

# Menselijke perceptie Versus Computerperceptie

Hiram van Paassen

Begeleider: Stefan van der Stigchel

Gewicht: 7,5 ECTS

24 januari 2011

# Inhoudsopgave

<b>1</b>	<b>Inleiding</b>	<b>3</b>
1.1	Perceptie . . . . .	3
1.2	Visuele perceptie . . . . .	3
1.3	Kunstmatige Intelligentie . . . . .	4
<b>2</b>	<b>Hoofdvraag</b>	<b>4</b>
<b>3</b>	<b>Kader</b>	<b>4</b>
3.1	Vergelijkingscriteria . . . . .	5
3.1.1	Prestatie . . . . .	5
3.1.2	Receptoren . . . . .	5
3.1.3	Structuur van informatie . . . . .	6
3.1.4	Symbolisch versus sub-symbolisch . . . . .	6
3.1.5	Datadichtheid . . . . .	7
3.2	Turing-compatibiliteit . . . . .	7
3.3	Marr's drie perspectieven . . . . .	7
<b>4</b>	<b>Theoretische verschillen</b>	<b>8</b>
4.1	Implementatie . . . . .	8
4.2	Algoritmisch . . . . .	8
4.3	Computationeel . . . . .	9
<b>5</b>	<b>Concrete verschillen</b>	<b>9</b>
5.1	Bewegingscontrast . . . . .	10
5.2	Resolutie . . . . .	10
5.3	Gezichtsherkenning . . . . .	11
<b>6</b>	<b>3DFACE, een gezichtsherkeningsalgoritme</b>	<b>13</b>
6.1	3DFACE in vogelvlucht . . . . .	14
6.2	Informatieverwerving . . . . .	15
6.3	Voorverwerking . . . . .	15
6.4	Herkenning . . . . .	16
<b>7</b>	<b>Conclusie</b>	<b>16</b>
<b>8</b>	<b>Discussie</b>	<b>16</b>
8.1	Toekomstig onderzoek . . . . .	17
8.2	Vergelijkingscriteria . . . . .	17
8.3	Alternative modaliteiten . . . . .	17
	<b>Referenties</b>	<b>18</b>

# 1 Inleiding

Het maken van een computer die de wereld kan waarnemen, zoals mensen de wereld waarnemen, is een droom van veel Kunstmatige-Intelligentiestudenten. Een computer die dat kan, is namelijk in staat om mensen vervelende, moeilijke of gevaarlijke taken uit handen nemen. Dit kan echter alleen als een computer genoeg informatie heeft om de taak te kunnen volbrengen. Om de benodigde informatie te verzamelen moeten computers dus hun omgeving kunnen waarnemen.

Hoewel er enige successen zijn op dit gebied, is het de vraag of die goed genoeg zijn om een toekomstige intelligente computer te voorzien in zijn informatiebehoefte. Er zijn heel veel taken waarbij perceptie van belang is en waar mensen geen moeite mee hebben, bijvoorbeeld het herkennen van personen en objecten. Computers presteren vaak slecht op dat soort taken. Naast dit verschil in prestatie is er ook een verschil in implementatie, aangezien computers werken op basis van transistoren en de menselijke hersenen op basis van neuronen.

De oorzaak van het prestatieverschil kan zitten in dit verschil in implementatie. De hardware van computers maakt het wellicht niet mogelijk om efficiënte perceptie te implementeren. Een andere mogelijkheid voor het prestatieverschil is dat de verschillen in aanpak en algoritmes —de manier waarop de hardware gebruikt wordt— ervoor zorgen dat de prestatie onderdoet voor mensen.

## 1.1 Perceptie

In de studie van Kunstmatige intelligentie (KI) is perceptie vaak het ondergeschoven kindje. De algemene gedachte is dat intelligentie voortkomt uit een ingewikkeld proces dat zich in de hersenen afspeelt en perceptie "maar" domme voorverwerking is. In het model van invoer  $\rightarrow$  verwerking  $\rightarrow$  actie is verwerking de plek waar intelligentie zijn oorsprong vindt. Deze gedachte maakt perceptie en actie tot een minder interessant fenomeen. De geschiedenis van KI-onderzoek laat echter ook zien dat veel onderzoek naar computer-"intelligentie" stuk loopt op de complexiteit van de echte wereld [Brooks, 1991]. Intelligente systemen falen zodra ze uit de "perfecte" wereld gehaald worden waarin ze zijn ontstaan. Een complex perceptiesysteem is cruciaal voor intelligentie in een complexe wereld als de onze. Zonder goede computerperceptie is een nuttige intelligente computer —een computer die efficiënt taken van mensen kan overnemen— nog ver weg.

## 1.2 Visuele perceptie

Visuele perceptie is in vele opzichten interessanter dan andere vormen van perceptie. Ten eerste omdat het de dominante vorm van waarneming voor het merendeel van de mensheid is. Veel taken die men een intelligent systeem zou willen laten doen zijn gedeeltelijk of helemaal gebaseerd op visuele perceptie. Ten tweede heeft het visuele systeem de directste link met de omgeving en aangezien die omgeving voor computer en mensen vergelijkbaar moet zijn, lijkt het aannemelijk dat enige overeenkomsten in het perceptuele systeem het sterkst zijn. Daarnaast is het visuele systeem binnen zowel psychologie als informatica een veel gebruikt onderwerp en is er relatief veel onderzoek beschikbaar. Deze

scriptie zal zich om deze redenen richten op visuele perceptie, hoewel er voor andere modaliteiten wellicht ook een interessante vergelijking mogelijk is.

### 1.3 Kunstmatige Intelligentie

Dit onderwerp is een typisch voorbeeld van een echt KI-onderwerp. Niet alleen omdat het interdisciplinair is, —het leent van psychologie, filosofie en informatica— maar ook omdat deze interdisciplinariteit precies is wat KI een waardevol vakgebied maakt. Te veel KI-onderzoek wordt gekenmerkt door een gebrek aan interdisciplinariteit. Meerdere disciplines gebruiken is namelijk niet per se interdisciplinair; dat is multidisciplinair. Interdisciplinariteit ontstaat uit de individuele disciplines en geeft toegevoegde waarde ten opzichte van de oorspronkelijke disciplines. Het onderwerp van deze scriptie is voor de afzonderlijke disciplines niet per se interessant of vernieuwend. De waarde zit in het feit dat interdisciplinariteit in staat stelt om vanuit een nieuw perspectief naar bestaande zaken te kijken en zo tot nieuwe inzichten te komen. Psychologie is vooral een beschrijvende wetenschap. Psychologen proberen te beschrijven hoe mensen werken. Wát mensen doen is vaak duidelijk en van ondergeschikt belang. Informatica is een hele andere tak van sport. Voor informatici is de precieze beschrijving van de werking van het onderwerp vaak oninteressant. Wát het doet, is daarentegen juist heel interessant. Kunstmatige intelligentie combineert beide perspectieven om tot een beter begrip van intelligentie te komen. Een mogelijkheid tot het vergelijken van de onderwerpen van beide vakgebieden, mensen en computers, is hierbij cruciaal. Het creëren van deze mogelijkheid is waar deze scriptie aan hoopt bij te dragen.

## 2 Hoofdvraag

Op dit moment zijn computers nog niet in staat de juiste informatie te verkrijgen uit hun omgeving om taken waar mensen geen moeite mee hebben, even goed als mensen te volbrengen.

Is het zinlijk om mensen met computers te vergelijken en welke lessen kunnen we daar uit leren?

Om deze vraag te kunnen beantwoorden moet een vergelijking worden gemaakt tussen twee wellicht totaal verschillende systemen. Het is zinlijk om eerst te bepalen wat er nodig is om een goede vergelijking te maken. Hiervoor zal in het eerste hoofdstuk een kader geschetst worden. Daarna is het noodzakelijk te beschrijven welke verschillen er inherent bestaan tussen menselijke perceptie en computerperceptie. Deze verschillen zullen aan de hand van dit kader beschreven worden in hoofdstuk 4. Als deze verschillen niet onoverkomelijk blijken, dan kan men kijken naar de verschillen die er in de praktijk zijn en wat daar de oorzaak van is. Dit kan dan licht werpen op de mogelijke lessen die we kunnen toepassen om computerperceptie beter te maken. Over deze concrete verschillen tussen de soorten perceptie zullen de hoofdstukken 5 en 6 gaan.

## 3 Kader

Het beschrijven van twee totaal verschillende onderwerpen in twee totaal verschillende vakgebieden is niet gemakkelijk. Doordat informatica en psychologie

zulke verschillende manieren van werken en beschrijven hebben, is het lastig om interdisciplinair te werken. Eigenschappen die binnen het ene vakgebied relevant worden geacht, kunnen in het andere vakgebied overbodig zijn. Het is dus de uitdaging om "common ground" te vinden van waaruit interdisciplinair onderzoek gestart kan worden.

Gelukkig zijn er binnen de verschillende disciplines wel degelijk resultaten die breder inzetbaar zijn en gebruikt kunnen worden als basis voor een vergelijking tussen een psychologieonderwerp en een informaticaonderwerp. In dit hoofdstuk zullen enkele van die resultaten aan bod komen om het mogelijk te maken menselijke perceptie met computerperceptie te vergelijken.

In het boek *Vision* van David Marr [Marr, 1982] probeert Marr op een interdisciplinaire manier een verklaring te geven hoe visuele perceptie werkt. Hiervoor poseert hij de drie-perspectieven-hypothese. Dit schetst een kader vanuit welk perspectief men een informatieverwerkend systeem kan beschrijven. Dit is van belang, omdat de verschillende perspectieven niet met elkaar te vergelijken zijn. Het negeren van dit verschil in perspectief bij een beschrijving lijdt dan dus tot verwarring en het vergelijken van appels met peren. Sectie 3.3 zal hierover verder gaan.

Daarnaast is het nodig om te kunnen constateren of beide systemen vergelijkbare limieten hebben. De gedachte dat computers een vergelijkbare prestatie kunnen leveren als mensen komt voor een deel voort uit de aanname dat computers en hersenen theoretisch gezien dezelfde functies kunnen vervullen. Als menselijke perceptie en computerperceptie niet equivalent zijn in hun theoretische mogelijkheden is een verdere vergelijking minder interessant, omdat dan al van tevoren vaststaat dat één van de twee systemen minder krachtig is. Deze zogenaamde Turing-compatibiliteit zal verder besproken worden in sectie 3.2

## 3.1 Vergelijkingscriteria

Omdat er echter meer meespeelt dan theoretische grenzen is het nodig om ook een indicatie te geven van de waarden waarop men een vergelijking wil uitvoeren. In deze sectie zal een aantal eigenschappen van informatieverwerkende systemen en specifiek perceptie aan bod komen. Deze onderwerpen zijn bedoeld om een aantal relevante eigenschappen toe te lichten en bloot te leggen. Deze lijst is echter niet uitputtend.

### 3.1.1 Prestatie

Een belangrijke maat om perceptie te vergelijken is prestatie. Dit is immers voor praktische toepassingen de meest relevante maat. Zolang een perceptief systeem niet net zo goed of beter presteert als menselijke perceptie, is het economisch gezien niet handig om een mens te vervangen. Beter presteren definiëren we hier als: een perceptiesysteem bezitten die de mogelijkheid geeft om alle taken die mensen goed kunnen uitvoeren net zo goed of beter te kunnen doen.

### 3.1.2 Receptoren

Computers zijn beter in het registreren van absolute waarden en het streven is vaak dat ze geen of een minimale bias voor verschillende stimuli hebben. Dit maakt ze zeer geschikt om te meten en om een geheel spectrum of blikveld

uniform te registreren. Zo heeft een digitale camera een homogene resolutie voor zijn gehele blikveld. Microfoons zijn er op gebouwd om elke frequentie zo goed mogelijk te registreren. De menselijke zintuigen zijn daarentegen veel minder homogeen. De resolutie in de fovea, het midden van het netvlies, is vele malen hoger dan die in periferie. Het nadeel is dat mensen slecht zijn in het absoluut waarnemen van stimuli. Het voordeel is echter dat deze bias ervoor zorgt dat er minder informatie verwerkt hoeft te worden. De resolutie is het hoogst waar dat nodig is en delen van het spectrum die minder relevant zijn, zijn minder prominent in de data terug te vinden. Dit scheelt een hoop voorverwerking die bij computers wel nodig is.

### 3.1.3 Structuur van informatie

Een probleem dat perceptie oplost is de transformatie van de structuur van informatie naar een structuur die bruikbaar is voor verwerking door "Intelligentie". Om een voorbeeld te geven. Een foto is representatie van een scene met een tweedimensionale structuur waarbij elke informatie eenheid, een pixel, spatieel gerelateerd is aan de anderen. De weinig mogelijke algemene classificaties beperken zich tot classificaties zoals "Alle blauwe pixels". Een andere representatie voor dezelfde scene zou een representatie kunnen zijn waarbij de eenheden van informatie bestaan uit de objecten die aanwezig zijn met enige semantische labeling zoals bijvoorbeeld de soort. Deze objecten kunnen nog steeds spatieel aan elkaar gerelateerd zijn, maar er zijn meer relaties mogelijk dan met pixels. Bijvoorbeeld *voor(x, y)* of *achter(x, y)* relaties. Ook zijn er meer classificaties mogelijk zoals "Alle stoelen".

### 3.1.4 Symbolisch versus sub-symbolisch

Een grote verdeling binnen de KI is de verdeling tussen symbolische KI en sub-symbolische KI [Honavar, 1994]. De symbolische KI houdt zich bezig met logische regels en ondeelbare symbolen die gemanipuleerd worden. De sub-symbolische KI houdt zich juist bezig met statistische technieken en neurale netwerken. Deze verdeling betekent dat er een continue discussie is over de manier waarop intelligentie geïmplementeerd moet worden. De grote vraag voor perceptie is of stimuli als gehele elementen verwerkt worden (holistisch), of dat ze opgedeeld worden in elementen die een expliciete betekenis en relatie tot elkaar hebben (features).

De initiële verwerking binnen het visuele systeem is sub-symbolisch, verwerking gebeurt in eerste instantie op basis van informatie-eenheden die spatieel tot elkaar gerelateerd zijn. Pas later in de verwerking zijn er tekenen van symbolische verwerking waarbij informatie-eenheden uit de stimulus gegroepeerd en afgebakend worden. Dit levert nieuwe informatie-eenheden op die als zodanig een semantische waarde krijgen [Quiroga et al., 2005].

De koppeling tussen symbolische en sub-symbolische systemen is in computerperceptie een lastig probleem, veel van de bestaande methodes zijn niet robuust en vergen veel onderhoud. De voordelen zijn echter relatief groot. Lastige invoer is door middel van een sub-symbolisch systeem makkelijker te verwerken terwijl er in de verwerking van de verkregen gegevens gebruik gemaakt kan worden van symbolische systemen die in staat zijn tot redeneren met de verkregen informatie [Markus, 2009, McGarry et al., 1999].

### 3.1.5 Datadichtheid

Statistische of sub-symbolische methoden voor het verwerken van informatie gedijen bij een overvloed aan informatie. Symbolische methoden zijn vaak juist gebaad bij een selectieve hoeveelheid informatie. De hoeveelheid informatie of, beter gezegd, de dichtheid van de informatie is van invloed op de efficiëntie van informatieverwerkende systemen. Bij mensen lijkt het erop dat ze goed gedijen bij een overvloed aan licht gestructureerde natuurlijke informatie [Bach-y-Rita and Kerckel, 2003]. Bij computers hangt het vooral van het gebruikte algoritme af of een hoge datadichtheid gewenst is. Zodra de looptijd van een algoritme exponentieel afhangt van de lengte van de invoer, wordt het onwenselijk om een grote invoer te gebruiken vanwege de lange looptijd.

## 3.2 Turing-compatibiliteit

Een universele Turingmachine is een model voor een computer die alle mogelijke algoritmen kan uitvoeren. Alle moderne computers zijn Turing-compatibel, wat betekent dat ze dezelfde problemen kunnen oplossen als een universele Turingmachine. Er zijn echter problemen die ook een Turingmachine niet kan oplossen, zoals het halting-probleem. Dit probleem houdt in dat er geen Turingmachine kan bestaan die voor elk mogelijk programma kan bepalen of het ooit stopt met uitvoeren of dat het oneindig doorloopt.

Het is een open vraag of er systemen mogelijk zijn die dergelijke problemen wel kunnen oplossen. Als een dergelijk systeem in de menselijke hersenen aanwezig zou zijn, dan zou dat betekenen dat mensen fundamenteel meer kunnen dan computers. Ze zouden dan immers problemen kunnen oplossen die computers niet kunnen oplossen. Het heeft er echter alle schijn van dat een dergelijke "hyper-computer" alleen onder zeer speciale onaanvaardbare omstandigheden voor kan komen. Dit zou betekenen dat computers dezelfde mogelijkheden moeten hebben als mensen in termen van wat ze kunnen berekenen. Hier volgt weer uit dat in principe, de menselijke hersenen in een computer geëmuleerd kunnen worden of dat een computer dezelfde problemen kan oplossen als een mens. Een mens kan dus fundamenteel niet meer dan een computer

## 3.3 Marr's drie perspectieven

Om een beeld te kunnen vormen van het verschil tussen menselijke perceptie en computerperceptie, is het wenselijk een kader te schetsen waarbinnen informatieverwerkende systemen vergeleken kunnen worden. David Marr geeft in *Vision* een dergelijk kader. Deze zogenaamde drie-level-hypothese maakt een onderscheid tussen het computationele perspectief, het algoritmische perspectief en het implementatieperspectief. Deze drie perspectieven hangen nauw met elkaar samen. Echter, om de verschillen tussen menselijke perceptie en computerperceptie duidelijk te beschrijven, is het noodzakelijk om onderscheid te maken.

**Computationeel** Wat is het doel van het systeem en onder welke omstandigheden is het systeem relevant.

**Algoritmisch** Wat is de representatie van invoer en uitvoer? Wat is het algoritme dat de transformatie doet?

**Implementatie** Hoe wordt de representatie en het algoritme fysiek geïmplementeerd?

Het voorbeeld dat Marr ter verduidelijking van zijn theorie gebruikt, is een beschrijving op drie niveaus van een kassa. Het computationele niveau beschrijft het wat en waarom: Een kassa telt bedragen op; hiervoor is kennis van optellen voor nodig. Op het algoritmische niveau gaat het over de soort transformatie van invoer naar uitvoer en welke symbolen gebruikt worden. Bijvoorbeeld, gebruik je Arabische cijfers of wordt er geturfd? Het implementatieniveau beschrijft vervolgens welke fysieke implementatie hieraan ten grondslag ligt: is de kassa elektronisch of mechanisch?

## 4 Theoretische verschillen

In het vorige hoofdstuk werd een kader geschepd om een vergelijking tussen menselijke perceptie en computerperceptie mogelijk te maken. Dit hoofdstuk zal een theoretische vergelijking maken om te kijken op welke niveaus er overeenkomsten zijn waarop een concrete vergelijking uitgevoerd kan worden. Om dit te kunnen doen, is het nodig om een scheiding te maken tussen de perspectieven en om vanuit de verschillende perspectieven te kijken wat de verschillen en overeenkomsten zijn.

### 4.1 Implementatie

Het grootste verschil tussen mens en computer is vanzelfsprekend de implementatie. Waar mensen hun rekenkracht putten uit neuronen in specifieke configuraties, zijn computers aangewezen op elektronische componenten met als basis de transistor. Hoewel dit een significant verschil is, is het voor de werking van het systeem niet per se van belang. In het verleden zijn er computers op basis van radiobuizen en mechanica zijn geweest, deze waren een aantal ordes van grootte langzamer. Toch deden ze niet onder in mogelijkheden, omdat ze allemaal Turing-compatibel waren. Omdat we ervan uit kunnen gaan dat de menselijke hersenen ook Turing-compatibel zijn, zijn verschillen qua implementatie niet onoverbrugbaar. Dit zorgt er wel voor dat een vergelijking op implementatieniveau, alleen verschillen zal opleveren. Toch zijn deze verschillen niet geheel irrelevant aangezien ze voor veranderingen zorgen waar op het algoritmische niveau rekening mee gehouden moet worden. Zo is het mogelijk dat de looptijd van algoritmes sterk verandert op basis van de implementatie. Veel taken vereisen een snelle reactietijd, dus moet perceptie ook snel plaatsvinden en zijn korte looptijden nodig. Hierdoor wordt het belangrijk om het juiste algoritme te kiezen passend bij een implementatie, ondanks dat alternatieve algoritmes wellicht uiteindelijk hetzelfde antwoord geven.

### 4.2 Algoritmisch

Door het verschil in implementatie is er op het algoritmische niveau ook een aantal belangrijke verschillen. Doordat computers al hun rekenwerk doen door middel van één of enkele processors, is het nodig dat als er meer taken zijn dan processors, deze om de beurt worden uitgevoerd. Menselijke neuronen werken onafhankelijk van elkaar en kunnen daardoor meerdere taken tegelijk verwerken.



Dit uit zich niet alleen in de mogelijkheid dat hersenen tegelijk met bijvoorbeeld perceptie en planning bezig kunnen zijn, het stelt de hersenen ook beter in staat om temporele informatie te verwerken. De hersenen lijken in hun verwerking van informatie wat dat betreft op een lopende band in een fabriek. Terwijl aan het eind van de band het product wordt afgemaakt, zijn er tegelijk aan het begin van de band mensen bezig die de eerste handelingen voor het volgende product uitvoeren. Een computer lijkt in dat opzicht veel meer op een fabriek waar de werknemers per persoon geheel verantwoordelijk zijn voor één product.

### 4.3 Computationeel

Het computationele perspectief vertoont daarentegen juist weer grote overeenkomsten tussen de twee typen perceptie. Een computationele beschrijving voor een perceptueel systeem is: een systeem dat stimuli uit de echte wereld verwerkt tot informatie waarmee een complex systeem taken kan oplossen. Deze definitie gaat op voor elk perceptiesysteem, ongeacht de algoritmes en implementatie. Om deze definitie meer inhoud te geven, is het nodig om ervan uit te gaan dat de informatie die een computer en een mens nodig hebben nagenoeg gelijk zijn. Immers, een systeem dat alleen dag en nacht hoeft te herkennen, zal met eenvoudige visuele perceptie af kunnen. Omdat de wens is dat een intelligente computer taken van mensen kan overnemen, is het waarschijnlijk dat ze overeenkomstige informatie nodig hebben en dat de onderdelen van hun perceptiesysteem computationeel op elkaar lijken.

Omdat computationeel gezien de verschillen tussen menselijke perceptie en computerperceptie niet zo groot zijn, is het wenselijk om vanuit dat perspectief naar concrete implementaties te gaan kijken. Verschillen in het computationele perspectief zijn namelijk niet inherent aan het verschil in implementatie en geven een hint in welke richting we verbeteringen in computerperceptie kunnen vinden. Daarnaast blijft het zinvol om ook naar het algoritmische perspectief te kijken. Hoewel hier significante verschillen te verwachten zijn, moeten er ook overeenkomsten zijn. Computationeel vergelijkbare algoritmes zullen dezelfde soort en hoeveelheid informatie opleveren. Dit zorgt ervoor dat er overeenkomsten in de representatie van die informatie kan zitten.

## 5 Concrete verschillen

Na een kader en een theoretische vergelijking is het tijd om te kijken naar een aantal concrete voorbeelden van verschillen en overeenkomsten tussen menselijke perceptie en computerperceptie. Het doel is een beeld krijgen van de verschillen die er concreet zijn tussen menselijke perceptie en computerperceptie. In plaats van een volledige vergelijking zal een aantal losse onderwerpen aan bod komen. Dit omdat er nog geen complete beschrijving is van het hele menselijke visuele systeem en omdat sommige onderdelen interessanter zijn dan andere. Maar ook binnen die onderwerpen is het niet altijd mogelijk om volledig te zijn. Wat relevant of bekend is in het ene vakgebied, is soms oninteressant of onbekend in het andere. Hierdoor is soms niets bekend over een fenomeen dat in het andere vakgebied uitgebreid besproken wordt.

## 5.1 Bewegingscontrast

Een van de eerste gebiedjes in het visuele systeem, V1, vervult een belangrijke taak, omdat het op basis van lichtintensiteitscontrasten randdetectie uitvoert [Marr, 1982]. Deze functie ligt aan de basis van de menselijke visuele perceptie. Alle visuele perceptie maakt gebruik van de voorverwerking in dit gebied [Sekuler and Blake, 2006]. Randdetectie in computers is een veel bestudeerd probleem waarvoor inmiddels goede algoritmes bestaan [Davis, 1975]. Computationeel gezien hebben deze algoritmes dezelfde functie, het detecteren van randen in de stimulus. Doordat de meeste computeralgoritmes serieel zijn, is de uitvoering van die algoritmes waarschijnlijk niet hetzelfde.

Een ander verschil dat minder voor de hand ligt, is dat voorverwerkingsalgoritmes vaak sterk verweven zijn met het algoritme waarvoor ze de voorverwerking uitvoeren [Bronstein et al., 2005]. Dit omdat het vaak efficiënter is meerdere dingen tegelijk te doen. Dit maakt het echter moeilijker om de in het voorverwerkingsalgoritme verkregen informatie voor andere algoritmes te gebruiken. Dit kan bijvoorbeeld komen omdat er door de verwevenheid van de algoritmes nooit een concrete representatie van de informatie uit het voorverwerkingsalgoritme beschikbaar gemaakt wordt.

Naast het kunnen waarnemen van randen op basis van intensiteit- en kleurcontrasten kunnen mensen ook randen detecteren in ander soort contrasten zoals beweging [Ullman, 1979]. Een sterk voorbeeld hiervan is het waarnemen van lijnen in random-dot-kinematogrammen. Deze kinematogrammen — filmpjes — zijn zo ontworpen dat een los beeldje uit de serie uit een patroonloze set pixels bestaat. In serie is echter beweging waar te nemen doordat de dots in verschillende delen van de stimulus een duidelijke bewegingsrichting hebben. In een stimulus waarbij twee vlakken met een contrasterende bewegingsrichting naast elkaar liggen, is een duidelijke scheidingslijn zichtbaar. Voor mensen met akinetopsia, het niet kunnen zien van beweging [Zihl et al., 1991], is dit contrast waarschijnlijk niet zichtbaar.

Het is opmerkelijk dat er in computeralgoritmen geen gebruik wordt gemaakt van andere contrasten. Ten eerste is het triviaal om de informatie over beweging om te zetten naar een formaat dat door de bestaande rand-detectiealgoritmes kan worden gebruikt. —Een voorbeeld van een simpel algoritme zou zijn om voor elke beeldpunt de vector van beweging om te zetten naar een kleurcodering waarbij de richting door een kleur gecodeerd wordt en de snelheid door de helderheid.— Ten tweede omdat het niet onwaarschijnlijk is dat onze perceptie van randen op basis van andere modaliteiten aanvullend en desambiguerend zou kunnen zijn, bovenop de informatie uit kleur en intensiteit. Dit kan zorgen voor een betere segmentatie in objecten. Een van de grote problemen binnen de computerperceptie is het segmenteren van invoer [Ullman, 1979, Horn, 1989]. Extra zekerheid levert betere informatie voor opvolgende algoritmes op en is dus zeer welkom.

## 5.2 Resolutie

Het menselijk oog heeft ongeveer 1.25 miljoen ganglioncellen. Dit zijn de cellen die hun signalen uiteindelijk via de oogzenuw naar de hersenen sturen ter verwerking [Sekuler and Blake, 2006]. Op een willekeurig moment kunnen er dus maximaal 1.25 miljoen stukjes informatie naar de hersenen verstuurd wor-

den. Vergelijk dit met een willekeurige camera en het oog legt het meestal af. De meeste videocamera's beginnen tegenwoordig bij één miljoen pixels en gaan door tot veel hogere resolutie. Toch is dit een verkeerde vergelijking. Het oog gebruikt namelijk allerlei strategieën om de hoeveelheid informatie die naar de hersenen gaat, in te dikken tot het hoogst nodige [Sekuler and Blake, 2006]. Het resultaat hiervan is dat hersenen meer kunnen met de slechts 1.25 miljoen "pixels" dan de computers die moeten werken met "scherpere" data uit moderne camera's. Het voordeel van minder informatie is duidelijk. Het scheelt verwerkingstijd, opslagruimte en uiteindelijk energie. Het grote probleem met het snijden in de binnenkomende informatie is dat er alleen overbodige informatie weggegooid mag worden. Om dat goed te doen, is kennis nodig over wat relevant is en wat niet, ander loopt men het risico om te veel informatie weg te snijden. Dit betekent dat je gespecialiseerde processen of algoritmes nodig hebt zoals die in het oog gebruikt worden.

**Resolutie** Aan de basis van de strategieën die in het oog gebruik worden ligt het feit dat het oog niet alles even scherp ziet. De resolutie in de fovea is vele malen hoger dan in de periferie van het oog [Sekuler and Blake, 2006]. Dit zorgt ervoor dat men altijd scherp genoeg kan zien zonder dat de hersenen het hele blikveld op een hoge resolutie moeten verwerken. Dit is iets wat computers soms parten speelt. Visuele-detectiealgoritmes werken namelijk veelal door middel van een venster op het blikveld te vergelijken met een model en dat venster steeds een stukje te verplaatsen of te draaien of van vorm te veranderen [Sung and Poggio, 2002]. Als het blikveld groter wordt, dan zijn er meer posities voor het venster en moet er vaker een 'dure' vergelijking uitgevoerd worden.

**Saccades** Het nadeel van een lage resolutie in de periferie is dat niet alles altijd scherp genoeg is. Om ervoor te zorgen dat mensen daar geen last van hebben, kunnen ze hun ogen bewegen. Door middel van saccades —korte snelle oogbewegingen— is het mogelijk om alle interessante stimuli toch scherp te krijgen [Sekuler and Blake, 2006]. Een belangrijke cue om naar een specifiek punt te kijken, zijn stimuli die ganglioncellen optimaal activeren. Detractoren zijn juist gradiënten met weinig contrast [Kienzle et al., 2009]. Dit is goed te modeleren in algoritmes. Saccades zorgen er dus voor dat we met ons beperkte blikveld toch de hele wereld scherp kunnen zien.

De techniek om de wereld te registeren levert voor computers gedetailleerdere informatie dan bij mensen. Dat lijkt beter omdat scherper zicht bij mensen immers ook als beter wordt ervaren, maar in dit geval is het juist een nadeel omdat computers nog te weinig gebruik maken van technieken die voorkomen dat ze overweldigd worden met informatie. De toepassing van die technieken hoeft echter niet op dezelfde manier te gebeuren als bij mensen. Eventuele softwarematige technieken die op een slimme en effectieve manier irrelevante informatie zouden kunnen snijden zijn waarschijnlijk een prima oplossing voor computers, tegenover de "hardwarematige" manier waarop mensen het doen.

### 5.3 Gezichtsherkenning

Binnen het vakgebied van computervisie is gezichtsherkenning één van de meest onderzochte onderwerpen [Zhao et al., 2003]. Dit onder andere omdat het een

van de meest commercieel toepasbare onderwerpen is. Er zijn namelijk veel situaties waarbij computers willen weten wie ze voor zich hebben, bijvoorbeeld bij bewakingscamera beelden, identificatie bij beveiligde instellingen of het inloggen op computers [Zhao et al., 2003]. Mensen gebruiken voor het herkennen van andere mensen gezichten en dus is het niet onlogisch om computers dat ook te laten doen. Op die manier is het zeker dat als mensen onderscheid kunnen maken, dat een computer genoeg informatie zou moeten hebben om dat ook te kunnen doen.

Er zijn een aantal verschillende soorten gezichtsherkenning algoritmes onder te verdelen in drie categorieën [Zhao et al., 2003].

**Holistisch** Bij de holistische methodes wordt herkenning gedaan op basis van het gehele gezicht [Turk and Pentland, 1991, Zhao et al., 2002]. Hierbij wordt geprobeerd om door middel van statische analyse de dimensionaliteit van de invoer te reduceren tot een bruikbaar en stabiel aantal dimensies. In feite is het een manier om alle redundante informatie te filteren [Zhao et al., 2003]. Eén van de gebruikte statische methodes is de zogenaamde principal-component-analysis. Deze methode is bedoeld om het aantal dimensies in een dataset te reduceren tot een minimale onoverlappende set. Het resultaat hiervan kan een zogenaamde set eigenfaces opleveren. Eigenfaces zoals in Figuur 1 karakteriseren de mogelijke vari-



Figuur 1: *Eigenfaces*

abelen die ten grondslag liggen aan alle mogelijke gezichtsvarianties. Een te herkennen gezicht wordt geanalyseerd aan de hand van de eigenfaces. Dit levert een set gewichten op die vergeleken kan worden met gewichten van bekende gezichten. Het idee is dat dezelfde gezichten altijd vergelijkbare gewichten opleveren ten opzichte van de vaste set eigenfaces [Zhao et al., 2003].

**Features** Bij de feature-gebaseerde methodes worden de onderdelen van het gezicht geëxtraheerd en hun kenmerken en relaties tot elkaar geanalyseerd. Hierbij wordt bijvoorbeeld gekeken naar de breedte van het voorhoofd of de afstand tussen de ogen [Zhao et al., 2003, Manjunath et al., 1992].

**Hybride** In de laatste categorie vallen de methodes die methodes uit de voorgaande categorieën combineren. [Penev and Atick, 1996]

De verschillen tussen menselijke perceptie en computerperceptie liggen vooral in het algoritmische perspectief. De meeste gezichtsherkenning algoritmes werken op basis van een afbeelding van de werkelijkheid. Dit kan een foto of video zijn en bevat soms ook diepteinformatie. Wel is het zo dat relatief veel algoritmes bedoeld voor video eigenlijk de losse beelden als invoer hebben en weinig tot geen gebruik maken van de temporele informatie die bewegende beelden kunnen bevatten [Zhao et al., 2003]. Dit betekent dat de invoer voor gezichtsherkenning algoritmes alleen spatieel gestructureerd is en dat de structuur van

de informatie weinig te maken heeft met gewenste informatie zoals welke gezichten er te zien zijn. Het verwerken en herkennen van gezichten door mensen is een complex proces dat op meerdere plekken in de hersenen gelokaliseerd is. Het voor gezichtsidentificatie belangrijkste gebiedje hiervan is de laterale fusievorm gyrus, ook bekend als het fusievormgezichtsgebied [Haxby et al., 2002]. Dit gebied verkrijgt zijn informatie uit het vroege visuele systeem. Onbekend is wat de precieze representatie is waarmee gewerkt wordt, maar omdat er al voorverwerking plaatsgevonden heeft, is het niet onwaarschijnlijk dat de structuur van de informatie beter toegespitst is op de benodigde informatie en dat onnodige informatie ontbreekt of makkelijk te filteren is.

Doordat computergezichtsherkenning vaak gebruik maakt van geïntegreerde voorverwerking, is het lastiger om die voorverwerking op basis van verschillende bronnen te doen. Al deze bronnen moeten namelijk geïntegreerd worden en doordat klassieke computeralgoritmes vaak gebruik maken van voor hun specifieke datastructuren voor representatie, is het niet altijd gemakkelijk of mogelijk om informatie uit verschillende bronnen samen te voegen. Dit is een nadeel voor de huidige computerperceptiealgoritmes, aangezien mensen dit waarschijnlijk wel goed kunnen. Algoritmes die op basis van tweedimensionale stimuli werken, zoals foto's en film, zouden gebruik kunnen maken van een heel scala aan diepte-informatie die in de stimulus zit. Mensen hebben immers niet minder dan zes verschillende manieren om diepte-informatie te extraheren uit een stimulus; zelfs als ze maar van één oog gebruik maken [Sekuler and Blake, 2006]. Om die informatie te kunnen gebruiken, moet er echter wel voorverwerking plaatsvinden op een manier waarbij al die bronnen weer geïntegreerd kunnen worden.

Dat deze diepte-informatie belangrijk is, is begrijpelijk aangezien de vorm van het gezicht bij gezichtsherkenning belangrijk is. Meer diepte-informatie, ofwel meer zekerheid over de driedimensionale vorm van het gezicht, draagt aan een beter model van het te herkennen gezicht. Dit maakt het algoritme minder gevoelig voor irrelevante verschillen in de invoer en zorgt voor een accurater algoritme [Bronstein et al., 2005]. Hoe belangrijk driedimensionale informatie is, zal duidelijk worden in het volgende hoofdstuk waarin een nieuwste generatie algoritme wordt behandeld, dat gebruik maakt van driedimensionale informatie en een zeer goede prestatie levert.

## 6 3DFACE, een gezichtsherkeningsalgoritme

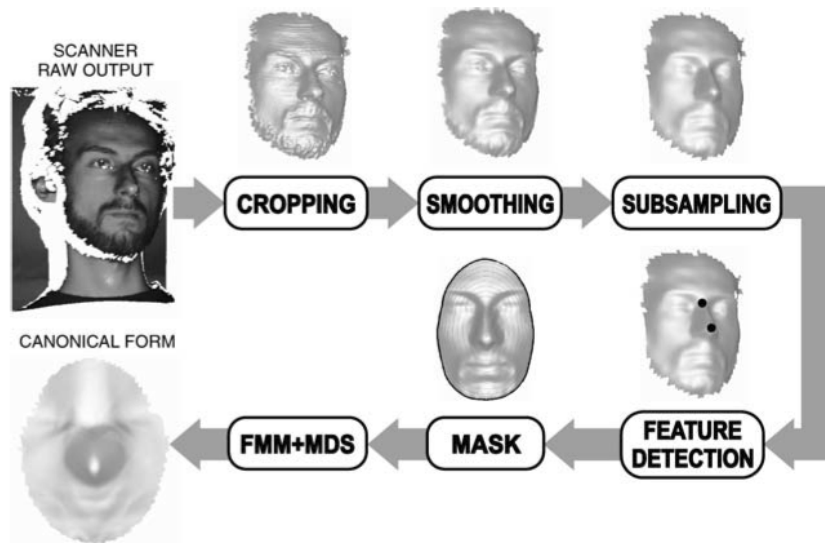
Een grote uitzondering op de regel dat computers slechter presteren in visuele taken is de nieuwste generatie gezichtsherkeningsalgoritmes. De laatste jaren zijn er nieuwe technieken bedacht die een honderdvoud beter presteren dan de gezichtsherkeningsalgoritmes van tien jaar geleden [Bronstein et al., 2005]. Dit komt doordat ze gebaseerd zijn op het herkennen van driedimensionale beelden. De reden waarom dat pas recent mogelijk is, is omdat er gebruik wordt gemaakt van 3D-cameras om diepteinformatie te kunnen detecteren. Om de vergelijking tussen menselijke perceptie en computerperceptie nog concreter te maken, zal dit hoofdstuk een vergelijking maken tussen dit nieuwe systeem —3DFACE— en menselijke perceptie [Bronstein et al., 2005].

3DFACE is een systeem dat in de categorie "holistische algoritmes" valt. Er wordt wel gekeken naar de onderdelen van het gezicht, maar dit is uitsluitend om een referentiepunt te bepalen en heeft verder geen rol in de herkenning van

gezichten.

## 6.1 3DFACE in vogelvlucht

Het 3DFACE-algoritme bestaat uit een serie stappen, zoals weergegeven in Figuur 2. De eerste drie stappen in het 3DFACE-algoritme zorgen voor de voor-

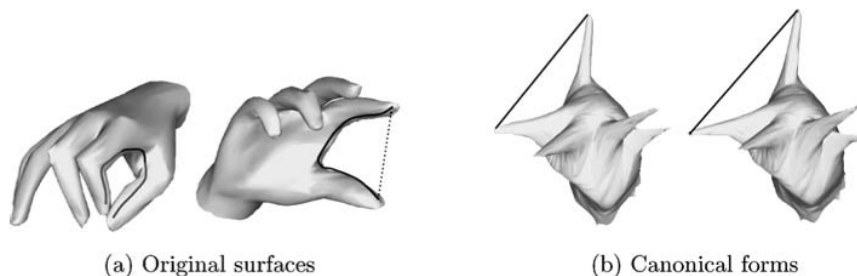


Figuur 2: De stappen in het 3DFACE algoritme

bewerking van de data van de 3D-camera. Deze dataset bestaat uit een verzameling punten in de ruimte. Om hieruit het gezicht te extraheren worden de punten in de achtergrond en punten die sterk afwijken door meetfouten weggefilterd. Vervolgens wordt het gezicht gladgestreken. Dit omdat de originele data een relatief "ruw" gezicht oplevert. Als derde stap wordt een aantal vaste punten op het gezicht gelokaliseerd. Dit gebeurt door naar de "hoeken" in het gezicht te kijken. Een aantal punten, zoals het topje van de neus en de ogen, worden hiermee gelokaliseerd. Hier zijn de bochten in het gezicht het scherpst. Het doel van de deze stap is uiteindelijk om het puntje van de neus en de bovenkant van de neus te lokaliseren. Deze punten zijn ongevoelig voor veranderingen door uitdrukkingen. De laatste stap in de voorverwerking is het bijsnijden van het gezicht tot een gebied rond het puntje van de neus waarbij alle punten in het gezicht binnen een bepaalde geodetische afstand —de afstand over de oppervlakte van het gezicht— tot het puntje van de neus liggen. Dit zorgt ervoor dat de gebruikte data altijd een consistente vorm heeft

Na de voorverwerking worden de gezichten getransformeerd naar een zogenaamde canonieke vorm. Deze vorm heeft de eigenschap dat hij relatief ongevoelig is voor veranderingen door uitdrukkingen. Deze canonieke vorm wordt verkregen door het model zo te veranderen dat de geodetische afstanden tussen de punten gelijk zijn aan de euclidische afstanden —de absolute afstanden

tussen twee punten— zoals getoond in Figuur 3. Deze canonieke vorm kan



Figuur 3: *Illustratie van de relatie tussen oorspronkelijke vorm (a) en canonieke vorm (b). De zwarte lijn is de geodetische afstand en de grijze lijn is de euclidische afstand. In de canonieke vorm zijn deze afstanden gelijk*

vervolgens vergeleken worden met gezichten in de database door middel van standaard oppervlakte-vergelijkingsalgoritmes.

## 6.2 Informatieverwerking

Er is een aantal duidelijke verschillen tussen het 3DFACE algoritme en menselijke gezichtsherkenning, te beginnen met de informatieverwerking. Hoewel mensen met twee ogen drie dimensies kunnen waarnemen, is het niet het geval dat ze die informatie direct uit het oog verwerven. Bij mensen gaat hier een aantal stappen aan vooraf waarbij verschillende stukjes informatie worden gebruikt om diepteinformatie uit te extraheren.[Sekuler and Blake, 2006] De opkomst van 3D-camera's die vanuit één kijkpunt ook diepteinformatie kunnen ontvangen, maakt het mogelijk om deze voorverwerkingsstappen over te slaan. Het is mogelijk, dat dit zelfs een betere manier is dan de menselijke manier, omdat deze inherent ambigue diepteinformatie levert in sommige omstandigheden. Aan de andere kant levert de noodzaak van een 3D-camera ook een limitatie op. 3D-camera's kunnen natuurlijk weinig met twee dimensionale foto's van mensen. Mensen kunnen daarentegen uit foto's vaak ook nog diepteinformatie halen.

## 6.3 Voorverwerking

Het is moeilijk te zeggen of de voorverwerking in het 3DFACE-algoritme enigszins lijkt op de menselijke voorverwerking. Het echter is moeilijk voor te stellen dat bij mensen er geen filter- of bijsnijdproces plaatsvindt om ervoor te zorgen dat niet meer informatie dan nodig door de laterale fusievorm gyrus verwerkt wordt. Dit zou namelijk betekenen dat elk proces in de hersenen alle informatie nodig zou hebben en dat lijkt onwaarschijnlijk. Een gezichtsherkenningproces hoeft immers niet per se te weten waar het gezicht zich bevindt. Hoe en waar een dergelijk filter proces plaatsvindt, is stof voor toekomstig onderzoek.

## 6.4 Herkenning

De manier waarop het 3DFACE-algoritme de gezichten transformeert tot een bruikbare dataset, is ook bruikbaar voor allerhande andere vormen. Dit vertoont overeenkomsten met het menselijke laterale fusieform gebied. Dit is namelijk niet gelimiteerd tot gezichten en na training ook inzetbaar voor visuele expertise op andere gebieden, zoals het herkennen van auto's, duiven en "greebles" [Tarr and Gauthier, 2000].

## 7 Conclusie

Een vergelijking tussen menselijke perceptie en computerperceptie is lastig omdat het over twee verschillende vakgebieden gaat. Om dit succesvol te doen is een kader nodig dat vastlegt wat de vergelijkingscriteria zijn. Daarnaast is de drie-niveau-hypothese van Marr essentieel omdat deze voorkomt dat verkeerde aspecten met elkaar vergeleken worden.

De theoretische vergelijking levert op dat computers en mensen qua implementatie totaal verschillen. Doordat we kunnen aannemen dat ze beide Turing-compatible zijn, levert dit geen onoverkomenlijke verschillen op. Een vergelijking op het algoritmische niveau is lastig omdat hier de verschillen een rol spelen, maar overeenkomsten ook aanwezig moeten zijn. Het computationele niveau zou echter bijna geheel vrij moeten zijn van verschillen.

Dit schept mogelijkheden om in het computationele perspectief een concrete vergelijking uit te voeren om te kijken of systemen in de praktijk ook overeenkomen. Hier worden grote overeenkomsten gevonden, maar ook verschillen.

Een deel van de verschillen tussen computerperceptie en menselijke perceptie is inherent aan het verschil in implementatie. Als de informatie van een 3D-camera afdoende is voor een intelligent systeem, dan is het niet noodzakelijk om in visuele voorverwerking alle cues voor driedimensionele informatie te gaan verwerken. Computationeel doen beide manieren hetzelfde, namelijk driedimensionale informatie leveren. Als dat de taak is, is het dus niet wenselijk om ondanks de verschillen te proberen om computers exact het evenbeeld van de mens te maken. Op die manier mis je namelijk de kans om gebruik te kunnen maken van de sterke punten van de implementatietechniek. Bijvoorbeeld het feit dat een 3D-camera wellicht beter en sneller is dan stereovisie in combinatie met voorverwerking.

Sommige verschillen lijken echter ingegeven door limieten aan kunnen, implementatietijd, kosten en gebrek aan kennis. Zoals het niet gebruikmaken van andere contrasten dan die van kleur en licht/donker. Het is wellicht wenselijk om deze niet-inherente verschillen weg te werken en —rekening houdend met de vorige alinea— zoveel mogelijk inspiratie bij mensen op te doen. Immers, de natuur heeft al miljoenen jaren gehad om het ontwerp te verbeteren en het is wellicht verspilde moeite om voor alles het wiel opnieuw uit te vinden.

## 8 Discussie

Ondanks dat mensen en computers qua implementatie zo verschrikkelijk verschillend zijn, is het opvallend dat vanuit het computationele perspectief er



zulke grote gelijkenissen zijn. Doordat deze gelijkenissen er zijn, blijft het interessant om mensen en computers met elkaar te vergelijken. Zeker omdat er in beide kennisgebieden nog significante gaten zitten en de overeenkomsten maken het mogelijk om oplossingen en concepten uit het ene vakgebied te gebruiken in het andere vakgebied.

## 8.1 Toekomstig onderzoek

Er is een aantal interessante mogelijkheden die voortvloeit uit deze scriptie voor toekomstig onderzoek. Doordat het zwaartepunt van deze scriptie op visuele perceptie lag, zijn er genoeg mogelijkheden om overeenkomstige vergelijkingen te maken voor andere modaliteiten van perceptie. Het onderzoek kan zelfs uitgebreid worden naar andere aspecten van informatieverwerkende systemen, zoals de verwerking en uitvoer.

Een concrete optie voor psychologieonderzoek is onderzoeken of andersoortige contrasten, zoals bewegingscontrasten, op dezelfde manier verwerkt worden in V1 als licht- en kleurcontrasten.

## 8.2 Vergelijkingscriteria

In het kader is geprobeerd om een aantal criteria te benoemen die relevant zijn voor een vergelijking tussen perceptie bij mens en computer. Deze lijst is helaas niet uitputtend gemaakt, dat wel doen is een onderzoek an-sich en buiten de scope van deze scriptie.

## 8.3 Alternative modaliteiten

Veel computerperceptietechnieken gebruiken dezelfde modaliteiten als mensen om aan hun informatie te komen, ondanks dat er allerlei alternatieven zijn zoals bijvoorbeeld sonar of infraroodzicht. Dit is begrijpelijk. Ten eerste omdat deze modaliteiten bekend terrein zijn voor mensen, aangezien ze zelf ook in die modaliteit kunnen waarnemen. Dit vergemakkelijkt het ontwerpen van de benodigde componenten. Als iemand bijvoorbeeld wil controleren of een camera werkt, is een blik op een computerscherm vaak voldoende. Bij andere modaliteiten is een vertaalslag naar een bekende modaliteit nodig voordat iemand zelf kan waarnemen wat de betreffende receptor waarneemt. Een andere reden is dat andere modaliteiten onderzoek vereisen om te verifiëren dat de modaliteit bruikbaar is. Bij onze aangeboren modaliteiten is het makkelijk om te verifiëren of de benodigde informatie voor een taak aanwezig is.

Hoewel we veel lessen kunnen leren van menselijke perceptie om toe te passen op computerperceptie, sluit dit niet uit dat er effectieve vormen van computerperceptie zijn die gebruik maken van modaliteiten die mensen niet bezitten. Dit valt echter buiten de grenzen van deze scriptie.

## Referenties

- [Bach-y-Rita and Kercel, 2003] Bach-y-Rita, P. and Kercel, S. W. (2003). Sensory substitution and the human-machine interface. *Trends in Cognitive Sciences*, 7(12):541–546.
- [Bronstein et al., 2005] Bronstein, A. M., Bronstein, M. M., and Kimmel, R. (2005). Three-Dimensional face recognition. *International Journal of Computer Vision*, 64(1):5–30.
- [Brooks, 1991] Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, 47(1-3):139–159.
- [Davis, 1975] Davis, L. S. (1975). A survey of edge detection techniques. *Computer Graphics and Image Processing*, 4(3):248–270.
- [Haxby et al., 2002] Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological psychiatry*, 51(1):59–67.
- [Honavar, 1994] Honavar, V. (1994). Symbolic artificial intelligence and numeric artificial neural networks: towards a resolution of the dichotomy. *Computational Architectures Integrating Neural and Symbolic Processes*, page 351–388.
- [Horn, 1989] Horn, B. K. (1989). Obtaining shape from shading information. In *Shape from shading*, page 123–171.
- [Kienzle et al., 2009] Kienzle, W., Franz, M. O., Schölkopf, B., and Wichmann, F. A. (2009). Center-surround patterns emerge as optimal.
- [Manjunath et al., 1992] Manjunath, B. S., Chellappa, R., and der Malsburg, C. V. (1992). A feature based approach to face recognition. In *1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92.*, page 373–378.
- [Markus, 2009] Markus, T. (2009). A hybrid approach to symbol grounding.
- [Marr, 1982] Marr, D. (1982). *Vision: A computational approach*. Freeman & Co., San Francisco.
- [McGarry et al., 1999] McGarry, K., Wermter, S., MacIntyre, J., and Campus, S. P. S. P. (1999). Hybrid neural systems: from simple coupling to fully integrated neural networks. *Neural Computing Surveys*, 2(1):62–93.
- [Penev and Atick, 1996] Penev, P. S. and Atick, J. J. (1996). Local feature analysis: a general statistical theory for object representation. *Network: computation in neural systems*, 7(3):477–500.
- [Quiroga et al., 2005] Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107.
- [Sekuler and Blake, 2006] Sekuler, R. and Blake, R. (2006). *Perception*. McGraw-Hill.

- [Sung and Poggio, 2002] Sung, K. K. and Poggio, T. (2002). Example-based learning for view-based human face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(1):39–51.
- [Tarr and Gauthier, 2000] Tarr, M. J. and Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *nature neuroscience*, 3:764–770.
- [Turk and Pentland, 1991] Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86.
- [Ullman, 1979] Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 203(1153):405.
- [Zhao et al., 2002] Zhao, W., Chellappa, R., and Krishnaswamy, A. (2002). Discriminant analysis of principal components for face recognition. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, page 336–341.
- [Zhao et al., 2003] Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003). Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 35(4):399–458.
- [Zihl et al., 1991] Zihl, J., Von Cramon, D., Mai, N., and Schmid, C. (1991). Disturbance of movement vision after bilateral posterior brain damage: further evidence and follow up observations. *Brain*, 114(5):2235.