

Een systeem dat tweedimensionale vormen kan leren herkennen

E. H. J. Persoon

Ondanks de grote vooruitgang op het gebied van de beeldbewerking in de laatste twintig jaar zijn er nog vele 'taken' in dit gebied aan te wijzen die een mens moeiteloos afdaan, maar die nog niet op bevredigende wijze door een machine kunnen worden verricht. Eén daarvan is het herkennen van voorwerpen aan hun vorm in twee dimensies [1]. We denken daarbij aan min of meer platte voorwerpen op een vlakke ondergrond, die willekeurig gerangschikt zijn en elkaar eventueel raken of zelfs ten dele over elkaar heen liggen. Deze herkenningstaak is onder meer van belang voor het automatiseren van assemblage. De machine zou onderdelen die ongeordend op een lopende band zijn gestrooid na herkenning kunnen grijpen en positioneren. Bij het inspecteren van lopende-bandprodukten of bij het identificeren en tellen van voorwerpen is het herkennen van de juiste vorm eveneens essentieel.

In dit artikel beschrijven we een herkenningssysteem dat door ons is ontwikkeld en dat twee markante eigenschappen heeft. De eerste is deze, dat het systeem een te herkennen vorm eerst *zelf leert kennen*. Door deze eigenschap is het systeem zeer flexibel: binnen zekere grenzen kan het voor het herkennen van elke willekeurige vorm worden gebruikt. Ten tweede kan het systeem een voorwerp óók herkennen als het door andere voorwerpen geraakt of zelfs gedeeltelijk bedekt wordt.

Om duidelijk te maken waar de moeilijkheden liggen bij het ontwerpen van een systeem dat gedeeltelijk bedekte voorwerpen kan herkennen, beschouwen we in het kort twee herkenningssystemen aan de hand van fig. 1. We veronderstellen dat de rechthoek van fig. 1a op de één of andere manier in het geheugen van het systeem is opgeslagen; van het systeem wordt gevraagd of de rechthoek in de beelden *b* en *c* aanwezig is.

Bij de eerste methode verschuift en verdraait men de rechthoek over de beelden en gaat na hoe goed de 'passing' is. Zowel in *b* als in *c* vindt men dat er één stand is waarin de rechthoek bijzonder goed past, en 'dus' aanwezig is. Met deze methode kan men dus — in principe — bedekte voorwerpen wel herkennen; de uitwerking ervan per computer vergt echter te veel tijd. Laten we aannemen dat het om zwart-witbeelden gaat op een beeldscherm met 100×100 beeldpunten; elk

beeldpunt kan dus de waarde 1 of 0 hebben (zwart of wit zijn). 'Passing' betekent dat een zekere functie f van de stand van de rechthoek een scherp minimum moet vertonen; f is b.v. het aantal beeldpunten dat in de te vergelijken beelden van waarde verschilt. Welnu, in de eerste plaats is het berekenen van een nieuwe stand — het uitvoeren van de transformaties 'verschuiving' en 'draaiing' voor elk punt van de rechthoek of van de rand van de rechthoek — al zeer bewerkelijk. In de tweede plaats is het aantal mogelijke standen zeer groot: globaal 100 oriëntaties bij elk van 100×100 posities. Een eerste verfijning zou hieruit kunnen bestaan dat men aan transformaties de voorrang geeft waarbij f kleiner wordt, maar ook dit heeft vaak geen succes doordat de rechthoek in een *relatief* minimum van f terecht komt, en niet in het gezochte *absolute* minimum.

De tweede methode die we hier willen aanstippen is veel effectiever, maar alleen bruikbaar als de voorwerpen in de scène elkaar niet raken. Men bepaalt de coördinaten van de beeldpunten van de voorwerpen,

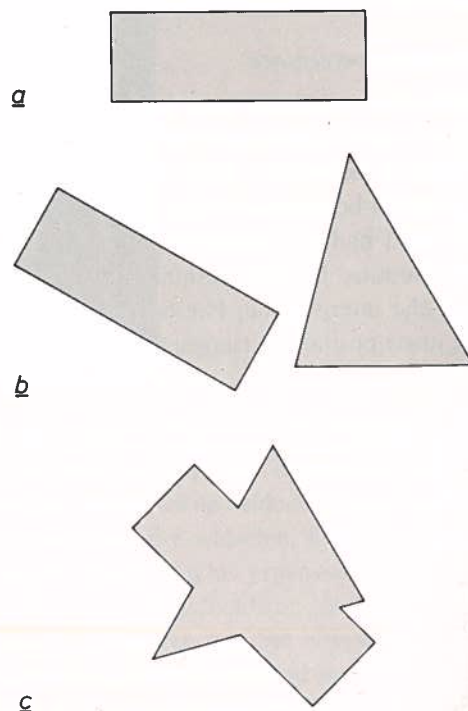


Fig. 1. Voorbeeld van een herkenningsprobleem. Men tracht systemen te ontwerpen die de rechthoek (a) in de beelden (b) en (c) herkent.

bij voorbeeld door aftasting van de rand ^[2], en berekent daaruit bepaalde vormkenmerken, zoals de grootste afstand van het zwaartepunt tot de rand of de 'circulariteit' (d.i. de verhouding van de oppervlakte tot het kwadraat van de omtrek — een getal dat voor een cirkel het grootst is). Bij voldoende overeenstemming van deze kenmerken met die van het gezochte voorwerp is dit 'herkend'. Als de voorwerpen elkaar echter raken, is deze methode onbruikbaar: er is geen criterium dat uitmaakt wanneer men van het ene voorwerp op het andere overgaat.

In ons systeem wordt een voorwerp in scènes als die van fig. 1c in eerste instantie herkend aan duidelijke 'lokale' kenmerken ter plaatse van de onbedekte stukken, b.v. de rechte hoeken 1, 2, 3 en 4 in fig. 2. We zullen zulke lokale kenmerken in het vervolg *vormelementen* noemen. Het is echter duidelijk dat, als het systeem eenmaal geconstateerd heeft dat de getoonde scène een aantal van zulke elementen met het gegeven voorwerp gemeen heeft, dit toch slechts een zwak criterium voor herkenning zou vormen als men niet tevens de relatieve posities en oriëntaties van de herkende elementen in aanmerking neemt. Een vorm wordt in ons systeem derhalve in *twee fasen* vastgelegd: in de eerste fase worden de vormelementen opgespoord, in de tweede fase wordt de relatieve ligging van de gevonden elementen vastgelegd. Dezelfde twee fasen komen terug bij het analyseren van een scène. Eerst wordt nagegaan of

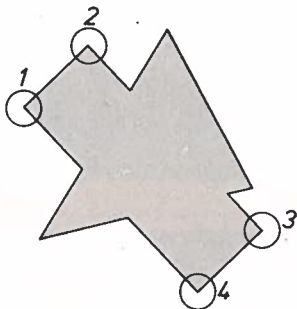


Fig. 2. De rechte hoeken 1, 2, 3 en 4 vormen markante, 'lokale' kenmerken van de rechthoek die voor de herkenning kunnen worden gebruikt, ondanks het feit dat de rechthoek gedeeltelijk door een driehoek wordt bedekt.

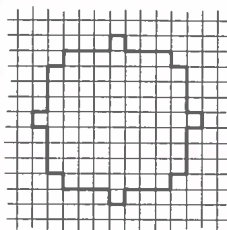


Fig. 3. Het venster waarmee vormelementen worden opgespoord: de beste benadering van een cirkel met een diameter van elf beeldpunten.

deze de gegeven vormelementen bevat en vervolgens of zij de juiste relatieve ligging hebben. Deze procedure maakt het mogelijk met betrekkelijk weinig elementen per voorwerp een vorm toch goed vast te leggen. Bovendien komen bij het herkennen van meer dan één vorm dezelfde elementen vaak weer terug, zodat de vereiste geheugencapaciteit niet evenredig toeneemt met het aantal te herkennen vormen.

Het opsporen en herkennen van de vormelementen

De door ons gebruikte vormelementen zijn beeldfragmenten die door het venster van fig. 3 uit het beeld worden gesneden. Het gekozen venster is de beste benadering van een cirkel met een diameter van elf beeldpunten op het orthogonale raster. De keuze van de diameter is een compromis: kiest men het venster te klein, dan kan men geen vormen — rechte randen, scherpe hoeken, rechte hoeken, enzovoort — meer onderscheiden; kiest men het te groot, dan wordt het aantal mogelijke vormelementen te groot. Te zamen met de volgende maatregelen blijkt het venster van fig. 3 een geschikte graad van onderscheidend vermogen te leveren.

Het is duidelijk dat niet elk beeld in dit venster als vormelement bruikbaar is: het venster heeft 81 beeldpunten en er zijn derhalve 2^{81} mogelijke beelden. Het aantal mogelijke vormelementen wordt drastisch beperkt door een vensterbeeld pas als 'vormelement' te accepteren als het zwaartepunt van de door het venster geselecteerde randpunten van het volledige beeld in het centrum van het venster ligt.

Om zulke vormelementen te vinden, wordt het beeldscherm van links naar rechts en van boven naar beneden door het venster afgetast. Zodra in het venster één of meer zwarte punten verschijnen, wordt het zwaartepunt van de geselecteerde randpunten berekend. Het venster wordt vervolgens zó verschoven dat dit punt in het centrum terecht komt. Doordat het beeld in het venster daarbij verandert, is na deze stap de rand meestal nog niet geheel gecentreerd, maar na nog één of twee zwaartepuntsberekeningen en verschuivingen is dit wel het geval. Bij dit alles verstaan wij onder 'randpunten' van een beeld de zwarte en de witte punten die resp. een wit en een zwart punt in horizontale of verticale richting tot buur hebben (zie fig. 4).

[1] M. Yachida en S. Tsuji, A versatile machine vision system for complex industrial parts, IEEE Trans. C-26, 882-894, 1977.
W. A. Perkins, A model-based vision system for industrial parts, IEEE Trans. C-27, 126-143, 1978.

B. Neumann, Interpretation of imperfect object contours for identification and tracking, Proc. 4th Int. Joint Conf. on Pattern recognition, Kyoto 1978, blz. 691-693.

[2] E. H. J. Persoon en K. S. Fu, Shape discrimination using Fourier descriptors, IEEE Trans. SMC-7, 170-179, 1977.
P. Saraga en J. A. Weaver, Een experiment met flexibele automatisering, dit nummer, blz. 340.

De coördinaten x_c, y_c van het zwaartepunt worden uit de matrix $D(x, y)$ van de rand (zie fig. 4) als volgt berekend:

$$x_c = \frac{\sum D(x, y)x}{\sum D(x, y)}, \quad y_c = \frac{\sum D(x, y)y}{\sum D(x, y)},$$

waarin de sommaties (Σ) zich uitstrekken over alle vensterwaarden van x en y . In fig. 4 geldt: $x_c = 0$, $y_c = 2\frac{1}{2}$. Een rand heet gecentreerd als de afstand van het zwaartepunt tot het centrum van het venster kleiner is dan of gelijk aan $\frac{1}{2}$.

Vormelementen moeten met elkaar vergeleken kunnen worden om uit te maken of zij 'gelijk' dan wel 'verschillend' zijn. Bij het vergelijken van twee beelden P en Q telt men gewoonlijk de beeldpunten die in P een andere waarde hebben dan in Q . Het gevonden aantal, $d(P, Q)$, is de conventionele 'afstandsmaat' tussen P en Q . De twee beelden worden dan als verschillend beschouwd als $d(P, Q)$ groter is dan een gegeven getal. Dit criterium doet echter geen recht aan het volgende belangrijke verschijnsel aan de rand. Van de binaire beelden die wij verwerken, zijn de beeldpunten wit of zwart naargelang het videosignaal een gegeven drempel al of niet overschrijdt. Bij een rand ligt het signaal dicht bij de drempel. Door kleine variaties in de drempel of ruis in het signaal, en eveneens door de variatie van de plaats van bemonstering ten opzichte van het beeld, kunnen de punten van de rand gemakkelijk van waarde veranderen. De beelden A en B in fig. 5 kunnen dus zeer wel van hetzelfde stukje van een gegeven scène afkomstig zijn. Voor de beelden A en C is dat zeer onwaarschijnlijk. Toch is de conventionele afstandsmaat tussen A en B gelijk aan die tussen A en C : $d(A, B) = d(A, C) = 5$. Wij hanteren daarom een andere afstandsmaat, $d'(P, Q)$. Daarin worden, evenals in $d(P, Q)$, beeldpunten niet meegeteld als ze in P en Q dezelfde waarde hebben; punten met verschillende waarde worden echter *ook niet meegeteld als ze in beide beelden randpunten zijn*. Ons criterium voor de gelijkheid van twee beelden P en Q luidt nu: $d'(P, Q) = 0$. Dit is dus ruimer dan het overeenkomstige conventionele criterium ($d = 0$): het laat waardeverschillen van randpunten toe. Volgens het nieuwe criterium zijn in fig. 5 A en B gelijk, maar A en C verschillend: $d'(A, B) = 0$, $d'(A, C) = 5$.

Samengevat verstaan wij onder *gelijke vormelementen*: vensterbeelden waarin het zwaartepunt van randpunten gecentreerd is en waarvoor d' gelijk aan nul is. Bij het leren worden de opgespoorde vormelementen opgeslagen in het geheugen als zij verschillen van de reeds eerder opgeslagen elementen. Bij het herkennen wordt nagegaan welke elementen in het geheugen ook in de aangeboden scène voorkomen. Dit zijn in het kort de eerste fase van het leren en de eerste fase van het herkennen.

Ten einde het aantal vormelementen dat moet worden opgeslagen verder te beperken, maken wij bij het

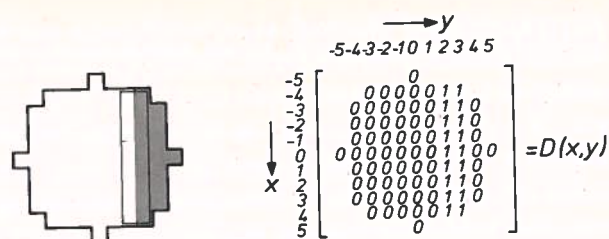


Fig. 4. Randpunten (omlijst) in een eenvoudig vensterbeeld. Rechts de bijbehorende matrix $D(x, y)$.

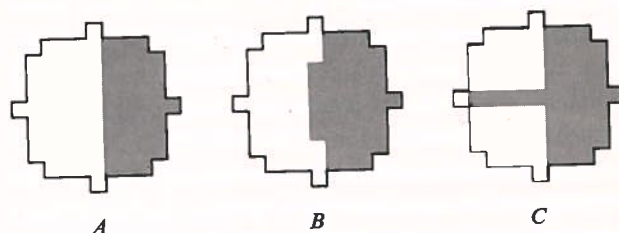


Fig. 5. Drie vensterbeelden. De conventionele 'afstandsmaat' d (het aantal beeldpunten dat in waarde verschilt) tussen A en B is gelijk aan die tussen A en C ($d = 5$). Met deze maat gemeten, zijn A en C gelijk als men A en B als gelijk rekent ('elementen zijn gelijk als $d \leq 5$ is'); omgekeerd zijn A en B verschillend als men A en C als verschillend rekent ('elementen zijn verschillend als $d \geq 5$ is'). In het hier gebruikte systeem worden A en B als gelijk beschouwd, omdat het verschil slechts randpunten betreft. Dit geldt niet voor A en C , die zijn dus 'verschillend'.

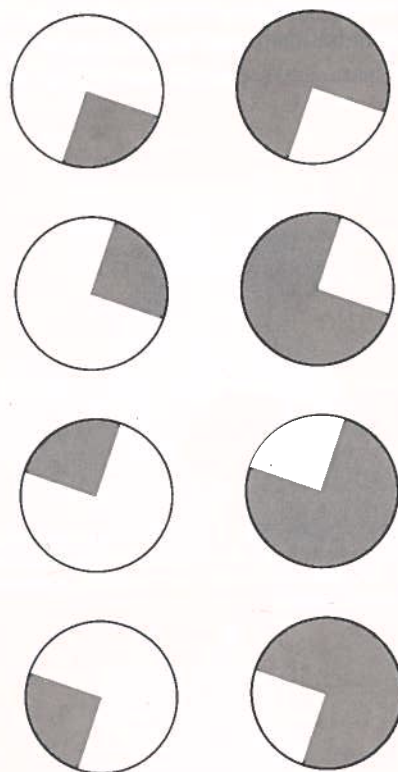
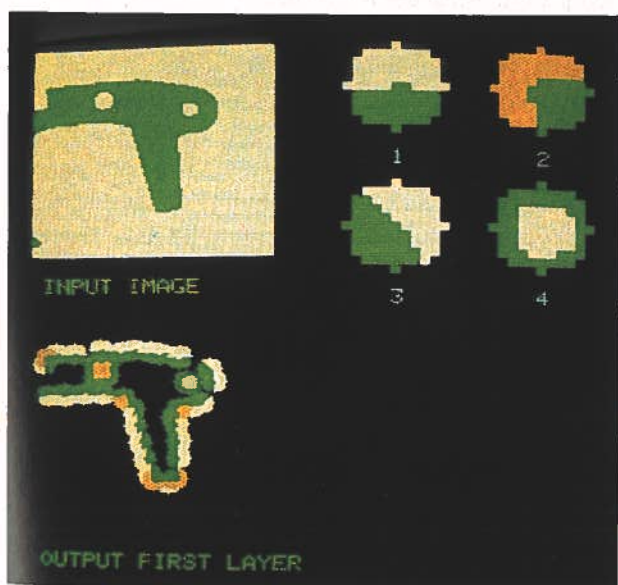
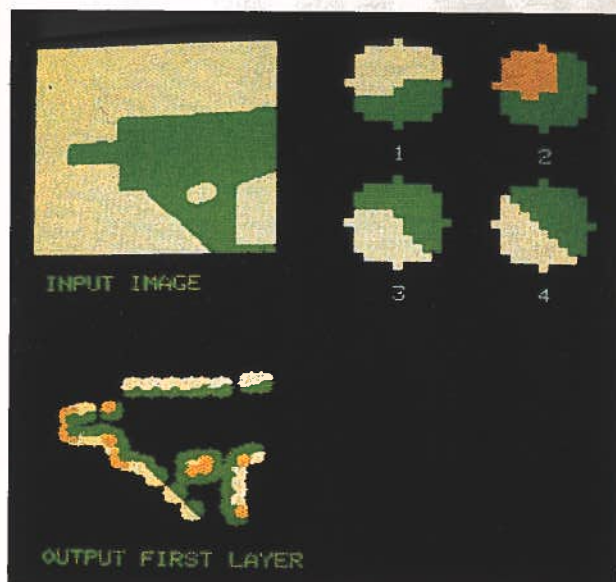
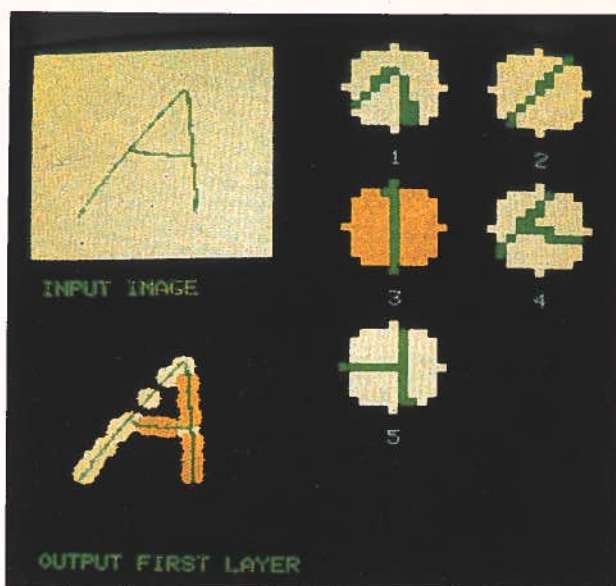


Fig. 6. Een vormelement (links boven) en de daarvan afgeleide elementen. Het venster is in deze figuur als een cirkeltje getekend. Als het element links boven eenmaal in het geheugen is opgeslagen, worden nieuwe elementen niet alleen met dit element, maar effectief ook met de andere zeven vergeleken. Dit bespaart ongeveer een factor 8 in vereiste geheugencapaciteit. Met het centreren van de rand is in deze figuur geen rekening gehouden.



identificeren van elementen nog gebruik van enige, voor de computer zeer eenvoudige en korte, bewerkingen: het roteren over 90° , 180° en 270° en het invertieren van de beeldpuntwaarden (0 wordt 1 en 1 wordt 0). Bij het vergelijken worden deze bewerkingen op het opgeslagen element toegepast; een nieuw element wordt dus in feite niet alleen met de opgeslagen elementen zelf vergeleken, maar ook met de elementen die via deze operaties daaruit worden afgeleid. Als dus b.v. het element in *fig. 6* links boven eenmaal is opgeslagen, wordt het eveneens gebruikt voor het identificeren van de andere zeven in deze figuur. De maatregel reduceert de vereiste geheugencapaciteit (ongeveer) met een factor 8. (In *fig. 6*, en ook in *fig. 9* en *13*, is het venster gemakshalve als een cirkeltje getekend.)

Uit *fig. 7* moge blijken dat men op deze wijze met een verrassend gering aantal elementen uitkomt. In elk van de foto's ziet men links boven het binaire beeld dat op een scherm van 100×100 punten ter lering werd aangeboden. Rechts boven ziet men, sterk vergroot, de vormelementen die daarbij werden gevonden; slechts vijf in de eerste en vier in de beide andere foto's. Het beeld links onder werd verkregen door elk element — weer op de oorspronkelijke schaal — af te beelden op de plaats en in de oriëntatie- en inversietoestand waarin het werd opgespoord of herkend. In de eerste foto hebben we element 3 een aparte kleur gegeven; daardoor vindt men het gemakkelijk terug, ook in het horizontale segment. In de andere foto's herkent men op dezelfde wijze gemakkelijk het element 2, al of niet gerooteerd of geïnverteerd.

De gevonden elementen zijn in het algemeen niet bruikbaar voor het herkennen van de afgebeelde voorwerpen als deze over een willekeurige hoek zijn gerooteerd, of als ze zijn omgekeerd zodat het spiegelbeeld ontstaat. Daartoe zouden de elementen zelf eveneens gerooteerd en gespiegeld moeten worden. Dit zijn ingewikkelde operaties, waarvan wij hebben afgezien. In plaats daarvan laten wij bij de eerste fase van het leren het voorwerp en zijn spiegelbeeld in een groot aantal orientaties in één kwadrant zien. In *fig. 8* ziet men de vormelementen die op deze wijze van een voorwerp zijn verkregen. Dat ondanks het grote aanbod het aantal zo gering blijft, is te danken aan de beperking van het onderscheidend vermogen, waardoor bij draaiing meestal slechts vijf standen per kwadrant worden onder-

Fig. 7. Driemaal de analyse van een beeld in vormelementen. Rechts boven telkens de elementen die in het beeld links boven werden gevonden. De elementen (11 beeldpunten in diameter) zijn sterk vergroot ten opzichte van het beeld (100×100 beeldpunten). Links onder de reconstructie van het beeld met de gevonden elementen. Door een aparte achtergrondkleur is in de eerste foto het element 3 en in de beide andere het element 2 gemakkelijk terug te vinden, al of niet gerooteerd over 90° , 180° of 270° en al of niet geïnverteerd.

scheiden. Onze maatregel om waardeverschillen van randpunten niet te tellen, draagt krachtig bij tot de beperking van het aantal elementen. Het aantal elementen dat bij een poging tot herkenning met die van fig. 8 kan worden geïdentificeerd, is weer (bijna) acht maal zo groot.

De relatieve posities der vormelementen

Voordat wij nu de tweede fase van het leren en van het herkennen nader bezien, merken we op dat vormen óók herkend moeten worden als ze wel gelijkvormig maar niet even groot zijn, dit onder meer met het oog op gemakkelijk optredende verschillen in optische instelling bij het bekijken van leer- en herkenningsscènes. Dit punt hebben we tot nog toe buiten beschouwing gelaten, omdat het bij het herkennen van vormelementen een ondergeschikte rol speelt. Bij het herkennen van voorwerpen als geheel moet er echter van het begin af aan rekening mee worden gehouden.

Wij bespreken de tweede fase aan de hand van het voorbeeld van fig. 9. In fig. 9a ziet men de te leren vorm, in fig. 9b de te analyseren scène. Als resultaat van de eerste fase van het leren zijn onder meer de elementen 1, 2, 3, 4 en 5 van fig. 9c in het geheugen opgeslagen. Als resultaat van de eerste fase van het herkennen is in fig. 9b onder meer het element nr. 2 in de punten A, B, C, D, E, F, G, H, K geïdentificeerd, hetzij direct, hetzij in één van de geroteerde toestanden 2a, 2b of 2c.

Wij bespreken nu eerst de tweede fase van het herkennen en komen pas later terug op die van het leren, omdat daarbij ook het herkennen wordt gebruikt. We nemen dus even aan dat de tweede fase van het leren achter de rug is. Het resultaat daarvan is een aantal 'leerlijsten' waarop de posities van de vormelementen van de rechthoek zijn genoteerd voor elk van de vijf standen van fig. 9d. Zoals wij straks zullen zien, behoeven dit niet exact de standen te zijn waarbij resp. de elementen 1, 2, 3, 4 en 5 in de eerste fase zijn geleerd. Er bestaat dus onder meer een leerlijst L_2 met de vormelementen van de rechthoek in stand 2 van fig. 9d en hun posities:

L_2	vormelement	...	2	2a	2b	2c	
	positie	...	P	Q	R	S	

Behalve de vermelde vormelementen bevat L_2 nog rechte-rand-elementen, die op een groot aantal posities voorkomen als de rechthoek groot is. Het geheugen bevat voorts lijsten L_1 , L_3 , L_4 , L_5 voor de andere oriëntaties. Als het voorwerp niet spiegelsymmetrisch zou zijn, zouden er nog eens vijf lijsten van de gespiegelde vorm zijn.

Voor herkenning moet de rechthoek tot dekking kunnen worden gebracht met de rechthoek in fig. 9b, en om

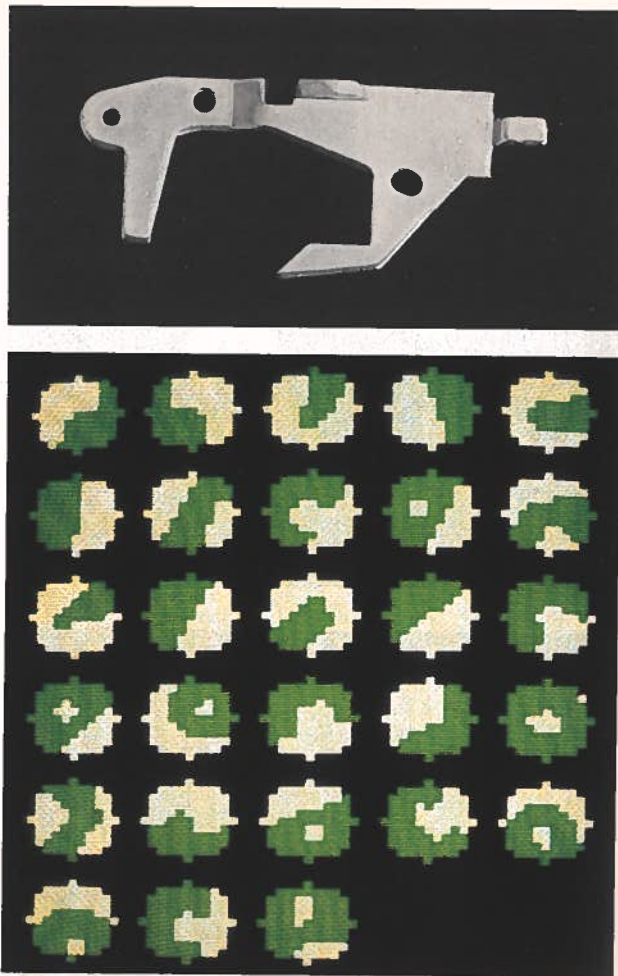


Fig. 8. Boven: een voorwerp dat in een groot aantal oriëntaties, ook omgekeerd ('gespiegeld'), ter lering werd aangeboden. Onder: de vormelementen die daarbij werden gevonden.

na te gaan of dit kan, gaan we vanzelfsprekend uit van de stand die vormelementen met de scène gemeen heeft, stand 2. Om echter dekking te verkrijgen, moet de rechthoek vanuit deze stand, behalve een translatie en een grootte-verandering, in het algemeen ook nog een rotatie over een kleine hoek ondergaan. Want hoewel er geen onderscheid is tussen de vormelementen der beide rechthoeken, kunnen na de beste translatie en grootte-verandering, de posities nog duidelijk merkbaar verschillen doordat de hoekdiscriminatie voor posities veel scherper is dan voor elementen (1 op 50, dus ca. 1° , voor vormen van 50 beeldpunten lengte). Doordat echter de rotatie tot een kleine hoek beperkt blijft, is de berekening eenvoudig. De algemene berekening voor rotaties,

$$\begin{aligned}x' &= x \cos \varphi - y \sin \varphi, \\y' &= x \sin \varphi + y \cos \varphi,\end{aligned}$$

kan nl. worden vervangen door

$$\begin{aligned}x' &= x - \varphi y, \\y' &= y + \varphi x.\end{aligned}$$

De poging tot herkenning verloopt in het gegeven voorbeeld nu verder als volgt. Bij de analyse van de scène van fig. 9b, worden de herkende vormelementen in een 'scènelijst' L_s genoteerd op de posities waar zij zijn gevonden:

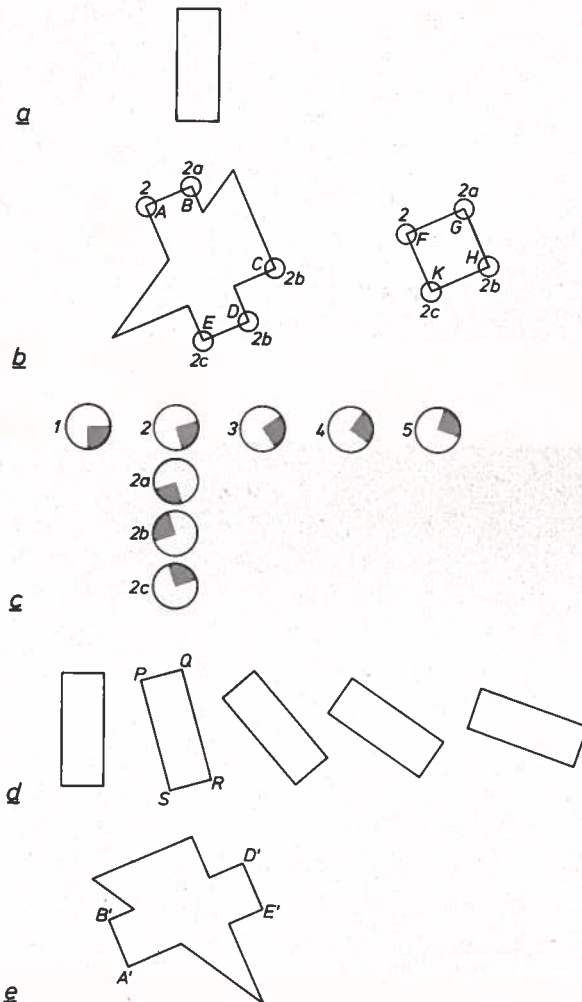


Fig. 9. De twee fasen van het leren kennen en van het herkennen, gedemonstreerd aan een eenvoudig voorbeeld. a) De op te sporen rechthoek. b) De scène die moet worden geanalyseerd. c) Enige van de vormelementen van de rechthoek die tijdens de eerste fase van het leren in het geheugen worden opgeslagen. Met behulp van nr. 2 kunnen ook de elementen 2a, 2b en 2c worden geïdentificeerd. In de scène (b) is dit gebeurd in A, B, C, D, E, F, G, H en K (eerste fase van het herkennen). d) Oriëntaties van de rechthoek waarvoor in de tweede fase van het leren de elementen met hun posities op 'leerlijsten' worden gezet. Zo ontstaat er b.v. een lijst L_2 die de items '2 in P', '2a in Q', ... bevat. In de tweede fase van het herkennen wordt van de scène in (b) een 'scènelijst' L_s opgemaakt: ..., '2 in A en F', '2a in B en G', ... Het systeem constateert dat L_s en L_2 (minstens) drie elementen (2, 2a en 2b) gemeen hebben, en zoekt naar transformaties die de bewuste elementen van L_2 in die van L_s over doen gaan. Dit lukt niet voor de elementen in A, B en C of in A, B en H, maar wel voor die in A, B en D. Als met de gevonden transformatie minstens de helft van alle elementen van L_2 (waaronder in dit voorbeeld vele rechte-rand-elementen) in elementen van L_s overgaan, is de rechthoek 'herkend'. e) Een scènelijst wordt effectief ook geconfronteerd met de 'over 90°, 180° en 270° geroteerde' leerlijsten. Zo wordt de rechthoek A'B'D'E' eveneens via stand 2 in (d) herkend.

L_s	positie		A	B	C	D	E	F	G	H	K	
	element		2	$2a$	$2b$	$2b$	$2c$	2	$2a$	$2b$	$2c$	

Het systeem zoekt nu naar gemeenschappelijke elementen van L_s en de leerlijsten. In het voorbeeld blijken L_s en L_2 een aantal elementen gemeenschappelijk te hebben. Daarvan kiest het systeem er drie op bepaalde plaatsen in de scène, b.v. 2 in A, 2a in B en 2b in C, en het onderzoekt of de corresponderende elementen van L_2 ermee tot dekking gebracht kunnen worden door een translatie, een rotatie en een vergroting of verkleining. Met de elementen van L_s in A, B en C lukt dat niet; een poging met A, B en H mislukt eveneens, maar die met A, B en D gelukt wel. Ons criterium voor 'dekking' van twee posities met de coördinaten x, y en x', y' is dat $|x - x'|$ en $|y - y'|$ beide kleiner zijn dan 2. Na dit succes worden de gevonden transformaties toegepast op alle elementen van L_2 . Als het percentage van deze elementen dat daardoor tot dekking komt met één van de elementen van L_s groter is dan 50%, geeft het systeem het signaal „herkend” door. De elementen van L_s die in het voorbeeld aldus worden bedekt, zijn de rechte-hoek-elementen in A, B, D, E en een groot aantal rechte-rand-elementen.

Voor het bepalen van de vereiste transformaties is het ene drietal elementen van L_2 geschikter dan het andere. De hoekelementen zijn b.v. geschikter dan rechte-rand-elementen, omdat zij karakteristieker zijn. Op de keuze van een 'geschikt' drietal gaan wij hier niet verder in.

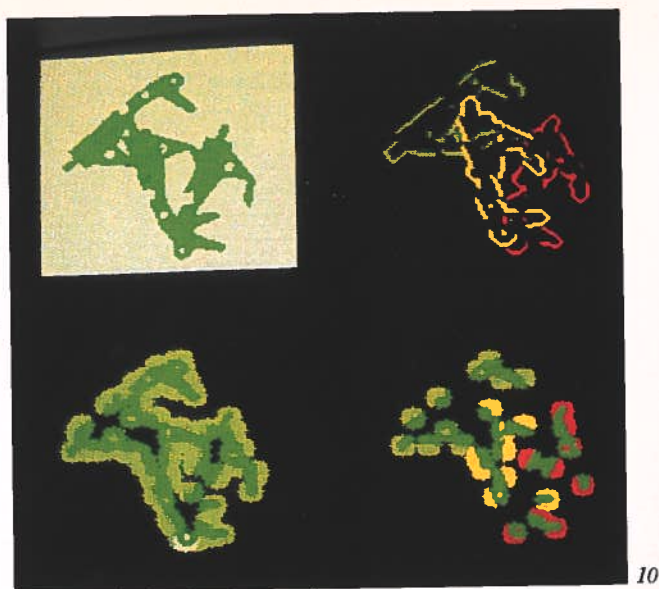
Rest ons nog te vermelden hoe de leerlijsten worden verkregen. Het te leren voorwerp wordt om te beginnen in één stand getoond. De vormelementen worden opgespoord, vergeleken met het in de eerste fase gevonden stel en, bij herkenning, met de bijbehorende coördinaten genoteerd. Dit levert de eerste lijst (L_1). Ter beperking van de vereiste geheugencapaciteit en rekentijd wordt de lijst beperkt tot een redelijk aantal, b.v. 15, min of meer gelijkelijk over de omtrek verdeelde elementen. Vervolgens wordt het voorwerp over een kleine hoek, b.v. ca. 2°, gedraaid en wordt nagegaan of het met lijst L_1 kan worden herkend. Zo ja, dan wordt het voorwerp verder gedraaid. Voor de eerste stand waarin de herkenning mislukt, wordt een nieuwe lijst gemaakt. Meestal is dit na omstreeks 20°, doordat de vormelementen zelf niet meer herkend worden. Zo krijgen we in één kwadrant dus meestal vijf lijsten, en nog eens vijf voor het spiegelbeeld.

De rechthoeken van fig. 9d liggen qua oriëntatie in het zelfde kwadrant als de rechthoek van fig. 9b. Voor het herkennen van vormen in de andere kwadranten wordt een scène effectief niet alleen geconfronteerd met de 10 bedoelde lijsten, maar zo nodig ook met de 'over 90°, 180° en 270° geroteerde' lijsten, die tijdens de con-

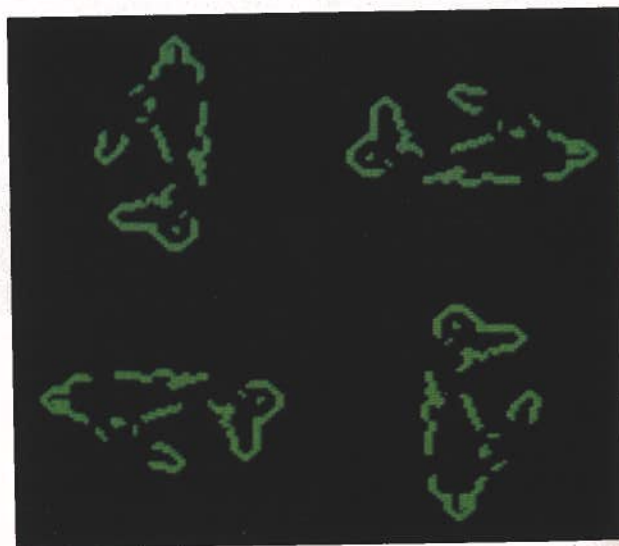
frontatie door een korte berekening uit de oorspronkelijke volgen. Zo zou ook de rechthoek van fig. 9e via stand 2 in fig. 9d worden herkend. Deze maatregel bespaart niet alleen geheugencapaciteit, maar verkort ook de leertijd in de tweede fase.

Fig. 10 geeft een meer realistisch voorbeeld van een herkenningsprobleem. Het systeem heeft de vorm van fig. 8 geleerd en krijgt de opdracht het beeld links boven te analyseren. In de eerste fase van het leren zijn de vormelementen van fig. 8 opgeslagen. In de tweede fase zijn telkens een aantal van deze en van de daaruit afgeleide elementen, met de bijbehorende coördinaten, op tien leerlijsten gezet. In fig. 11 ziet men een representatie van één der lijsten voor elk kwadrant; de figuurtjes bestaan uit 15 van de 28 elementen van fig. 8. Het beeld links onder in fig. 10 is een representatie van de scènelijst L_s : de elementen van de scène die met elementen in het geheugen werden geïdentificeerd, gereproduceerd ter plaatse van de identificatie. Rechts onder de elementen van L_s die, nadat de vereiste transformaties van een lijst op grond van drie elementen waren gevonden en toegepast, ook wat positie betreft met een leerlijst-element overeenstemden. Elke achtergrondkleur correspondeert met één lijst. Ten slotte rechts boven een visuele representatie van de drie gevonden lijsten, na de bedoelde transformaties. Daaruit blijkt dat het systeem de drie voorwerpen in de getoonde scène heeft herkend.

Voor eenvoudige vormen kan de rekentijd veel korter uitvallen dan voor ingewikkelde vormen. Een ronde schijf, bij voorbeeld, behoeft slechts in één oriëntatie te worden getoond en levert dus slechts één leerlijst op. Fig. 12, die op dezelfde wijze is samengesteld als fig. 10, geeft van een scène met plastic bolletjes een analyse die gebaseerd is op deze ene leerlijst.



10

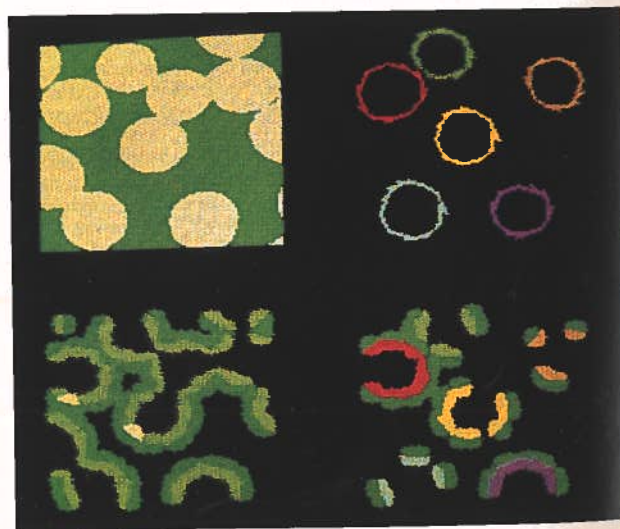


11

Fig. 10. Het zoeken van het voorwerp van fig. 8 in een scène. *Links boven*: de scène die moet worden onderzocht. *Links onder*: een visuele representatie van de scènelijst L_s : de elementen van fig. 8 die in de scène zijn herkend, ter plaatse van de herkenning. *Rechts onder*: de elementen van L_s die door de elementen van een leerlijst door transformatie konden worden bedekt. Elke achtergrondkleur correspondeert met één leerlijst. *Rechts boven*: visuele representatie van de herkende lijsten, na de bedoelde transformatie.

Fig. 11. Visuele representatie in vier kwadranten van één leerlijst (15 elementen) van het voorwerp van fig. 8.

Fig. 12. Zoeken en tellen van ronde schijven. De samenstelling van de figuur is geheel overeenkomstig die van fig. 10. Door de eenvoudige vorm van het gezochte voorwerp is er slechts één leerlijst.



12

Slotopmerkingen; verder onderzoek

Het bovenstaande geeft slechts in grote trekken de werking van het systeem weer; vele details zijn weggelaten. Zo worden b.v. de elementen die in de eerste fase van het leren worden gevonden, onder meer getoetst op de mogelijkheid ze terug te vinden, alvorens ze worden opgeslagen. Stoot men b.v. bij het opsporen 'bij toeval' op het element van *fig. 13a*, dan is de kans het terug te vinden niet zo groot, doordat de centrering tot een ander element leidt als men het iets anders benadert (*fig. 13b,c*). Door dergelijke toetsen levert de eerste fase een serie elementen die geschikt uitgangsmateriaal vormen voor de tweede fase. We stippen nog een ander detail aan. Om te voorkomen dat het systeem bij het analyseren van een scène een geleerd element 'rakelings' mist, wordt, als een opgespoord element niet

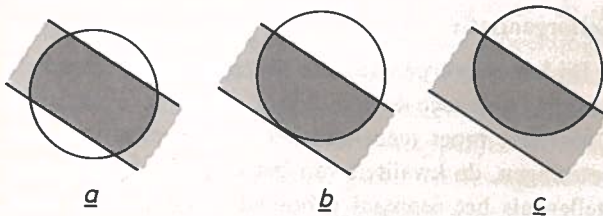


Fig. 13. De 'terugvindbaarheid' (retrievability) van vormelementen. Vindt men bij het leren 'bij toeval' het vormelement (*a*), dan vindt men het later vrijwel nooit meer terug doordat de centrering van de rand tot een ander element (*c*) leidt als men hetzelfde stukje scène iets anders benadert (*b*). Dergelijke moeilijk terug te vinden vormelementen worden niet in het geheugen opgenomen.

[3] Voor verdere details zie men E. H. J. Persoon, Principles for selforganisation and their application to picture recognition, Proc. 4th Int. Joint Conf. on Pattern recognition, Kyoto 1978, blz. 379-383.

wordt herkend, het venster ook op de acht rondom het centrum gelegen beeldpunten gecentreerd en worden de daar gevonden elementen achtereenvolgens met de geheugeninhoud vergeleken, tot eventueel herkenning plaatsvindt [3].

Het verdere onderzoek is er in de eerste plaats op gericht de bewerkingstijd te verkorten. Daartoe moet op speciale taken gerichte apparatuur worden gebouwd, die een aantal bewerkingen parallel uitvoert. Voor de eerste fase is dat al gelukt. Voorts onderzoeken we thans de mogelijkheid de besproken methode toe te passen op beelden die niet zwart-wit zijn, maar ook grijswaarden bevatten.

Samenvatting. Voor het machinaal herkennen van voorwerpen aan hun vorm in twee dimensies is een systeem ontworpen dat in twee fasen werkt, zowel bij het leren kennen van een voorwerp als bij het herkennen ervan in een getoonde scène. In de eerste fase worden 'vormelementen' opgespoord. Dit zijn beelden in een klein venster (diameter 11 beeldpunten), waarbij het zwaartepunt van de door het venster geselecteerde 'randpunten' in het centrum van het venster ligt. Vormelementen waarvan slechts randpunten in waarde verschillen, worden als 'gelijk' beschouwd. Bij het leren kennen wordt het voorwerp, ook omgekeerd, in een groot aantal oriëntaties getoond. Door de betrekkelijk kleine afmetingen van het venster is het onderscheidend vermogen tussen vormelementen beperkt, en daardoor is het aantal ervan, ook bij gecompliceerde vormen en ondanks het grote aanbod bij het leren, niet erg groot. In de tweede fase legt het systeem de positierelaties tussen de elementen vast op hoogstens ongeveer 10 'leerlijsten', vijf voor vijf oriëntaties van het voorwerp in elk kwadrant en hetzelfde aantal voor de gespiegelde vorm (omgekeerde voorwerp). Bij het herkennen wordt een 'scènelijst' opgemaakt van vormelementen die in de eerste fase zijn herkend, met hun posities. Als ongeveer de helft of meer van de elementen van één leerlijst met overeenkomstige elementen van de scènelijst door transformaties tot dekking kan worden gebracht, is het voorwerp herkend. Met deze methode kunnen voorwerpen óók worden herkend als zij andere voorwerpen raken of daar zelfs gedeeltelijk door worden bedekt.