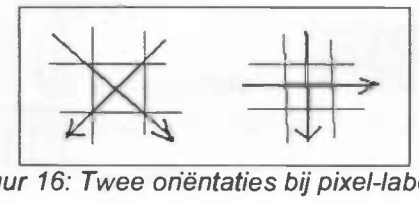
Figuur 15: De gebruikte labels bij topografische pixel labelling. Een pixel wordt als Hill bestempeld als het niet voldoet aan één van de andere voorwaarden.

De labels *peak* en *ridge* representeren voorgrondpixels, de labels *pit* en *ravine* representeren achtergrondpixels. Een *saddle* vertegenwoordig veelal een smalle dip tussen twee voorgrond objecten. Vandaar dat dit label ook tot de voorgrond wordt gerekend. De visualisaties in dit hoofdstuk zijn zodoende gepresenteerd in drie niveaus: voor-, achtergrond en neutrale pixels.

4.2.2.3 Label algoritme

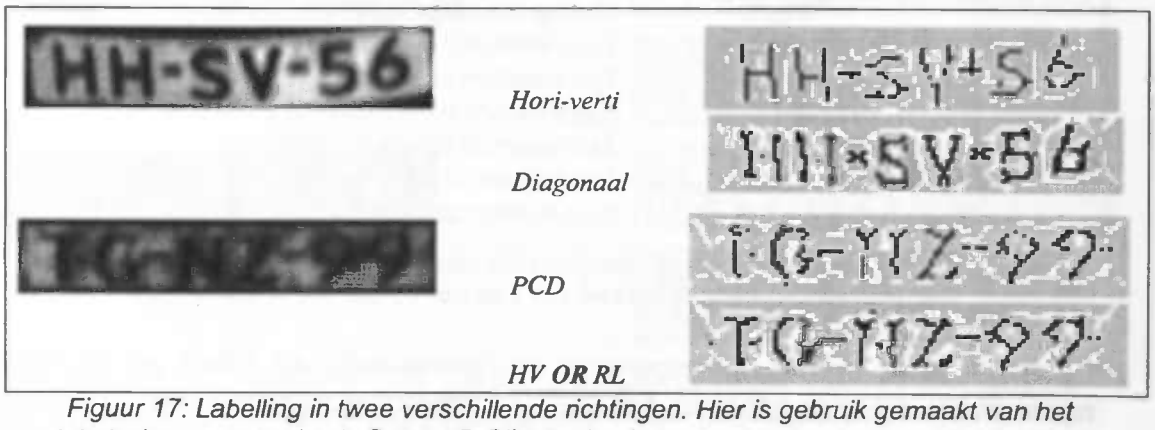
Voor de totstandkoming van de pixellabelling zijn verschillende algoritmen voorhanden. Wang en Pavlidis gebruiken discrete Chebychev polynomen om het oppervlak locaal te benaderen. Met behulp van eerste en tweede afgeleiden worden pixels vervolgens topografisch geclassificeerd.[14] Lee en Kim laten zien dat dit kan leiden tot misclassificaties. Zij introduceren een andere methode om de beoogde classificatie te bewerkstelligen.[15] Hierin speelt de *principal curvature direction* een belangrijke rol.

De principal curvature direction is de richting waarin het oppervlak de grootste buiging heeft. Deze richting wordt per pixel vastgesteld en kan worden gebruikt om de vier buren van het pixel vast te stellen aan de hand waarvan een label wordt toegewezen. Bij continue oppervlakken zijn er vele oriëntaties, maar bij discrete images wordt dit beperkt tot twee mogelijke oriëntaties. (figuur 16)



Figuur 16: Twee oriëntaties bij pixel-labelling

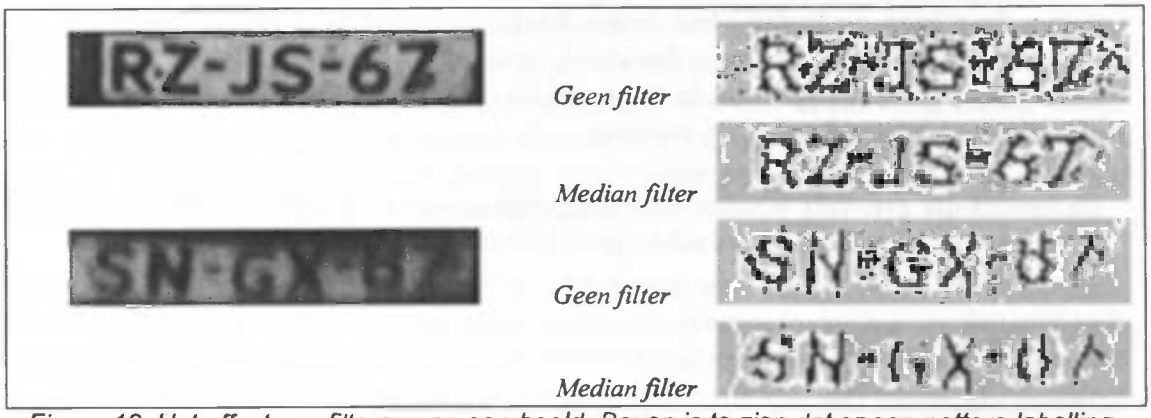
Op deze manier ontstaan drie manieren om een image te labellen. Namelijk twee met alleen één van de gekozen orientaties, horizontaal-verticaal en links-rechts-diagonaal, en, één waarbij de oriëntatie per pixel afhankelijk is van de principal curvature direction. Bij gebruik van de principal curvature direction wordt soms topografische informatie niet als zodanig gevonden terwijl deze wel aanwezig was. De reden hiervoor is dat als één van de richtingen van de principal curvature direction nul is, de oriëntatie *45* graden wordt gedraaid. Dat kan ervoor zorgen dat bij bijvoorbeeld een diagonale *ridge*, de andere oriëntatie wordt gekozen. Dat is goed te zien in figuur 17 bij de letter *'N'*. Bij weglating van deze oriëntatiedraaiing in het algoritme wordt vrijwel altijd de diagonale oriëntatie gekozen. Dit leidt tot verlies van sommige horizontale of verticale lijnsegmenten.



Figuur 17: Labelling in twee verschillende richtingen. Hier is gebruik gemaakt van het labelschema gevonden in figuur 15. Bij labelling in horizontale-verticale richting worden diagonale lijnsegmenten moeizaam gevonden en vica versa. In plaats van per pixel een keuze te maken voor de richting van labelling, levert een combinatie van twee labels per pixel over het algemeen het meest leesbare resultaat op.

4.2.2.4 Preprocessing en pixel-threshold

Voordat het label-algortime wordt aangewend kan een image worden gefilterd met een *3x3* median filter. Dit heeft als voordeel dat het oppervlak gladder wordt en dus mooiere labelling kan veroorzaken. (zie figuur 18) Nadeel is dat kleine oneffenheden worden gladgestreken. Een grotere neighbourhood zou dat effect vergroten. Deze kleine oneffenheden kunnen essentiële overgangen zijn tussen voor en achtergrond. Gezien de ruisige input moet hier een afweging worden gemaakt tussen het behouden van informatie en het veroorzaken van een mooiere labelling.

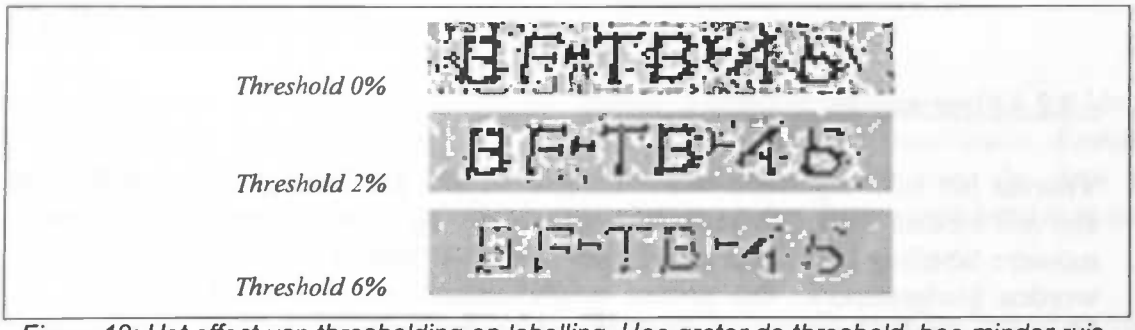


Figuur 18: Het effect van filteren van een beeld. Boven is te zien dat er een nettere labelling wordt veroorzaakt, onder valt voorgrondinformatie weg.Over het algemeen leidt het filteren van het beeld tot een leesbaarderkenteken.

Het toewijzen van een label aan een pixel gaat gepaard met vergelijkende operatoren tussen twee verschillende pixels. In deze operaties kan een threshold worden ingebouwd om kleine overgangen niet te misclassificeren. Het is hierbij zaak de threshold zo te kiezen dat grote overgangen behouden blijven voor classificatie. Een grote threshold leidt tot verlies van informatie, een kleine threshold tot ruis in de data. Vanwege de variable illuminatie wordt de threshold dynamisch per nummerbord vastgesteld. De threshold bij pixel-vergelijkende operaties wordt vastgesteld aan de hand van een percentage van het verschil tussen maximale en minimale grijswaarde van alle geïsoleerde karakters. Dit leidt afhankelijk van het gekozen percentage soms tot een threshold van nul. Dat is niet gewenst en daarom wordt de threshold minimaal op één gezet. In formule:

$$T_{nummerbord} = Max(1 \quad , \quad (t_p \cdot (g_{\max} - g_{\min})))$$

Met een thresholdpercentage van *2%* wordt bedoeld $t_p = 0{,}02$. In figuur 19 is het effect van thresholding op labelling te zien.



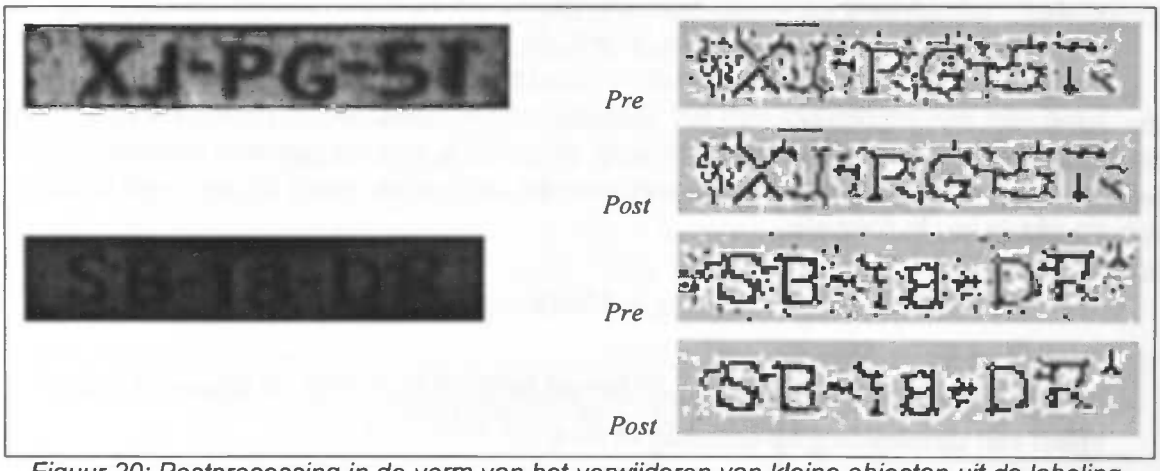
Figuur 19: Het effect van thresholding op labelling. Hoe groter de threshold, hoe minder ruis, maar meer informatieverlies.

De twee procedures werken allebei ruisonderdrukkend. Er bestaat een duale relatie tussen deze twee methoden. Aan de ene kant vermindert het gebruik van een filter de toepassing van een ingebouwde threshold, omdat het oppervlak gladder wordt. Aan de andere kant kan een ingebouwde threshold na filtering ook noodzakelijk zijn omdat er nog steeds misclassificatie kan ontstaan.

De keuzes die gemaakt worden voor het bereiken van de karakteristieke basis, worden verderop voor zowel voor, als achtergrond verschillend gemaakt.

4.2.2.5 Post processing

In een gelabeld image kunnen kleine objecten worden verwijderd. Door middel van de tijdsefficiënte klassieke methode van union-find [6] worden de connected components van de voorgrond gevonden. Deze connected components zijn opgebouwd uit pixels met voorgrondlabels. Alle componenten met een beperkt aantal pixels worden uit het beeld gefilterd door ze te vervangen door neutrale labels. De mininum toegestane objectgrootte moet niet te groot worden gekozen, om essentiële karakterdelen te behouden. (Figuur 20)



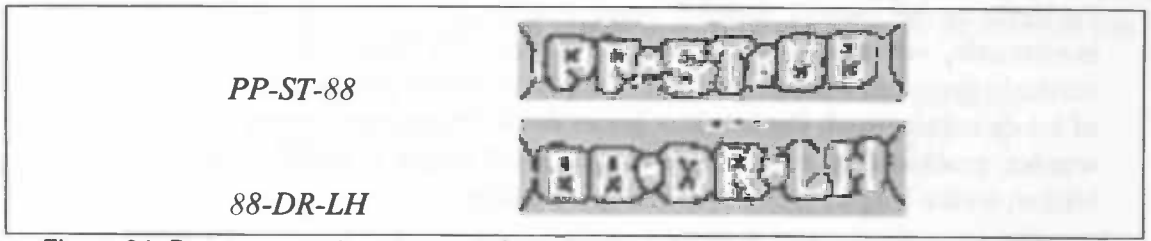
Figuur 20: Postprocessing in de vorm van het verwijderen van kleine objecten uit de labeling. Over het algemeen blijkt dat het filteren van objecten kleiner gelijk objectgrootte 2 leidt tot een leesbaardere labelling.

4.2.2.6 Karakteristieke basis

Om een optimale feature basis te verkrijgen moeten verschillende keuzes worden gemaakt. Deze keuzes betreffen het wel of niet toepassen van een filter, een juiste waarde voor de threshold, het labelalgortime en de minimale grootte van een toegestaan object bij postprocessing. Deze waarden zijn voor de voor- en achtergrond apart experimenteel bepaalt.

Om de voorgrond effectief te segmenteren, blijken twee methoden te concurreren. Toepassing van een median-filter in combinatie met een lage threshold van *1%* levert vergelijkbare prestaties na het labellen op, in vergelijking met een iets hogere threshold van *2%* zonder filtering. De voordelen van een gladder oppervlak wegen niet volledig op tegen het gebruik van de originele data. Vandaar dat beide methoden zullen worden aangewend bij de herkenningsfase. Om geen informatie te verliezen betreffende de voorgrond worden twee in verschillende richting gelabelde images gecombineerd met een logische ***OR*** operatie. Hierbij wordt in het resulterende image een pixel tot de voorgrond gerekend als dat ook het geval is in één van de twee gegeven gelabelde images. Met behulp van postprocessing worden alle objecten kleiner of gelijk 2 pixels uit het beeld verwijderd.

De vorm van ravijnen in de achtergrond vertelt wat over de vorm van de voorgrond. Zoals uit figuur 21 blijkt, bevatten gesloten curven kruisende ravijnen en bevindt zich tussen twee lijnsegmenten een ravijn. De hoeveelheid en grootte daarvan vertellen iets over het karakter. Een goed geïsoleerd image omsluit het karakter nauw. Dit betekent dat vooral de achtergrond informatie die zich *in* het karkater bevind zal worden gebruikt. Dit zou problemen op kunnen leveren bij bepaalde karakters. De locatie van een ravijn bij de *'T'* in het voorbeeld is afhankelijk van de afstand tot de voorgrond van de al dan niet aanwezige buren. Dat kan ervoor zorgen dat het ravijn net buiten het geïsoleerde image valt. Hetzelfde geldt voor de *'J'* en de *'L'*. Deze informatie wordt voor kennisgeving aangenomen. Na experimenteren blijkt dat links-recht diagonale labelling het best de achtergrond representeert, nadat het beeld een median filter heeft gepasseerd. Er wordt hierbij geen threshold ingevoerd bij de pixel vergelijkende operaties. Met behulp van postprocessing worden alle objecten kleiner of gelijk 2 pixels uit het beeld verwijderd.



Figuur 21: De vorm van de achtergrond wordt gevonden na labelling in diagonale richting. Over het algemeen resulteerde dat in de best leesbare *basis.*

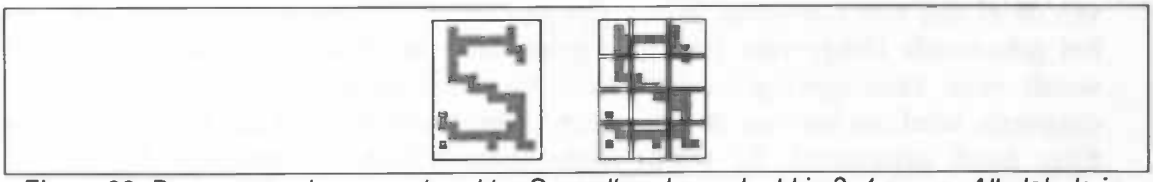
In appendix C is een lijst met nummerborden opgenomen met daarop karakters die na het thresholdonderzoek in klasse B belanden. Daar worden de resultaten van topografische labelling tegenover gezet, zodat een voorlopige indruk kan worden verkregen over mogelijke prestaties.

4.2.2.7 Topografische features extraction

Nu bekend is op welke wijze de karakteristieke basis is verkregen, zullen van deze basis features worden ontleend die zich lenen voor neurale classificatie. Feature extraction wordt toegepast om een reductie in de grootte van de data te bewerkstelligen en om bruikbare karakteristieken te vinden. In principe bestaat er een groot scala aan feature extractors die kunnen worden toegepast.

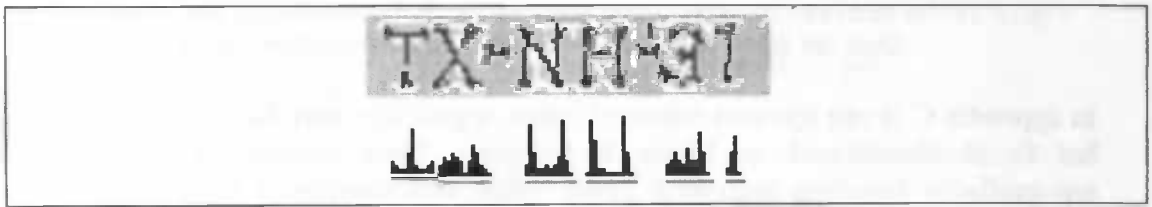
De selectie van een feature extractor is uitermate belangrijk voor het bereiken van grote herkenning. Vooraf kan er weinig tot niets worden gezegd over de resultaten van een bepaalde feature extractor, mede doordat neurale classificatie niet direct inzichtelijk is. Trier, Jain en Taxt beschrijven verschillende mogelijkheden die op uiteenlopende karakter representaties kunnen worden losgelaten [5]. Daarvan zijn de zernike of geometrische momenten, zoning en projectie toepasbaar op de ontwikkelde karakteristieke basis. Pseudo Zernike momenten worden verderop al toegepast op grijswaardenplaatjes en zullen hier niet worden gebruikt. Van geometrische momenten zijn de prestaties reeds bekend. Zoning en projectie zullen wel worden toegepast.

Bij zoning wordt een grid over een karakter gelegd. (Zie figuur 22) De dimensies van het grid worden kleiner gekozen dan de dimensies van een karakter. Alle voor- of achtergrond pixels die in een zone vallen worden bij elkaar opgeteld. Dit veroorzaakt een reductie in data, die bepaald wordt door de grootte van de zones. Er zullen verschillende zonegroottes worden gebruikt. Na classificatie blijkt dan welke zonegrootte het best kan worden toegepast. Elk gelabeld karakter levert na zoning twee feature vectoren op: één voor de voorgrond en één voor de achtergrond.



Figuur 22: De voorgrond van een karakter S wordt onderverdeeld in 3x4 zones. Alle labels in een zone worden opgeteld. Zodoende wordt een feature vector gevonden.

Projectie is het proces waarbij geselecteerde pixels worden geprojecteerd op een horizontale, verticale of diagonale as. (Figuur 23) Hier zullen alleen horizontale en verticale projectie worden gebruikt. De geselecteerde pixels behoren tot de voorgrond of tot de achtergrond. Per karakter levert dat 4 feature vectoren op. De vectoren zullen worden geschaald naar een vaste lengte. Deze lengte is variabel en na het testen zal blijken welke lengtes prevaleren boven de andere.



Figuur 23: Horizontale projectie van voorgrondpixels. Deze histogrammen worden geschaald naar een vaste omvang en doen dan dienst als een feature vector.

4.2.3 Pseudo Zernike momenten

Deze paragraaf behandelt de implementatie van pseudo Zernike momenten. Hierbij spelen verschillende afwegingen een rol. Visueel gelijkende karakters zouden een vergelijkbare feature vector moeten krijgen. Maar de geïsoleerde karakters zijn verschillend in omvang en belichting. Om zo generiek mogelijk pseudo zernike momenten te kunnen toepassen, is het noodzakelijk deze twee eigenschappen te invariaren over de dataset. Invariantie in omvang kan gemakkelijk worden ingebouwd bij de daadwerkelijke berekening van de pseudo Zernike momenten. Invariantie in belichting vergt wat meer contemplatie. Daarna volgt de vaststelling omtrent welke ordes en lengtes van momenten zullen worden toepast. Dit levert het implementatie traject op, zoals vermeld in figuur 24.

Figuur 24: Implementatie traject voor pseudo Zernike momenten. Nadat een geïsoleerd grayscale karakter invariant is gemaakt onder belichting worden via pseudo Zernike momenten een feature vector gevonden.

4.2.3.1 Illuminantie Invariantie

Door beelden invariant te maken onder belichting, worden ze onderling beter vergelijkbaar op inhoud. Dit blijkt geen eenvoudige taak. Eén van de weinige operaties die illuminatie invariantie bewerkstelligt is het gebruik van gradiënten van beelden.

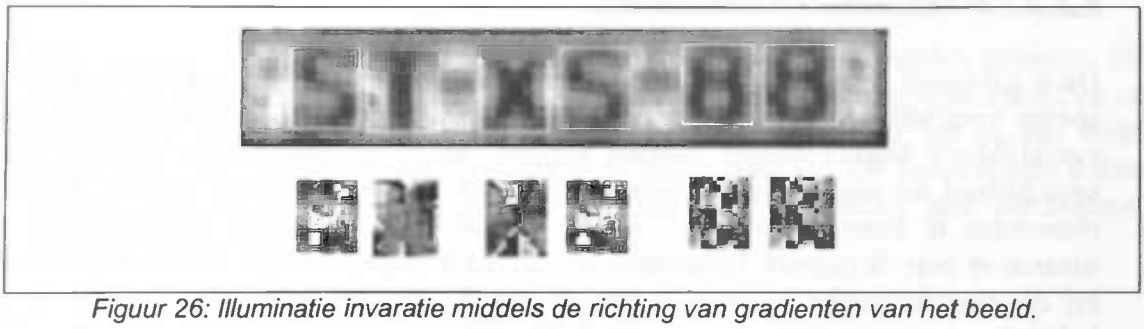
$$s_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad s_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Figuur 25: Verticale en Horizontale Sobel-kernels

Het gradiënt van een beeld wordt berekend door twee Sobel convolutie uit te voeren. Eén verticaal en één horizontaal. (zie figuur 25) Gradiënten hebben twee eigenschappen. De richting van een gradiënt wijst in de richting van de grootste oplopende curve. De magnitude van een gradiënt is een maat voor de steilheid in die richting.

$$Richting \quad \theta_{gradient} = \arctan\left(\frac{s_y}{s_x}\right) \quad , \quad Magnitude \quad M_{gradient} = \sqrt{s_y^2 + s_x^2}$$

In tegenstelling tot de richting van een gradiënt verhult de magnitude geen informatie over de topologie van een beeld. De magnitude komt niet van toepassing bij illuminatie invariantie. De richting van gradiënt, vanaf nu gradiënt genoemd, wijst in de richting van een edge en altijd bergop. De richtingen op een vlak karakteriseren een beeld en zijn onafhankelijk van de belichtingsintensiteit. (zie figuur 26) De distributie van richtingen over een vlak representeren een karakter. Van deze distributie kunnen features worden ontleend die het beeld representeren.



Figuur 26: Illuminatie invaratie middels de richting van gradienten van het beeld.

4.2.3.2 Pseudo Zernike momenten

De basis van pseudo Zernike momenten is de set van orthogonale pseudo Zernike polynomen, gedefinieerd als poolcoördinaten over de eenheidscirkel. De 2-dimensionale pseudo Zernike momenten van orde p met herhaling q van een beeld wordt gedefinieerd door formule 1. Er zijn $(n+1)^2$ polynomen van orde $\leq n$.

$$Z_{pq} = \frac{p+1}{\pi} \int_0^1 \int_{-\pi}^{\pi} R_{pq}(r) e^{-jq\theta} f(r,\theta) r dr d\theta \qquad (1)$$

waar

$$r = \sqrt{x^2 + y^2} \quad , \quad \theta = \tan^{-1}\left(\frac{y}{x}\right) \quad , \quad -1 < x, y < 1 \quad ,$$

$$R_{pq}(r) = \sum_{s=0}^{p-|q|} (-1)^s \frac{(2p+1-s)!}{s!(p+|q|+1-s)!(p-|q|-s)!} r^{p-s}$$

$$0 \leq |q| \leq p \quad , \quad p \geq 0$$

Omdat het gemakkelijker werken is met reële functies wordt Z_{pq} opgedeeld in een reëel en een imaginair gedeelte Z_{pq}^C en Z_{pq}^S. Voor de uiteindelijke uitkomst van een moment wordt dan de euclidische afstand berekend. $PZ_{pq} = \sqrt{\left(Z_{pq}^C\right)^2 + \left(Z_{pq}^S\right)^2}$

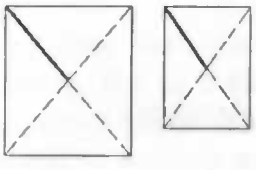
$$Z_{pq}^C = \frac{2(p+1)}{\pi} \int_0^1 \int_{-\pi}^{\pi} R_{pq}(r) \cos(q\theta) f(r,\theta) r dr d\theta \qquad (3)$$

$$Z_{pq}^S = \frac{2(p+1)}{\pi} \int_0^1 \int_{-\pi}^{\pi} R_{pq}(r) \sin(q\theta) f(r,\theta) r dr d\theta \qquad (4)$$

waar $p \geq 0$ en $q > 0$

4.2.3.3 Implementatie

Voor de implementatie is gebruikt gemaakt van code die is afgeleid van [10]. Alhoewel deze code gebruikt wordt voor momenten van binaire beelden is deze na een kleine aanpassing geschikt voor grijswaardenbeelden. De pseudocode is te vinden in appendix D.



Figuur 27: Het complete beeld wordt geschaald op de eenheidscirkel. Hierbij wordt schaalinvariatie bereikt doordat elk beeld, ongeacht zijn grootte, wordt geschaald.

Schaalinvariatie wordt bereikt doordat de beelden worden geprojecteerd naar de eenheidscirkel. Hierin wordt elke afstand tot het centrum gedeeld door de maximale afstand aanwezig in het beeld. Dat is de halve diagonaal. (zie figuur 27)

Er is reeds geëxperimenteerd met optical character recognition via pseudo Zernike momenten. Hierbij zijn de ordes 6 tot en met 8 met alle bijbehorende herhalingen toegepast op binaire beelden.

4.2.4 Overzicht van features

De keuze van twee feature extractors leidt na toespitsing op het werkdomein tot een keur aan feature extraction methoden.

				# features
Pseudo Zernike Features	Ordes tot en met 6			27
	Ordes tot en met 7			35
	Ordes tot en met 8			44
Topografische Features	Zoning	3x4	Fg	12
			Fg + Bg	24
		4x5	Fg	20
			Fg + Bg	40
		5x6	Fg	30
			Fg + Bg	60
		6x7	Fg	42
			Fg+Bg	84
	Projectie	8x10	Fg	18
			Fg + Bg	36
		10x12	Fg	22
			Fg + Bg	44
		12x14	Fg	26
			Fg + Bg	52

Tabel 5: Overzicht van features. Er is een grote diversiteit vanwege het feit dat vooraf niets kan worden voorspeld over het succes van features bij neurale klassificatie.

Van de pseudo zernike features zullen de ordes 6 tot en met 8 worden gebruikt. Bij zoning en projectie van topografische labels zullen verschillende vectorgroottes en zonegrootes in acht worden genomen. In het volgende hoofdstuk zal na een reeks testen duidelijk worden welke methoden het best presteren. De features die zullen worden gebruikt om karakters neuraal te classificeren, worden hier samengevat gepresenteerd in tabel 5.

4.2.5 Principal Component Analysis

Zoals duidelijk is geworden uit tabel 5 kunnen de cijferreeksen die ontstaan na feature extraction grote lengtes aannemen. Zo'n cijferreeks kan redundantie bevatten. Dit kan worden onderzocht middels principal component analysis. PCA is een wiskundige methode die een aantal gecorreleerde variabelen transformeert naar een kleinere set van ongecorreleerde variabelen, genaamd principal components. Deze principal components bevatten zoveel mogelijk variabiliteit van de data als gespecificeerd. Hier wordt gekozen om de feature reeksen ontstaan na toepassing op de trainset (verderop omschreven) te reduceren tot *99%* van de variatie in de data. Zodoende wordt de data gereduceerd tot een handbare set, terwijl de variatie in de dataset behouden blijft. De reductie die dat tot gevolg heeft staat vermeld in tabel 6.

			# features	# pc
Pseudo Zernike Ordes tot en met 6			27	24
Pseudo Zernike Ordes tot en met 7			35	30
Pseudo Zernike Ordes tot en met 8			44	37
Zoning	3x4	Fg	12	12
		Fg + Bg	24	23
	4x5	Fg	20	20
		Fg + Bg	40	28
	5x6	Fg	30	29
		Fg + Bg	60	28
	6x7	Fg	42	31
		Fg+Bg	84	25
Projectie	8x10	Fg	18	18
		Fg + Bg	36	28
	10x12	Fg	22	21
		Fg + Bg	44	28
	12x14	Fg	26	24
		Fg + Bg	52	28

Tabel 6: Overzicht van features en de verkregen principal components

De features verkregen door pseudo Zernike momenten vertonen relatief weinig redundante informatie. Gezien de aard van deze momenten (een versimpelde datarepresentatie) niet onverwacht.

Wat opvalt is de redundantie van achtergrond informatie in de zones en de projectievectoren. Dit is echter logisch wanneer men zich realiseert dat er een verband bestaat tussen de voorgrond van een karakter en de achtergrond van een karakter. De verwachting kan dan ook uitgesproken worden dat de variatie in herkenning bij zoning en projectie apart niet veel zal gaan verschillen.

Voordat principal component analysis kan plaatsvinden wordt de data eerst genormaliseerd naar normale verdeling. Het gemiddelde van één feature wordt naar $\mu = 0$ getransformeerd en de deviatie wordt geschaald naar $\sigma^2 = 1$. De transformatie die gepaard gaat met de principal component analysis wordt bewaard in een matrix. Deze transformatie variabelen worden opgeslagen per trainset per feature, zodat andere datasets (validatie en testsets) dezelfde transformatie kunnen ondergaan.

4.2.6 Tussenbeschouwing

De beantwoording van deelvraag 2 heeft nu de 2^e fase afgerond. In fase 1 zijn methoden geselecteerd die features van grayscale beelden kunnen ontwaren. In fase 2 zijn er keuzes gemaakt omtrent de implementatie van deze methoden. Zodoende werden ze toegespitst op het domein van nummerbordherkenning. Dit heeft geleid tot een set van features.

Deze set van features wordt getest op bruikbaarheid in de volgende fase, die wordt ingegaan in de volgende paragraaf.

4.3 Neurale classificatie

4.3.1 Inleiding

Door de gekozen features neuraal te classificeren wordt inzicht verkregen in het succes van de gekozen features en de bijdrage aan karakter en nummerbordherkenning. Dat zal dan ook in deze paragraaf plaatsvinden. Voordat er resultaten gepresenteerd kunnen worden zal hier eerst een opzet voor worden gemaakt. Het blijkt dat er vele mogelijkheden zijn om te behandelen. Hieruit zal een keuze worden gemaakt door een strategie te kiezen, met behulp waarvan de resultaten de meeste betekenis bevatten.

4.3.2 Opzet

Het doel van dit deelonderzoek is het bepalen van het succes van grijswaarden feature extraction. De route daarnaar toe is tweeledig. Eerst wordt er toegewerkt naar twee succesvolle neurale netwerken, voor elke soort feature extractor één. Vervolgens worden deze twee netwerken getest.

Een neuraal netwerk, in de vorm van een feedforward multilayer perceptron, is een conventioneel instrument om classificaties mee te bewerkstelligen. [17] Het bestaat uit een input layer van neuronen, met voor elk inputneuron een element van de inputvector, een hidden layer, en een output layer, met voor elke klasse een outputneuron. Modellen met meerdere hidden layers zijn zowel denkbaar als mogelijk, maar niet noodzakelijk. [18] Elk input neuron is verbonden met alle hidden neuronen, elk hidden neuron is verbonden met alle output neuronen. (Fully connected) Een input neuron verdeelt de input van dat neuron over de hidden layer. Alle andere neuronen gebruiken een overgangsfunctie voor het berekenen van hun output. Als overgangsfunctie wordt gekozen voor de sigmoid. Van deze functie is bekend dat hij gebruikt kan worden in combinatie met patroon herkenning. De gewichten van alle verbindingen en bias's van alle neuronen worden gevonden door het terug berekenen van fouten in de outputs. (Error back-propagation) Er bestaan verschillende trainingsmethoden met verschillende karakteristieken en prestaties. Hieruit zal een keuze worden gemaakt.

Het aantal neuronen in de hidden layer ligt niet vast. Door te variëren met deze aantallen kan een succesvolle structuur worden geselecteerd. Dit is bovendien noodzakelijk vanwege het vooraf oncontroleerbaar fenomeen van generalisatie in het netwerk. Te veel hidden neuronen kan leiden tot het *onthouden* van aangeboden input in plaats van het ontdekken van een *regel* tot classificatie. Te weinig hidden neuronen kan leiden tot een netwerk dat te weinig complexiteit bevat om überhaupt een *regel* aan te leren. Experimenteren met verschillende aantallen hidden neuronen zou moeten leiden tot een tevredenstellende selectie van structuur.

Het aantal output neuronen is tevens aan variatie onderhevig. Op het eerste gezicht zou elke klasse een eigen outputneuron kunnen krijgen. Dat levert, met uitsluiting van op nummerborden ontbrekende letters, 27 klassen op. Vanwege de ambivalente relaties tussen sommige letters en cijfers, in combinatie met de a priori kennis omtrent de syntax van nummerborden kunnen ook minder klassen worden gekozen. Wanneer elk van de B/8, 2/Z, 5/S en 0/D combinaties tot één klasse worden gerekend levert dat in totaal 23 klassen op.

4.3.3 Strategie

Implementatie van alle mogelijke netwerken is mogelijk, maar tijdrovend. Daarnaast kan met behulp van een goede strategie een selectie worden gemaakt. Hierbij is het doel om twee netwerken te selecteren die het best presteren. Deze netwerken worden vervolgens uitgebreid getest. De selectieprocedure zal als volgt plaatsvinden.

(1) Selectie van trainingsmethoden.
(2) Reductie van feature extractors
(3) Selectie van beste feature extractors in combinatie met aantal hidden layers

(1)
Er staan verschillende error backpropagation methoden ter beschikking om feedforward netwerken mee te trainen.[19] Al deze train-methoden zijn batch error-propagation methoden. Dat betekent dat eerst alle patronen worden aangeboden alvorens de gewichten worden aangepast. Op voorhand kan geen enkele worden afgewezen vanwege zijn (wan)prestatie op patroonherkenning. Door alleen de trainingsmethode te variëren en andere variabelen vast te leggen wordt een zo duidelijk mogelijk overzicht verkregen over de prestatie van een trainingsmethode op de gekozen netwerktopologie. Hieruit worden de beste trainingsmethoden geselecteerd. De rest wordt niet meer gebruikt.

Gradient Descent		
	(gdx)	gradient descent with momentum and adaptive learning rate
Conjugate Gradient Algorithms		
	(cgf)	Fletcher-Reeves update
	(cgp)	Polak-Ribiére update
	(cgb)	Powell-Beale restarts
	(scg)	Scaled Conjugate Gradient
Quasi Newton		
	(bgf)	BFGS Algorithm
	(oss)	One Step Secant Algorithm
Anderen		
	(rp)	resilient backpropagation
	(lm)	Levenberg-Marquardt

Verschillende error back propagation neural network trainmethoden. [19]

(2)
De topografische feature extractors kunnen worden toegepast op voorgrond informatie alleen of op voor- én achtergrond informatie. Door middel van een simpele test met de beste trainingsmethode en gebruik van verschillende hidden neuronen kan inzicht worden verkregen of het gebruik van achtergrond informatie daadwerkelijk iets toevoegt aan de herkenning. Dit reduceert het aantal topografische feature extractors tot de helft.

(3)
Alle geselecteerde feature extractors worden nu getraind met de beste trainingsmethoden en met verschillende opmaak van de hidden layer. Hieruit worden de twee beste feature extractors geselecteerd, eentje voor zernike features en eentje voor topografische features

4.3.4 Methode

4.3.4.1 Netwerkselectie

De data worden willekeurig opgedeeld in drie sets. Een trainingset, een verificatieset en een testset, in de verhouding *(trainset : validatieset : testset) = (7 : 1 : 1)*. Elk soort karakter zal dan gemiddeld ongeveer 96 keer in de trainingsset en 14 keer in de test en verificatieset voorkomen.

De verificatieset wordt ingezet om een praktisch en relevant netwerk op te leveren. Een training wordt gestopt als tijdens het trainen de prestaties van het netwerk op de validatieset en trainingset te ver uit elkaar lopen. Dit voorkomt een netwerk dat te specifiek is afgestemd op de trainingset. Wanneer de prestaties van valdiatieset en trainingset ongeveer gelijk zijn betekent dat de prestaties van het netwerk in de praktijk ongeveer gelijk zal zijn.

Alle gepresenteerde resultaten verkregen in de netwerkselectie zijn percentages correct geclassificeerde karakters van de testset. Hiervoor is een klein reject criterium in het leven geroepen om te voorkomen dat uitkomsten worden vertroebeld door minimale verschillen. Dit criterium laat juist geclassificeerde winnende outputs ter minimale grootte *0,3* met een minimaal verschil met de runner-up van *0,1* doorgaan als correct.

Het trainen wordt tijdens de netwerkselectie stopgezet als de error *(mse)* onder de 0.01 duikt of als een validatiestop plaatsvindt. Deze stop vindt plaats wanneer het verschil in prestatie *(mse)* tussen training en validatie 0.01 bedraagt.

4.3.4.2 Testen

Nadat bekend is welke feature extractors in combinatie met een bepaald aantal hidden neuronen het best presteren, worden deze uitgebreid getest. Hiertoe vind een heropdeling van de dataset plaats, waarbij de testset uitgebreid wordt ten koste van de trainset. De verhouding *(trainset : validatieset : testset) = (6 : 1 : 2)*. Hierdoor heeft

de testset meer karakters zodat de tests betrouwbaarder worden. Nadeel is dat door minder aangeboden trainpatronen het netwerk niet generaliseert. Dit laat zich echter controleren door de reeds aanwezige resultaten.

Vervolgens worden de gekozen netwerkconfiguraties 5 keer getraind. Hierbij wordt gebruikt gemaakt van een lagere MSE-error als doel. Deze wordt *mse = 0.005* gekozen. De training wordt wederom gestaakt als de het verschil in prestaties (*mse*) tussen validatie en trainset te ver uit elkaar lopen *(>0.01)*. Uit de testset worden de 7 niet lege subsets van alle mogelijke combinaties van klassen uit het thresholdonderzoek geëxtraheerd. De gemiddelde prestaties van al deze subsets in combinatie met verschillende reject regions zullen worden onderzocht.

Het verdient benadrukking dat de reject regions in de strategie een andere functie hebben dan bij het testen. In de strategie wordt eerst de *error* bepaald en vervolgens *reject* en *correct*, ter voorkoming van een fotofinish winaar. Bij het testen wordt eerst *reject* bepaald en daarna *error* en *correct*, omdat dit de praktijkprestaties representeert. Om te onderzoeken of er winst valt te halen uit het reduceren van het aantal outputklassen worden de netwerken met zowel volledige output en als gereduceerde output getraind en getest.

Het doel tijdens het testen is het reduceren van de error van de netwerken. Dat gaat ten koste van meer rejections. De mate waarin errors gemaakt mag worden ligt aan de context van het gebruik in de praktijk. Sommige situaties vragen om een uitermate lage error. Dat drukt op de herkenning. Vanwege de context van nummerbordherkenning wordt gekozen om de error laag te houden. Dat houdt in maximaal enkele procenten.

4.3.5 Resultaten

4.3.5.1 Netwerkselectie

(1)
Vastgesteld wordt welke trainingsmethoden het best presteren. Daarvoor moeten de overige variabelen vast worden gekozen. De trainingen zullen plaatsvinden voor zernike features van orde 6 en topografische zoning van 3x4 met alleen voorgrond. Verder wordt het aantal hidden neuronen vastgesteld op 40. De resultaten zijn te vinden in tabel 7.

%	Gdx	Rp	Cgf	Cgp	Cgb	Scg	Bfg	Oss	Lm
pseudo zernike 6	***52.3***	***54.4***	*15.4*	*28.1*	*46.1*	*30.7*	*18.8*	*3.9*	*46.6*
zoning 3x4 FG	*63.7*	***66.9***	*17.2*	*58.3*	*40.6*	***66.7***	*15.1*	*1.3*	*61.5*

Tabel 7: Prestatie van verschillende trainmethoden.

Hieruit kan worden opgemaakt dat de trainingsmethoden resilient backpropagation, gradient descent met momenten en adaptive learning rate, scaled conjugate gradient en levenberg-marquardt bij features door zoning goed presteren. De laatste twee

presteren wat minder bij zernike features. Ondanks de prestaties van Levenberg-Marquardt trainingsmethode wordt deze niet gebruikt in verband met de enorm lange trainingstijden en geheugeneisen. Per soort feature extractor worden twee trainingsmethoden gekozen. Deze zijn in de tabel verdikt aangegeven.

(2)
Vervolgens wordt getest of achtergrond informatie verkregen door topografische analyse überhaupt bijdraagt aan de herkenning van karakters. Hiertoe worden de verschillende topografische features door zoning en projectie geclassificeerd met de trainingsmethode resilient backpropagation, in combinatie met verschillende aantallen hidden neuronen. De resultaten zijn te vinden in tabel 8.

	25	30	35	40	45	50
Zoning 5x6 FG	71.9	71.9	71.4	72.7	72.4	71.6
Zoning 5x6 FG+BG	**78.1**	**77.6**	**76.6**	**75.3**	**77.9**	**73.7**
Zoning 6x7 FG	71.9	72.9	74.2	73.7	74.7	74.5
Zoning 6x7 FG+BG	**78.4**	**74.2**	**75.8**	**77.3**	**77.9**	70.1
Projection (8x10) FG	43.8	47.4	47.7	49.5	45.8	48.7
Projection (8x10) FG+BG	**51.6**	**49.5**	**49.7**	**51.0**	**51.0**	**52.1**
Projection (10x12) FG	43.8	46.6	46.6	45.3	46.1	**50.8**
Projection (10x12) FG+BG	**49.0**	**49.2**	**52.9**	**53.6**	**49.5**	50.3

Tabel 8: Vergelijking tussen prestaties topografische features met en zonder achtergrondinformatie. Het gebruik van achtergrond informatie leidt tot een kleine winst in de herkenningspercentages.

Er kan worden geconstateerd dat de informatie over de achtergrond bijdraagt aan de classificatie van karakters. Op twee uitzonderingen na zijn alle herkenningspercentages bij achtergrondinformatie groter dan zonder. Vanaf nu wordt de informatie over de achtergrond toegepast bij verdere experimenten.

(3)
De verschillende feature extractors worden nu getraind met de geselecteerde trainingsmethoden en attributen. De zernike features worden getraind met resilient backpropagation en met gradient descent met momentum en adaptive learning rate. Hierbij varieert het aantal hidden neuronen met de grootte van de input. De neurale netwerken bij topografische features worden getraind met resiliant en scaled conjugate gradient backpropagation. Elke combinatie wordt 3 keer getraind. Tabel 9 laat de gemiddelden zien.

		TrainRP					TrainGDX				
	h	*h-4*	*h*	*h+4*	*h+8*	*h+12*	*h-4*	*h*	*h+4*	*h+8*	*h+12*
PZernike: 6	24	54.5	52.6	51.3	54.7	56.5	44.5	52.9	53.1	47.7	56.3
PZernike: 7	30	53.9	55.5	56.6	56.4	57.9	49.2	53.9	52.6	51.8	55.2
PZernike: 8	37	58.1	57.3	56.5	**59.6**	57.6	58.6	49.0	54.2	53.4	57.2

Tabel 9: Resultaten van pseudo Zernike features met verschillende aantallen hidden neuronen en trainmethoden.

TrainRP

	20	25	30	35	40	45	50
Zoning 3x4	68.3	69.9	70.9	69.9	70.8	70.1	71.6
Zoning 4x5	71.1	72.2	71.9	71.6	73.8	72.3	72.4
Zoning 5x6	77.7	76.2	77.7	77.4	78.5	**78.7**	78.1
Zoning 6x7	75.4	76.2	75.4	76.1	77.3	75.4	77.6
Projectie 8x10	47.5	50.6	52.2	50.3	51.0	51.3	52.2
Projectie 10x12	47.9	49.1	50.4	49.4	49.9	50.4	51.2
Projectie 12x14	50.1	50.7	52.1	51.0	51.5	50.4	50.7

TrainSCG

	20	25	30	35	40	45	50
Zoning 3x4	43.7	37.2	39.7	31.7	27.6	52.4	45.5
Zoning 4x5	45.1	26.8	45.2	38.5	57.2	24.6	29.2
Zoning 5x6	42.9	32.6	53.7	47.1	65.8	31.7	52.5
Zoning 6x7	53.0	47.3	34.5	35.9	42.4	74.6	53.9
Projectie 8x10	27.7	23.0	26.0	28.1	32.0	22.0	21.4
Projectie 10x12	27.1	23.1	24.9	17.9	20.1	27.7	28.8
Projectie 12x14	22.7	32.0	27.4	25.3	20.1	38.9	32.6

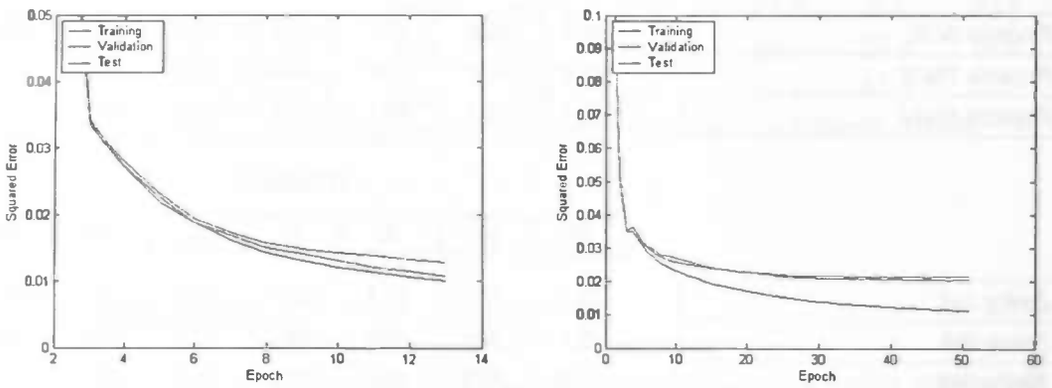
Tabel 10: Prestaties van topografische features met verschillende trainmethoden en hidden neuronen

De resilient backpropagation trainingsmethode levert de beste resultaten voor zowel de pseudo Zernike momenten als bij de topografische features. Uit deze test worden de twee beste netwerken geselecteerd. Alhoewel sommige verschillen minimaal zijn wordt er voor gekozen simpelweg de gemiddeld best presterende methoden te kiezen. Deze zijn in de tabel 10 aangegeven en worden hieronder opgesomd.

- Pseudo Zernike momenten
  - orde 8
  - hidden neuronen 45
  - features 44
  - features na pca 37
- Zoning van topografische features
  - Gebruik gemaakt van voor- en achtergrond
  - Zonegrid 5x6
  - Hidden neuronen 45
  - Features 60
  - Features na pca 28

Deze twee netwerken zullen vanaf hier respectievelijk Zernike en Zoning genoemd worden.

In figuur 28 staan twee grafieken met daarin de prestaties van de datasets tijdens de training van de twee netwerken. Hieruit kan worden opgemaakt dat de neurale netwerken convergeren naar een oplossing door de steeds kleiner wordende helling. Verder blijkt daaruit dat de gekozen opsplitsing in de 3 datasets goed gekozen is geweest. Weliswaar lopen de prestaties van trainset tegenover de validatie- en testset uiteen, de validatie en testset ontlopen elkaar niet veel. De reden dat de trainset een kleinere error kent is vanwege het feit dat het netwerk zich aanpast aan de trainpatronen en niet aan alle patronen.

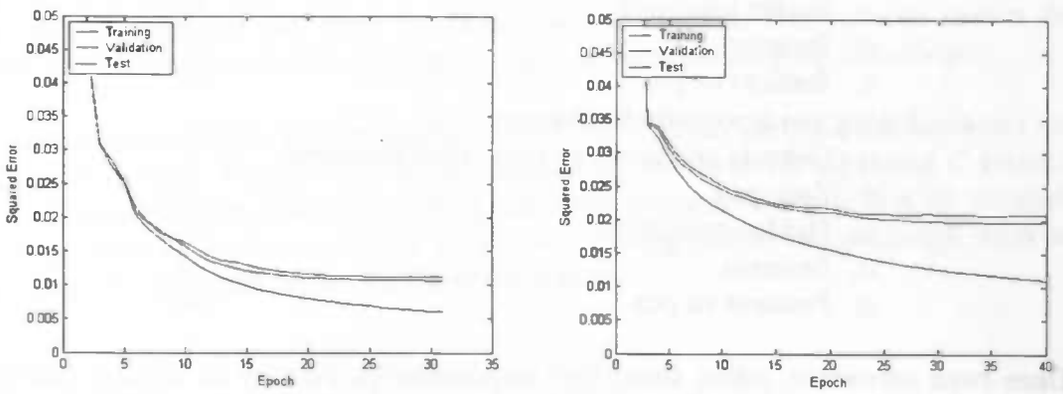


Figuur 28: Trainingprogress van respectievelijk Zoning en Zernike

4.3.5.2 Testen

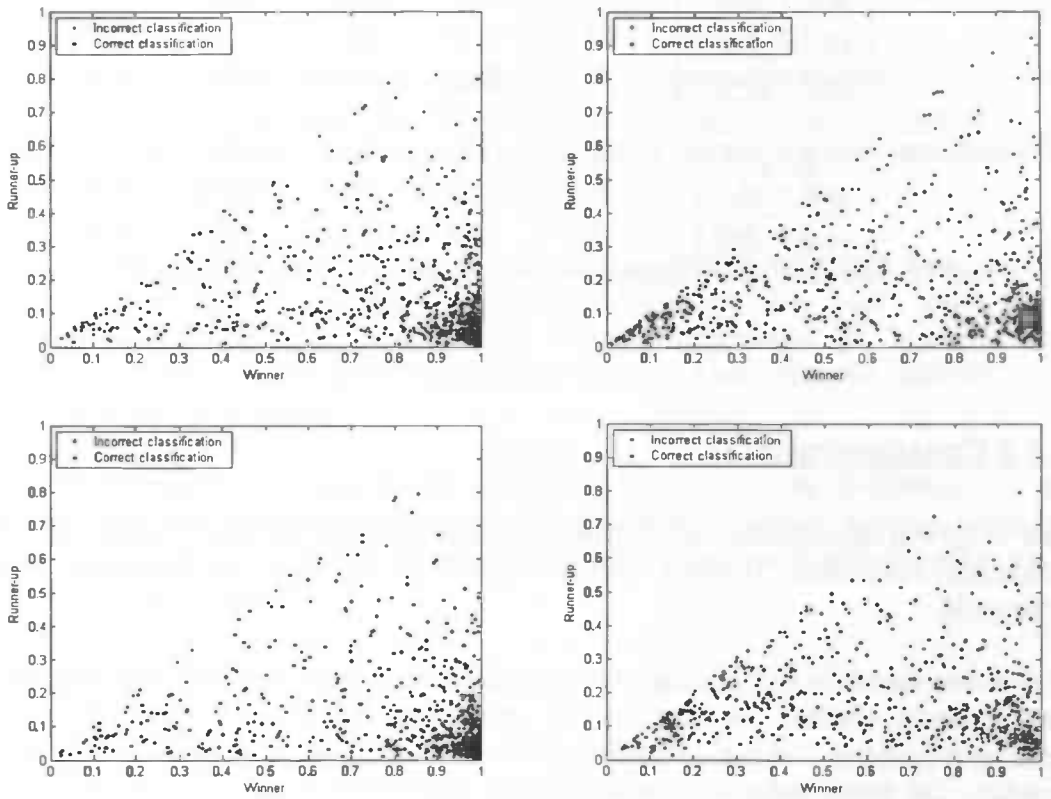
De twee geselecteerde netwerken zijn elk 5 keer geconstrueerd en getraind met de gevonden topologieën en van gebruikmaking van de herschikking van de data. Daarna heeft een zelfde proces plaatsgevonden, maar dan met een gereduceerde output. Vervolgens zijn ze getest met alle mogelijk niet lege combinaties (7 stuks) van klassen gevonden na het threshold onderzoek. Het trainingsproces zal worden besproken, alvorens de gemiddelde resultaten besproken zullen worden. Dat resulteert in een reeks constateringen, waarin alle opmerkelijkheden vermeld zullen worden.

Eerst zal opnieuw worden vastgesteld of de nieuwe dataopsplitsing goed gekozen is geweest door te kijken naar plots van de trainingprogress bij normale input. De validation- en testset ontlopen elkaar wederom weinig, oftewel een goed verdeling. (zie figuur 29)



Figuur 29: Trainingprogress van respectievelijk Zoning en Zernike na heropdeling datasets.

De prestaties van een neuraal netwerk laten zich gemakkelijk weergeven middels een RDO plot. (Reduced discriminant output). Hierin krijgt elk patroon uit de testset een stipje in het plot, waarin de waarde van de winnaar wordt uitgezet tegen de waarde van de runner-up. De kleur bepaalt de correctheid van de classificatie.



Figuur 30: RDOplots van de complete testset op vier soorten netwerken. In leesvolgorede Zoning met complete output, zoning met gereduceerde output, Zernike met complete output en Zernike met gereduceerde output.

Een hoge winner in combinatie met een lage runner-up is bij een goed getraind netwerk een indicatie van een correcte classificatie. De grote concentratie blauwe stippen rechts onderin de grafieken van figuur 30 bevestigen dat. Om de misclassificaties zoveel mogelijk te beperken kan een rejection criterium worden ingevoerd. Hierbij moet de winnaar voldoen aan een minimum waarde α en moet het verschil tussen winnaar en runnerup een minimale waarde β zijn. Aan de hand van een rdoplot kan een goed rejection criterium worden gekozen. Er wordt echter numeriek onderzoek gedaan naar de effecten van het gebruik van verschillende rejection criteria op de herkenning. Alle uitslagen daarvan zijn te vinden in Appendix E. In tabel 11 is een samenvatting te vinden van de prestaties van de complete testsets bij verschillende criteria.

		Zoning			Zernike		
		Correct	Rejected	Error	Correct	Rejected	Error
27 outputklassen	α=0.1 β=0.0	82.1	1.9	16.0	63.0	5.7	31.2
	α=0.3 β=0.1	78.5	12.3	9.2	55.7	28.7	15.6
	α=0.5 β=0.3	71.9	22.2	5.9	48.1	41.7	10.2
23 outputklassen	α=0.1 β=0.0	87.0	1.4	11.9	67.2	2.3	30.5
	α=0.3 β=0.1	81.5	11.5	7.0	59.3	25.9	14.8
	α=0.5 β=0.3	75.5	19.8	4.7	49.7	41.8	8.5

Tabel 11: Tabel met testresultaten van de complete testset bij verschillende criteria.

4.3.6 Constateringen

Hier volgt een opsomming van constateringen die gemaakt kunnen worden omtrent de verkregen resultaten. Hieraan zullen zo nodig oorzaken en gevolgen worden gekoppeld.

Wat direct opvalt is dat Zoning over de gehele linie beter presteert dan Zernike. De fundamentele oorzaak die hier aan te grondslag ligt is het feit dat de features verkregen middels Zoning de karakters robuuster representeren dan de features van Zernike. De twee volgende constateringen onderstrepen dat. Ten eerste verloopt classificatie van smerige karakters bij Zoning relatief beter dan bij Zernike. Ten tweede resulteert het reduceren van de output in combinatie met het opkrikken van de rejectregion in een groter verlies van herkenning bij Zernike dan bij Zoning. De prestatie van pseudo Zernike momenten op binaire beelden is, in vergelijking met de resultaten die hier verkregen zijn, goed te noemen. [10] Hoogstwaarschijnlijk is de wijze waarop illuminatie invariantie verkregen wordt niet effectief genoeg.

Het reduceren van output heeft een kleine prestatiewinst tot gevolg voor zowel Zoning als Zernike. De oorzaak is hiervan is dat het minieme onderscheid tussen ambivalente karakters niet meer gemaakt hoeft te worden. Het verschil tussen de onderlinge klassen wordt zodoende verhoogd. Dat vereenvoudigt de classificatie.

Niet geheel onverwacht worden karakters uit klasse A beter herkend dan karakters uit de klassen B en C. Verrassender is de constatering dat karakters uit klasse C relatief goed herkend worden. Dat er met deze karakters getraind wordt is niet de oorzaak daarvan, daarvoor is de smerigheid te veel willekeurig. Het is waarschijnlijk meer te wijten aan de robuustheid van een feature extractor.

4.3.7 Tussenbeschouwing

De beantwoording van deelvraag 2 heeft de 3e en laatste fase gepasseerd. In fase 1 zijn methoden geselecteerd die effectief features van grayscale beelden kunnen vinden. In fase 2 zijn deze geïmplementeerd en toegespitst op het domein der nummerbordverkenning.

Fase 3 omvatte het testen van bruikbaarheid van deze methoden. Via een selectiestrategie zijn 2 neurale netwerken gevonden die voor respectievelijk topografische pixellabelling en pseudo Zernike momenten de beste prestaties leverden, oftewel de hoogste herkenning kenden. Deze neurale netwerken zijn vervolgens uitvoerig getest.

Met de informatie verkregen in dit hoofdstuk kan vervolgens deelvraag 2 worden beantwoord. Dat gebeurt in de volgende paragraaf. Hieraan kunnen vervolgens conclusies worden verbonden, zowel in de context van karakterherkenning als van nummerbordherkenning.

4.4 Conclusies deelvraag 2

In hoofdstuk 2 werd geconstateerd dat minder leesbare karakters minder makkelijk classificeerbaar zijn. Vanwege het feit dat de invloed van thresholding niet direct meetbaar was, werd derhalve besloten de representatie van karakters na thresholding te onderzoeken. In het onderzoek werden de probleemgevallen geïsoleerd door gebruik te maken van drie verschillende thresholdmethoden.

Deelvraag 1:

Wat is de invloed van thresholding op de representatie van karakters?

Het antwoord op deelvraag 1 is dat de invloed van thresholding op de representatie van karakters is dat 14.4% van alle karakters na thresholding verminderd leesbaar zijn. Vanwege het verband tussen menselijk leesbaarheid en machinale leesbaarheid, in combinatie met het gegeven dat alleen de probleemgevallen zijn gevonden, kon dan worden gesteld dat thresholding van invloed is op de herkenning van karakters.

Deze uitkomst legitimeerde een onderzoek naar mogelijke verbeteringen. Deze verbeteringen konden gevonden worden in gebruikmaking van het menselijk leesbare originele grayscale beeld.

Deelvraag 2:

Wat is het resultaat van het gebruik van grayscale features op karakterherkenning?

De beantwoording van de tweede deelvraag heeft als volgt plaatsgevonden. In de literatuur is eerst gezocht naar geschikte kandidaten om karakteristieken van grijswaardenbeelden te ontrekken. Hieruit is een keuze gemaakt aan de hand van vastgestelde criteria, namelijk navolging in de literatuur, prestaties over het algemeen, toepasbaarheid op het domein en de mate van invariantie omtrent illuminatie. Dat resulteerde in twee geschikt geachte methoden: pseudo Zernike momenten en topografische pixellabelling. Deze twee methoden zijn vervolgens toegespitst op het domein ter hand en geïmplementeerd. Dat heeft geleid tot verschillende alternatieven om features van beelden te ontwaren. De ontwaarde features zijn gebruikt om karakters mee te kunnen classificeren. Hiertoe zijn neurale netwerken aangewend. De mogelijkheden zijn gereduceerd met behulp van een strategie tot er twee succesvolle neurale netwerken overbleven. Eén neuraal netwerk ter classificatie van pseudo Zernike features en één neuraal netwerk ter classificatie van topologische labels middels zoning. Deze twee netwerken zijn op verschillende manieren getest.

De maximale prestaties op grayscale karakters zijn behaald door middel van topografische labelling van pixels van het karakterbeeld. Deze basis is gebruikt om door middel van zoning features te verkrijgen. Door middel van neurale classificatie van deze features werd minimaal, dus met de grootste rejectregion, een herkenning behaald van 75.5% met 4.7% misclassificaties. Afhankelijk van de praktijktoepassing kan deze herkenning worden opgeschroefd ten koste van de error.

Het antwoord op deelvraag 2 kan dan gegeven worden. Het resultaat van het gebruik van grayscale features op de herkenning is dat, met in acht name van een klein errorpercentage, een herkenning kan worden bereikt van 75.5%. Bij acceptatie van hogere errorpercentages en gebruik van minder (ambivalente) outputklassen kan dit worden opgeschroefd tot 87.0%

4.5 Context: nummerbordherkenning

Het nummerbordherkenningsproces, dat is beschreven in hoofdstuk 1, herkent niet alle aangeboden nummerborden. Van de 960 niet herkende nummerborden die beschikbaar zijn gesteld voor onderzoek bleken 556 te voldoen aan de eisen van leesbaarheid en tevredenstellende isolatie van de individuele karakters.

Van alle 6 x 556 = 3336 karakters bleken 14.4% verminderd leesbaar na thresholding. Dit betrof alleen de probleemgevallen. In de context van nummerbordherkenning zal dit percentage minimaal gelijk, maar waarschijnlijk hoger zijn, omdat er per nummerbord 1 thresholdmethode wordt toegepast, en niet 3 thresholdmethoden.

Er is geconcludeerd dat er mogelijk verbeteringen zijn aan te brengen in het herkennen van karakters door het thresholden van beelden niet meer te gebruiken in het herkenningsproces. De prestaties hiervan zijn reeds bekend. De herkenning van de karakters middels binary feature extraction in het systeem van nummerbordherkenning is onbekend. Vandaar dat een vergelijking met dat systeem niet opgaat. Er kan echter een schatting worden gemaakt.

Thresholding met de methode van Niblack en Bernsen met postprocessing van Yanowitz en Bruckstein kent de beste prestaties volgens [4]. De variabelen van de methoden worden ingevuld zoals ze in hoofdstuk 2 gekozen zijn. Van de binaire karakters die zo verkregen worden, worden features gevonden door zoning met zonegroottes 5x6 te gebruiken. Na een normalisatie van de data wordt hiermee een neuraal netwerk getraind. De testresultaten die dit oplevert worden naast die van zoning van pixellabels gezet in tabel 12. In de vergelijking is gestreeft naar vergelijkbare errorpercentages, die in ieder geval onder de 10% liggen. Bij binary treedt dat op bij een rejectregion van (α, β) = (0.7, 0.4) en bij grayscale bij een rejectregion van (α, β) = (0.5, 0.2). Vergelijkingen met verschillende reject regions zijn te vinden in appendix E.

(α, β) = (0.7, 0.4)	Binary			Grayscale		
%	Correct	Rejected	Error	Correct	Rejected	Error
Klasse A	77.9	18.9	3.2	76.8	17.5	5.7
Klasse B	40.6	51.0	8.3	53.1	39.6	7.3
Klasse C	38.1	54.8	7.1	50.0	45.2	4.8

Tabel 12: Verschil tussen prestaties binary en grayscale feature extraction.

Hierdoor valt een vergelijking te maken tussen de twee methoden. Over de gehele linie lijkt er geen winst te halen in herkenning door gebruikmaking van topografische labels. De herkenning van karakters uit de klassen B en C verloopt echter robuuster. Dat blijkt uit grotere correcte herkenningspercentages in combinatie met een verlaging van het errorpercentage. De verwachting kan vervolgens uitgesproken worden, dat door gebruikmaking van topografische labels bij feature extraction er meer nummerborden zullen worden herkend.

Maar dat is niet het geval. Zie tabel 13 voor de resultaten op alle niet herkende nummerborden bij gebruikmaking van binary en grayscale features. Als het leveren van betere prestaties gedefinieerd wordt als het bereiken van een hoger correct herkenningspercentage, dan presteert binary feature extraction beter dan grayscale feature extraction. Als het leveren van betere prestaties gedefinieerd wordt als het reduceren van errorpercentages, dan presteert grayscale beter dan binary feature extraction. De beste prestaties worden geleverd door methoden die winnen in correcte herkenning door het verlagen van de error.

556 nummerborden	Correct	Rejected	Error
Binary Features	106	425	44
Grayscale Features	7	538	11
Conjunctiviteit	6	420	1

Tabel 13: Aantal herkende, verworpen en foutief herkende nummerborden bij gebruikmaking van zowel binary als grayscale feature extraction. Bovendien is de overlap vermeld.

5 Evaluatie

De twee deelvragen zijn beantwoord. Het onderzoek zal hier worden geëvalueerd. Eerst volgt er een korte uiteenzetting van het verloop van het onderzoek. Daarna worden alle relevante aspecten uit het onderzoek nog een keer genoemd, waarna de centrale vraag beantwoord kan worden.

Om te beginnen werd het werkdomein gedefinieerd. Er werd gekozen voor karakters afkomstig van niet herkende nummerborden, waarvan de karakters goed zijn geïsoleerd. Dat resulteerde in 3336 karakters waarmee het onderzoek plaats vond. Vervolgens werd onderzocht wat de invloed van thresholding op de representatie van karakters was. Dit leidde tot een verdeling van karakters over drie klassen: na thresholding goede en verminderd leesbare karakters en karakters waar thresholding werd verhinderd door objecten in het beeld. Thresholding had op 14.4% van de karakters invloed op de leesbaarheid. In combinatie met het verband tussen menselijke en machinale leesbaarheid, werd gesteld dat thresholding van invloed is op de herkenning van karakters. Dat legitimeerde een onderzoek naar het gebruik van het originele grayscale beeld in plaats van het gethresholde beeld bij feature extraction. Daartoe werden eerst kandidaat-methoden, gevonden in vakliteratuur, aan vier criteria getoetst, namelijk, de mate van invariantie omtrent illuminatie, de mate van toepassing op een vergelijkbaar domein, de navolging die een kandidaat krijgt in de vakliteratuur en de resultaten die zijn geboekt. Dit leidde tot een selectie van twee methoden, namelijk, topografische labelling en pseudo Zernike momenten. Deze werden toegespitst op het domein en vervolgens getest middels neurale classificatie. Features verkregen via zoning van topografische labels vertoonde de beste prestaties. Hieruit bleek dat bij inachtname van een laag errorpercentage (4.7%) en herkenning kon worden behaald van 75.5% van alle karakters in het domein. Bij een reductie van outputklassen door middel van samenvoegen van ambivalente karakters, in combinatie met het toelaten van een groter errorpercentage, kon een herkenning worden bereikt van 87.0% van alle karakters in het domein.

5.1 Relevante aspecten

Om de centrale hoofdvraag te beantwoorden, werd gekozen voor de beantwoording van twee deelvragen. Per deelvraag zullen de relevante aspecten op een rijtje worden gezet.

Deelvraag 1 betrof een onderzoek naar de invloed van thresholding op de representatie van karakters. Om tot een bevredigend antwoord te komen zijn vele keuzes gemaakt en verantwoord. Hiervan worden de drie belangrijkste aspecten vermeld.

- Dit onderzoek vond plaats op de invloed op de representatie en niet, zoals bij voorkeur gewenst, op de invloed op de herkenning. Dit had als reden dat, door het gebruik van verschillende thresholdmethoden en binary feature extractors in het nummerbordenherkenningssyteem, geen overzicht kon worden verkregen omtrent de individuele herkenning van karakters. Toch bleek de invloed van representatie in verband te staan met de herkenning, omdat er een verband bestaat tussen menselijke en machinale leesbaarheid. Dit verband wordt vervolgens bevestigd door het feit dat karakters uit klasse B bij binary feature extraction inderdaad een mindere herkenning tentoonspreiden dan de karakters uit klasse A.
- De term *leesbaarheid* werd gedefinieerd. Een binair karakter is niet leesbaar als na thresholding er informatie verschijnt of verdwijnt, op plaatsen in het beeld, waar dat *ogenschijnlijk* niet zou moeten gebeuren.
- Middels een human expert werden alle karakters getoetst op leesbaarheid na thresholding. Hierbij werden 3 thresholdmethoden per karakter gebruikt, om zo de probleemgevallen te isoleren.

Deelvraag 2 onderzocht de resultaten van het toepassen van grayscale features op de herkenning van karakters. Dit verliep in drie fasen. Elke fase vereiste een groot aantal keuzes. Hieruit worden de vier belangrijkste aspecten vermeld.

- Aan de hand van vier criteria werden kandidaat methoden uit de vakliteratuur geselecteerd voor implementatie. De keuzes van de criteria werden gestuurd door zowel het domein als de kandidaat in beschouwing te nemen.
- De karakteristiek topografische basis waar features van werden geëxtraheerd is niets anders dan een tussenstation in de herkenning, net zoals een binair beeld dat is. De winst die hiermee behaald wordt is niets anders dan een grotere verscheidenheid aan pixellabels, in plaats van alleen zwart en wit.
- De wijze waarop illuminatie invariatie wordt bereikt bij pseudo Zernike momenten is een belangrijk aspect bij herkening middels deze momenten.
- Vanwege de vele mogelijkheden die ontstaan bij neurale classificatie van features is gekozen voor een strategie die uiteindelijk leidt tot de twee neurale netwerken die het best presteren, één voor topografische features en één voor pseudo Zernike features.

5.2 Conclusies hoofdvraag

De mogelijkheid van het gebruik van features van grayscale beelden bij feature extraction bleek uit het karakterherkenningsproces zoals beschreven in hoofdstuk 2. Deze mogelijkheid resulteerde in de centrale vraag van dit onderzoek. Er werden twee deelvragen ontworpen die tezamen een bevredigend antwoord zouden geven op deze centrale vraag.

Centrale vraag:

Wat is het gevolg van het gebruik van het originele beeld bij feature extraction op de herkenning van karakters?

In hoofdstuk 2 werd geconstateerd dat minder leesbare karakters minder makkelijk classificeerbaar zijn. Vanwege het feit dat de invloed van thresholding niet direct meetbaar was, werd derhalve besloten de representatie van karakters na thresholding te onderzoeken. In het onderzoek werden de probleemgevallen geïsoleerd door gebruik te maken van drie verschillende thresholdmethoden.

Deelvraag 1:

Wat is de invloed van thresholding op de representatie van karakters?

Het antwoord op deelvraag 1 is dat de invloed van thresholding op de representatie van karakters is dat 14.4% van alle karakters na thresholding verminderd leesbaar zijn. Vanwege het verband tussen menselijk leesbaarheid en machinale leesbaarheid, in combinatie met het gegeven dat alleen de probleemgevallen zijn gevonden, kon dan worden gesteld dat thresholding van invloed is op de herkenning van karakters.

Deze uitkomst legitimeerde een onderzoek naar mogelijke verbeteringen. Deze verbeteringen konden gevonden worden in gebruikmaking van het menselijk leesbare originele grayscale beeld.

> ***Deelvraag 2:***
>
> *Wat is het resultaat van het gebruik van grayscale features op karakterherkenning?*

Het antwoord op deelvraag 2 kan dan gegeven worden. Het resultaat van het gebruik van grayscale features op de herkenning is dat, met in standhouding van een klein errorpercentage, een herkenning kan worden bereikt van 75.5%. Bij acceptatie van hogere errorpercentages en gebruik van minder (ambivalente) outputklassen kan dit worden opgeschroefd tot 87.0%

Door de beantwoording van deelvraag 1 kan gesteld worden dat, naast de mogelijkheid van grayscale feature extraction, het gebruik ervan tevens nut heeft. Na thresholding wordt 14.4 % van de karakters namelijk onleesbaar. De beantwoording van deelvraag 2 laat zien dat daadwerkelijke toepassing van grayscale features leidt tot een degelijke herkenning.

De centrale vraag kan vervolgens worden beantwoord. Het gevolg van toepassing van het originele beeld bij feature extraction op het herkennen van karakters is dat, naast het gebleken nut ervan, het originele beeld zich leent om tot een degelijke karakterherkenning te geraken.

5.3 Context: nummerbordherkenning

De gevolgen van dit onderzoek voor het herkennen van nummerborden kan worden gevat in de volgende twee constateringen

- Grayscale feature extraction heeft nut. Dit blijkt onder andere uit de verhoogde herkenning die dat tot gevolg heeft bij herkenning van dezelfde karakters uit klasse B, met inachtname van een vergelijkbare error.
- Het gebruik van grayscale features leidt niet tot meer herkende nummerborden. Het heeft wel tot gevolg dat het errorpercentage daalt ten koste van herkenning.

De winst van dit onderzoek op de herkenning van nummerborden blijft in het ongewisse, vanwege het feit dat exacte gegevens over de prestaties van de karakterherkenner in dit systeem niet bekend zijn. Een vergelijking met een willekeurig gekozen methode, namelijk zoning na thresholding middels Niblack en Bernsen, laat echter zien dat het gebruik van topografische labelling een robuuste methode is om karakters mee te herkennen.

Referenties

[1] George Nagy
Twenty Years of Document Image Analysis in PAMI
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 22, nr 1, pp 38-62, jan 2000

[2] M.H. ter Brugge; J. Nijhuis; L. Spaanenburg; J.H. Stevens
License-plate Recognition
Persoonlijk archief J.A.G. Nijhuis, 2003

[3] O. Trier, T. Taxt, A. Jain
Feature Extraction methods for character recognition- A survey
Pattern Recognition, vol 29, pp 641-662, 1996

[4] O. Trier, T. Taxt
Evaluation of Binarization Methods for Document Analysis
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 17, nr 3, pp 312-315, mar 1995

[5] O. Trier, A. Jain
Goal Directed evaluation of binarization method
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 17, nr 12, pp 1191-1201, dec 1995

[6] L. Shapiro, G. Stockman
Computer Vision
Prentice Hall, ISBN 0-13-030796-3, 2001

[7] H. Yoshimura, M. Etoh, K. Kondo, N. Yokoya
Gray-scale character recognition by Gabor Jets projection
Proceedings of the 15th International Conference on Pattern Recognition, vol 2, pp 335-338, 2000

[8] S. Paschalakis, p. Lee
Pattern Recognition in grey level images usding moment based invariant features
Image Processing and its Applications, Conference Publication nr. 465, pp 245-249, 1999

[9] R.J. Prokop, A.P. Reeves
A Survey of Moment-Based techniques for unoccluded object representation and Recognition.
CVGIP - Graphical Models and Image Processing, vol 54, nr 5, pp 438-460, 1992

[10] R. Bailey, M. Srinath
Orthogonal Moment Features for use with Parametruc and Non-Parametric Classifiers
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 18, nr 4, pp 389-399, apr 1996

[11] Sudobh Kumar
Surface Triangulation: A Survey
ACM Transactions on Graphics, vol 10, nr 2, pp 182-199, apr 1991

[12] M. Garland & P.S. Heckbert
Fast Polygonal Approximation of Terrains and Height Fields
Publication of Carnegie Mellon University of Pittsburgh, CMU-CS-95-181, September 19, 1995

[13] H. Nishida
Boundery Feature extraction from gray-scale document images
4th ICDAR 18-20 Augustus 1997, pp 132-136

[14] L. Wang, T. Pavlides
Direct Gray-Scale Extraction of Features for Character Recognition
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 15, nr 10, pp1053-1067, oct 1993

[15] S.W. Lee, Y.J. Kim
Direct Extraction of Topographic Features for gray Scale Character Recognition
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 17, nr 7, pp 724-729, july 1995

[16] O. Trier, T. Taxt, A. Jain
Recognition of Digits in Hydrographic Maps: Binary Versus Topographic Analysis
IEEE Transactions on Pattern Analysis and Machine Intelligence. vol 19 nr 4, pp 399-404, april 1997

[17] S. Haykin
Neural Networks: A comprehensive foundation
Prentice Hall, ISBN 0-13-895863-7, 1994

[18] Jos Nijhuis
An engineering approach to Neural System Design
Proefschrift nijmegen. ISBN 90-9005499-5, 1992

[19] Documentation for the Neural Network Toolbox for Matlab
Training Algorithms: *Speed and Memory Comparison*
http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/backpr13.shtml#3826

Appendices

Appendix A: Resultaten van classificatie naar threshold

```
Number of plates             :  556
Hence number of characters   : 3336

    Klasse A                 : 2660
    Klasse B                 :  479
    Klasse C                 :  197
```

Kwantitieve resultaten per karakter weergegeven voor elke klasse.

	kA	kB	kC	tot
A	1	0	0	1
B	96	60	3	159
C	4	0	0	4
D	130	7	5	142
E	0	0	0	0
F	118	21	17	156
G	101	16	8	125
H	100	14	2	116
I	0	0	0	0
J	86	7	13	106
K	30	6	13	49
L	84	9	17	110
M	0	0	0	0
N	125	9	0	134
O	1	0	0	1
P	115	28	9	152
Q	0	0	0	0
R	97	24	6	127
S	121	16	4	141
T	128	21	17	166
U	0	0	0	0
V	176	5	9	190
W	0	0	0	0
X	123	13	3	139
Y	44	3	4	51
Z	116	20	5	141

	kA	kB	kC	tot
0	79	5	5	89
1	86	17	19	122
2	137	29	8	174
3	95	10	2	107
4	92	10	1	103
5	78	21	5	104
6	71	22	3	96
7	76	9	14	99
8	83	48	2	133
9	67	29	3	99

Resultaten in verhouding met het totaal per karakter per klasse.

	kA	kB	kC		kA	kB	kC
A	1.00	0.00	0.00	0	0.89	0.06	0.06
B	0.60	0.38	0.02	1	0.70	0.14	0.16
C	1.00	0.00	0.00	2	0.79	0.17	0.05
D	0.92	0.05	0.04	3	0.89	0.09	0.02
E	0.00	0.00	0.00	4	0.89	0.10	0.01
F	0.76	0.13	0.11	5	0.75	0.20	0.05
G	0.81	0.13	0.06	6	0.74	0.23	0.03
H	0.86	0.12	0.02	7	0.77	0.09	0.14
I	0.00	0.00	0.00	8	0.62	0.36	0.02
J	0.81	0.07	0.12	9	0.68	0.29	0.03
K	0.61	0.12	0.27				
L	0.76	0.08	0.15				
M	0.00	0.00	0.00				
N	0.93	0.07	0.00				
O	1.00	0.00	0.00				
P	0.76	0.18	0.06				
Q	0.00	0.00	0.00				
R	0.76	0.19	0.05				
S	0.86	0.11	0.03				
T	0.77	0.13	0.10				
U	0.00	0.00	0.00				
V	0.93	0.03	0.05				
W	0.00	0.00	0.00				
X	0.88	0.09	0.02				
Y	0.86	0.06	0.08				
Z	0.82	0.14	0.04				

Verdeling van klassen onder nummerborden.

De cellen waar A+B geen 6 is staat de sequence i.c.m C	Klasse B						
	0	1	2	3	4	5	6
Klasse A 6	161						
5	75	136					
4	30	31	56				
3	3	2	12	26			
2			2	2	16		
1						2	
0							2

Appendix B: Nummerborden na thresholding en topolabelling

Sommige originele nummerborden waren zo donker dat om visuele redenen deze iets lichter zijn gemaakt. Boven de karakters in de threshold kolom staat in welke klasse ze zijn ingedeeld. De labelling voor de voorgrond is gedaan met een threshold van *2%*, zonder filtering en met postprocessing grootte 2. De labelling voor de achtergrond is gedaan zonder threshold na filteren en met postprocessing grootte 2.

Appendix C: Code voor pseudo Zernike momenten

```
/*****************************************************************************/
/**    Radial polynomial  as describe in page 391                           **/
/*****************************************************************************/

double Pseudo_Radial_Poly(int n, int l,double ruo)
{
  int k, i           ;
  double temp   = 1.0 ;
  double result = 0.0 ;

  for(k = l ; k <= n ; k++)
    {
      temp = 1.0;
      for(i  = l + k + 2;
          i <= n + k + 1;
          i++) temp *= i;

      temp /= NumberFactorial( n - k );
      temp /= NumberFactorial( k - l );
      temp *= pow( ruo , k );

      if( ( n-k ) % 2 == 1 )
        temp=-temp;
      result+=temp;
    }
  return result;
}



/*****************************************************************************/
/**        V is polynomial defined as equ(3) in paper                       **/
/*****************************************************************************/

void W_poly(int n,int m,double ruo,double sita,double* realpart ,double* impart)
{
   double temp;

   temp      = Pseudo_Radial_Poly( n, m, ruo );
   *realpart = temp * cos( m * sita );
   *impart   = temp * sin( m * sita );
}
```

```
/*****************************************************************************/
/**       description in  PAMI vol 18.No.4                                  **/
/*****************************************************************************/

double PseudoZernikeMoment (Matrix *in, int p, int q)
{

  double cx, cy          ;  // het geometrische centrum van de input
  double longest         ;  // de afstand van het centrum tot een hoek.
  double ruo             ;  // afstand van een pixel tot het centrum
  double sita            ;  // hoek tussen y-as en (centrum-pixel)lijn
  double moment          ;  // wortel uit de kwadraten van het reeel en
                            // imaginaire deel van het moment
  double realmoment = 0.0;
  double immoment   = 0.0;
  double realpart   = 0.0;
  double impart     = 0.0;
  double sum        = 0.0;
  int    x, y            ;

  cx      = in->cols / 2.0;
  cy      = in->rows / 2.0;
  longest = (1.0/2.0) * sqrt ((double)((in->cols * in->cols)+(in->rows * in->rows) ));
  //  longest = sqrt ( (lx-cx)*(lx-cx) + (ly-cy)*(ly-cy) );

  for(y=0; y<in->rows; y++)
    for(x=0; x<in->cols; x++)
      {
       Convert_X_Y_TO_Polar(x, y, cx, cy, &ruo, &sita);
       ruo = ruo / longest;

       W_poly(p, q, ruo, -sita, &realmoment ,&immoment);

       ////////////////////////////////////////////
       ////  Added a multiplication ///////////////
       ////////////////////////////////////////////
       realpart += ((double)in->buf[y][x]) * realmoment;
       impart   += ((double)in->buf[y][x]) * immoment;

      }

  realpart = realpart * ( p + 1 ) / PI;
  impart   = impart   * ( p + 1 ) / PI;
  moment   = sqrt ( (realpart * realpart ) + ( impart * impart ) );

  for(y=0; y<in->rows; y++)
    for(x=0; x<in->cols; x++)
      sum += in->buf[y][x];

  if(abs(sum) < 1e-6)
    return 0;
  else
    return (moment /= sum);
}
```

Appendix D: Testresultaten winnende neurale netwerken

Zoning met volledige output

α=0.1, β=0.0	Correct	Rejected	Error
A	85.0	1.3	13.7
B	71.9	4.2	24.0
C	66.7	4.8	28.6
AB	83.1	1.7	15.2
AC	83.7	1.5	14.8
BC	70.3	4.3	25.4
ABC	82.1	1.9	16.0

α=0.3, β=0.1	Correct	Rejected	Error
A	82.9	8.8	8.4
B	63.5	22.9	13.5
C	54.8	35.7	9.5
AB	80.0	10.8	9.1
AC	80.9	10.6	8.5
BC	60.9	26.8	12.3
ABC	78.5	12.3	9.2

α=0.5, β=0.2	Correct	Rejected	Error
A	76.8	17.5	5.7
B	53.1	39.6	7.3
C	50.0	45.2	4.8
AB	73.3	20.7	5.9
AC	74.9	19.4	5.6
BC	52.2	41.3	6.5
ABC	71.9	22.2	5.9

Pseudo Zernike winnaar met volledige output

α=0.1, β=0.0	Correct	Rejected	Error
A	68.2	4.6	27.1
B	41.7	9.4	48.9
C	42.9	11.9	45.2
AB	64.3	5.3	30.3
AC	66.4	5.1	28.4
BC	42.0	10.1	47.8
ABC	63.0	5.7	31.2

α=0.3, β=0.1	Correct	Rejected	Error
A	61.4	24.1	14.5
B	34.4	44.8	20.8
C	28.6	52.4	19.0
AB	57.5	27.1	15.4
AC	59.1	26.1	14.8
BC	32.6	47.1	20.3
ABC	55.7	28.7	15.6

α=0.5, β=0.2	Correct	Rejected	Error
A	53.9	36.4	9.6
B	24.0	63.5	12.5
C	26.2	61.9	11.9
AB	49.5	40.4	10.0
AC	52.0	38.2	9.8
BC	24.6	63.0	12.3
ABC	48.1	41.7	10.2

Zoning met gereduceerd output

α=0.1, β=0.0	Correct	Rejected	Error
A	90.4	0.5	9.1
B	74.0	3.1	22.9
C	71.4	4.8	23.8
AB	88.0	0.9	11.1
AC	89.0	0.8	10.1
BC	73.2	3.6	23.1
ABC	87.0	1.4	11.9

α=0.3, β=0.1	Correct	Rejected	Error
A	86.3	8.0	5.7
B	62.5	25.0	12.5
C	61.9	26.2	11.9
AB	82.8	10.5	6.7
AC	84.6	9.3	6.1
BC	62.3	25.4	12.3
ABC	81.5	11.5	7.0

α=0.5, β=0.2	Correct	Rejected	Error
A	81.0	14.6	4.3
B	56.3	37.5	6.3
C	45.2	47.6	7.1
AB	77.4	18.0	4.6
AC	78.6	16.9	4.5
BC	52.9	40.6	6.5
ABC	75.5	19.8	4.7

Pseudo Zernike winnaar met gereduceerd output

α=0.1, β=0.0	Correct	Rejected	Error
A	74.6	2.0	24.5
B	43.8	4.2	52.1
C	35.7	2.4	61.9
AB	69.2	2.3	28.5
AC	70.9	2.0	27.1
BC	41.3	3.6	55.1
ABC	67.2	2.3	30.5

α=0.3, β=0.1	Correct	Rejected	Error
A	66.6	21.3	12.1
B	32.3	41.7	26.0
C	23.8	52.4	23.8
AB	61.6	24.2	14.2
AC	63.6	23.4	13.0
BC	29.7	44.9	25.4
ABC	59.3	25.9	14.8

α=0.5, β=0.2	Correct	Rejected	Error
A	56.6	36.3	7.1
B	25.0	59.4	15.6
C	14.3	76.2	9.5
AB	52.0	39.6	8.4
AC	53.7	39.0	7.3
BC	21.7	64.5	13.8
ABC	49.7	41.8	8.5

Appendix E: Vergelijking binary en grayscale features

De rejectregions voor de resultaten na 5x6 zoning van binaire beelden staan in de tabellen aangegeven. Voor de resultaten na zoning 5x6 van voor en achtergrond informatie van topografische labels is gekozen voor een rejectregion van $(\alpha, \beta) = (0.5, 0.2)$.

$(\alpha, \beta) = (0.5, 0.2)$	Binary			Grayscale		
%	Correct	Rejected	Error	Correct	Rejected	Error
Klasse A	*82.1*	*13.6*	*4.3*	*76.8*	*17.5*	*5.7*
Klasse B	*52.1*	*35.4*	*12.5*	*53.1*	*39.6*	*7.3*
Klasse C	*40.5*	*42.9*	*16.7*	*50.0*	*45.2*	*4.8*

$(\alpha, \beta) = (0.6, 0.3)$	Binary			Grayscale		
%	Correct	Rejected	Error	Correct	Rejected	Error
Klasse A	*80.9*	*15.4*	*3.8*	*76.8*	*17.5*	*5.7*
Klasse B	*49.0*	*42.7*	*8.3*	*53.1*	*39.6*	*7.3*
Klasse C	*38.1*	*50.0*	*11.9*	*50.0*	*45.2*	*4.8*

$(\alpha, \beta) = (0.7, 0.4)$	Binary			Grayscale		
%	Correct	Rejected	Error	Correct	Rejected	Error
Klasse A	*77.9*	*18.9*	*3.2*	*76.8*	*17.5*	*5.7*
Klasse B	*40.6*	*51.0*	*8.3*	*53.1*	*39.6*	*7.3*
Klasse C	*38.1*	*54.8*	*7.1*	*50.0*	*45.2*	*4.8*