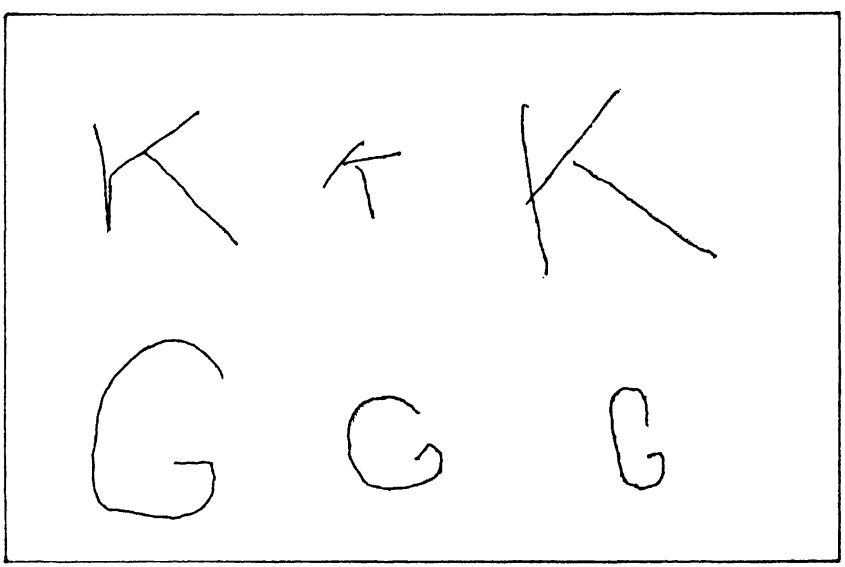
Psyke & Logos, 1988, *9*, 357-369

VISUEL GENKENDELSE VED TEMPLATE MATCHING

Axel Larsen og Claus Bundesen

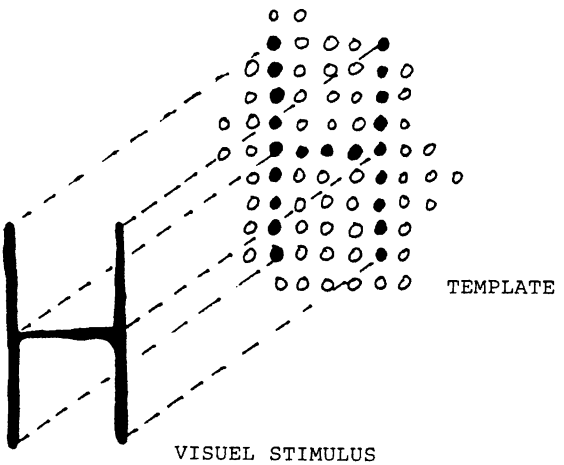
Der udvikles en beregningsmæssig model for visuel genkendelse ved sammenligning af synsindtryk med hukommelsesbilleder. I modellen antages genkendelse at være baseret på template matching i form af krydskorrelation. Positions-, størrelses- og orienteringsinvarians opnås ved aksetransformationer, mens støjproblemet tackles ved gaussisk filtrering. Modellen er implementeret som et computerprogram, hvis anvendelsesområde for øjeblikket omfatter genkendelse af enkeltvis præsenterede todimensionale mønstre såsom håndskrevne bogstaver.

Kig på de håndskrevne blokbogstaver i figur 1. Vi kan uden videre se, at der er tale om forskellige håndskrevne versioner af bogstaverne »K« og »G«. Hvordan bærer vi os ad? Den danske filosof og psykolog Harald Høffding (1889), som måske var den første, der behandlede genkendelsesproblemet under en fagpsykologisk synsvinkel, hævdede, at problemet består i at gøre rede for, hvorledes synsindtrykket af et objekt kommer i »kontakt« med et hukommelsesbillede af objektet.



Figur 1. Håndskrevne varianter af bogstaverne »K« og »G«.

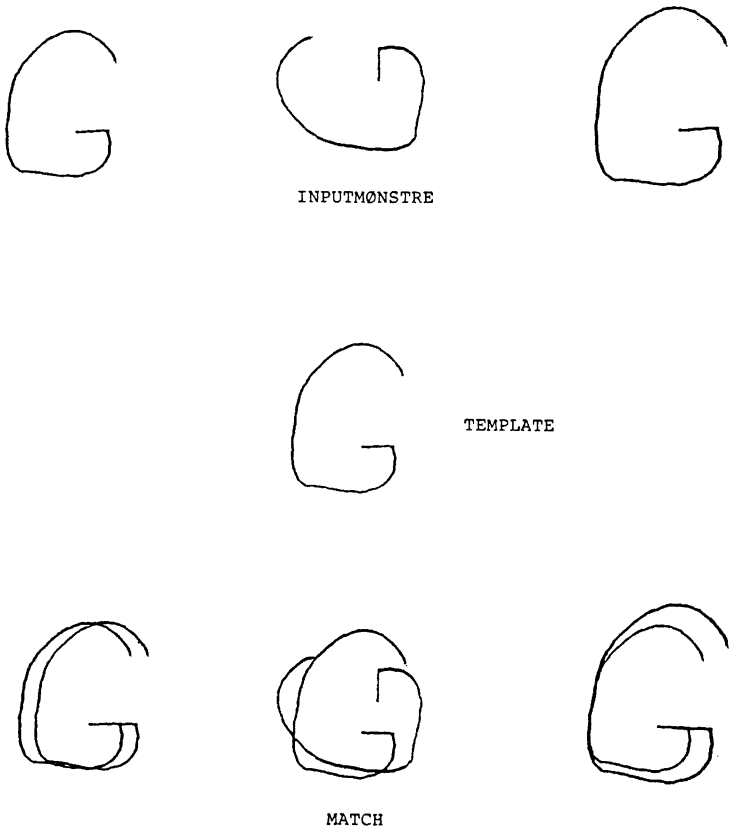
Høffdings antagelse, at visuel genkendelse af et objekt sker ved sammenligning af et synsindtryk af objektet med et tilsvarende hukommelsesbillede, har med tiden fået vid udbredelse (se f.eks. Neisser, 1967, om »the Höffding step«). Det er mere kontroversielt, hvilket format de involverede repræsentationer (synsindtryk og hukommelsesbilleder) har. Den simpleste hypotese synes at være, at både synsindtryk og hukommelsesbilleder er »billedmæssige« repræsentationer (jvf. Bundesen, 1986, pp. 4-5; Kosslyn, 1980), og at genkendelse således er baseret på sammenligning af billedmæssige repræsentationer (synsindtryk) med billedmæssige repræsentationer (hukommelsesbilleder). Hvis hukommelsesbillederne opfattes som en slags aftryk eller kopier af tidligere synsindtryk (se figur 2), beløber hypotesen sig til den antagelse, at genkendelse sker ved *template matching:* sammenligning af inputbilledet (synsindtrykket) med *templates* (hukommelsesbilleder) i form af kopier af tidligere inputbilleder med henblik på bestemmelse af graden af *match* (overlapning eller krydskorrelation).



Figur 2. Indkopiering af visuel stimulus som template ved aktivering af celler i en cellemosaik.

Det har længe været muligt at få maskiner til at genkende ensartede visuelle mønstre (f.eks. trykte bogstaver indlæst via et TV-kamera) ved hjælp af template matching (se f.eks. Duda & Hart, 1973; Uhr, 1973; Ullman, 1973). De simpleste template matching systemer har dog alvorlige begrænsninger. Som vist i figur 3 kommer de i vanskeligheder, blot inputbilledet afviger fra hukommelsesbilledet med hensyn til position, størrelse eller orientering. På grund af disse begrænsninger er template matching systemer blevet forkastet som seriøse modeller for visuel genkendelse hos mennesker. I årene omkring 1970 var forkastelsen næsten enstemmig blandt perceptionspsykologer

(se f.eks. Corcoran, 1971; Lindsay & Norman, 1972; Neisser, 1967; Reed, 1973). I denne artikel skal vi imidlertid forklare, hvorledes vanskelighederne synes at kunne overvindes ved enkle og plausible udbygninger af de simpleste systemer.



Figur 3. Template matching. Graden af overlapning eller match mellem inputmønster og template forringes alvorligt, hvis inputbilledet afviger fra hukommelsesbilledet med hensyn til position, størrelse eller orientering.

Forestillingsbilleder og billedtransformationer

Visuel genkendelse antages at kunne ske ved sammenligning af synsindtryk med repræsentationer lagret i en visuel korttidshukommelse (aktive *forestillingsbilleder)* så vel som ved sammenligning af synsindtryk med visuelle repræsentationer i langtidshukommelsen (egentlige *hukommelsesbilleder).* Genkendelse ved sammenligning af synsindtryk med forestillingsbilleder er

bedre undersøgt end genkendelse ved sammenligning af synsindtryk med hukommelsesbilleder, og de senere års undersøgelser har givet stærk støtte til den formodning, at visuelle forestillingsbilleder kan betragtes som *transformerbare templates* for synsindtryk af bestemte objekter (se Bundesen, 1986; Larsen, 1988; Shepard & Cooper, 1982).

Et eksempel på en transformation af et forestillingsbillede er en omdannelse af et forestillingsbillede af en genstand med en bestemt visuel størrelse og orientering i det tredimensionale rum til et forestillingsbillede af en genstand med samme form, men med en anden størrelse og i en ny orientering - en mental transformation af den repræsenterede genstands størrelse og orientering. Fænomenologisk synes det klart, at vi kan foretage sådanne transformationer af visuelle forestillingsbilleder, og forestillingsvirksomhed af denne art synes bl.a. at danne grundlag for visuel sammenligning af objekter. Bundesen, Larsen og Farrell (1981; se også Bundesen & Larsen, 1975; Larsen, 1985; Shepard & Metzler, 1971) fandt således reaktionstidsevidens for følgende perceptuelle procedure til sammenligning af objekter med hensyn til form, men uden hensyn til størrelse og orientering:

Har vi et synsindtryk af et objekt med en bestemt form, størrelse og orientering og et synsindtryk af et andet objekt med samme form og en given størrelse og orientering, kan vi identificere de to objekter som ensformede ved (a) at indkode det ene synsindtryk som et forestillingsbillede, ved (b) at transformere dette forestillingsbillede, så den repræsenterede størrelse og orientering kommer til at svare til det andet synsindtryks, og ved (c) at teste, hvorvidt dette synsindtryk stemmer overens med det omformede forestillingsbillede (matcher den transformerede template). På samme måde kan vi identificere forskelligt formede objekter som forskelligt formede.

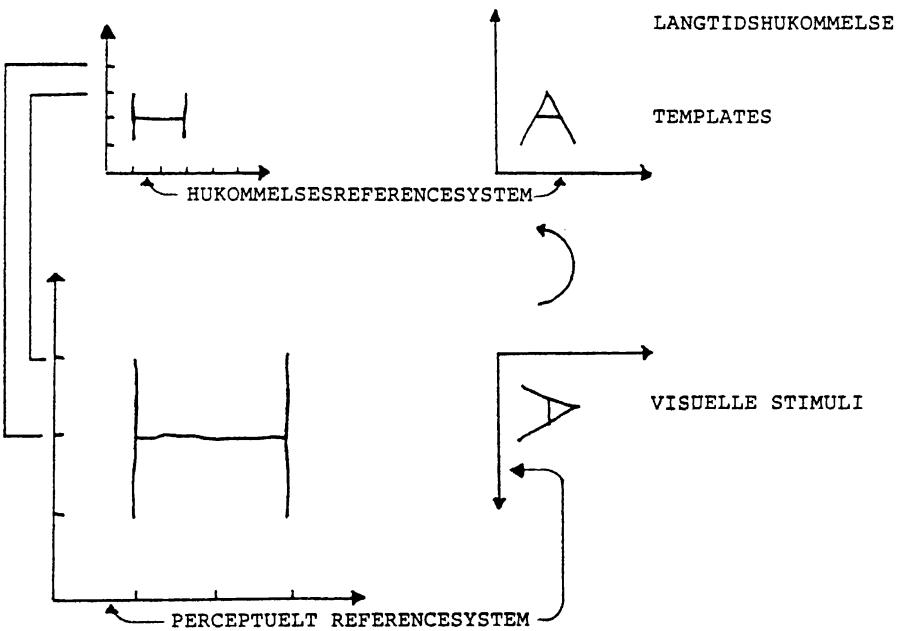
Hukommelsesbilleder og aksetransformationer

Genkendelse ved sammenligning af synsindtryk med hukommelsesbilleder (visuelle repræsentationer i langtidshukommelsen) er mindre velundersøgt end genkendelse ved sammenligning af synsindtryk med forestillingsbilleder (repræsentationer i visuel korttidshukommelse). Givet at forestillingsbilleder virker som templates for synsindtryk af bestemte objekter, er det imidlertid nærliggende at antage, at også genkendelse ved sammenligning af synsindtryk med hukommelsesbilleder kan forstås som template matching. De forholdsvis få relevante empiriske data, der foreligger, giver en vis støtte til denne antagelse. Til forklaring af resultaterne af en række reaktionstidsstudier foreslog Larsen og Bundesen (1978) således følgende teoretiske redegørelse for størrelsesinvarians ved genkendelse baseret på sammenligning af synsindtryk med hukommelsesbilleder:

Antag, at langtidshukommelsen rummer en mængde hukommelsesbilleder, som hver for sig specificerer et bestemt objekts geometri i relation til et

standardreferencesystem (et indre koordinatsystem, jvf. Marr, 1982). Genkendelse af stimulusobjekter repræsenteret i synsindtrykket ved sammenligning med disse hukommelsesbilleder forudsætter, at der etableres en korrespondance mellem positioner i hukommelsesreferencesystemet og positioner i det foreliggende synsfelt, eller med andre ord, at hukommelsesreferencesystemet interpreteres i synsfeltet. Kald den aktuelle interpretation af hukommelsesreferencesystemet i synsfeltet for det aktuelle *perceptuelle referencesystem,* og antag at det perceptuelle referencesystem er variabelt. Så kan invariant genkendelse under størrelsestransformation af et stimulusobjekt forklares ved størrelsestransformation af skalaen (akseenheden) i det perceptuelle referencesystem *(skalatransformation).*

Positions- og orienteringsinvarians kan forklares på lignende måde, d.v.s. ved at antage at det perceptuelle referencesystem kan tilpasses stimulusobjektet ved henholdsvis parallelforskydning og drejning af akserne. Proceduren forudsætter, at beregning af stimulusobjektets position, størrelse og orientering finder sted, *før* stimulusobjektet genkendes. Beregning af et objekts position og størrelse forud for genkendelse af objektet er relativt let at



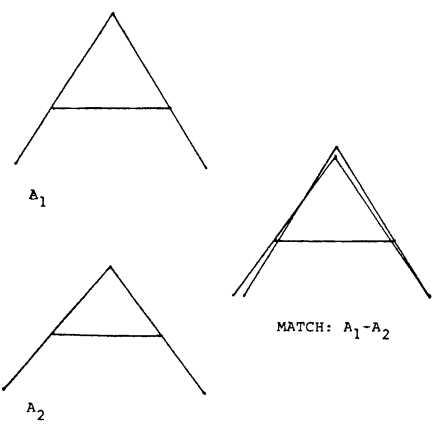
Figur 4. Størrelses- og orienteringsinvarians ved justering af perceptuelt referencesystem. Øverst vises to hukommelsesbilleder (»A« og »H«) i et hukommelsesreferencesystem. Nederst vises to inputmønstre (et størrelsestransformeret »H« og et drejet »A«). Ved at interpretere inputmønstrene i de angivne perceptuelle referencesystemer kan der etableres en passende korrespondance mellem visuelle inputmønstre og lagrede hukommelsesbilleder.

implementere. Beregning af objektets orientering kan ske forud for genkendelse af objektet, hvis den omhandlede »orientering« opfattes som en geometrisk defineret *indre retning* og ikke som f.eks. objektets afvigelse fra dets eventuelle »kanoniske« (d.v.s. sædvanlige) orientering i rummet (et objekts kanoniske retning kan almindeligvis først bestemmes, *efter* at objektet er blevet genkendt).

Vi skal senere se konkrete eksempler på, hvorledes de beskrevne beregninger kan realiseres. For nærværende er det afgørende at bemærke, at positions-, størrelses- og orienteringsinvarians kan opnås relativt enkelt ved at udbygge en templatebaseret genkendelsesmekanisme med et transformerbart perceptuelt referencesystem (koordinatsystem), og at et sådant skema (illustreret i figur 4) er en plausibel approksimation til en model for genkendelse ved sammenligning af synsindtryk med hukommelsesbilleder.

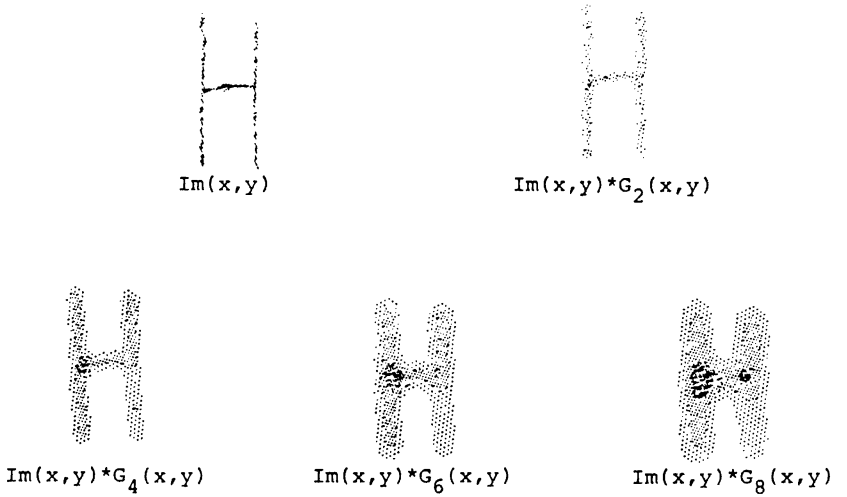
Støj og filtrering

Det er ikke tilstrækkeligt at kunne håndtere positions-, størrelses- og orienteringsinvarians. Figur 5 viser, hvordan et simpelt template matching system kommer i vanskeligheder i tilfælde, hvor det visuelle input (efter at være normaliseret med hensyn til position, størrelse og orientering) afviger helt bagatelagtigt fra det lagrede hukommelsesbillede: Hvis opløsningen af billeder i elementer (pixels) er rimelig høj, så ændres graden af match (overlapning) mellem to billeder drastisk ved ganske små ændringer i et af billederne - ændringer, som for det menneskelige øje er ligegyldige.



Figur 5. Template mismatch. Mønstre, der for det menneskelige øje er næsten ens, kan have ganske lille overlapning (grad af match) ved simpel template matching.

Et simpelt template matching system med en høj opløsningsgrad er altså yderst følsomt over for »støj« i form af ligegyldige småafvigelser mellem inputmønster og hukommelsesbillede. Paradoksalt nok synes støjproblemet at kunne løses ved systematisk at tilføre billederne endnu mere støj, f.eks. ved at filtrere billederne med et såkaldt gaussisk filter (den fra statistikken velkendte klokkeformede funktion). Resultatet af en sådan filtrering er illustreret i figur 6. Mens graden af match mellem to rå billeder ændres drastisk ved små ændringer i det ene af billederne, er graden af match mellem de tilsvarende filtrerede billeder stort set uafhængig af sådanne småændringer.



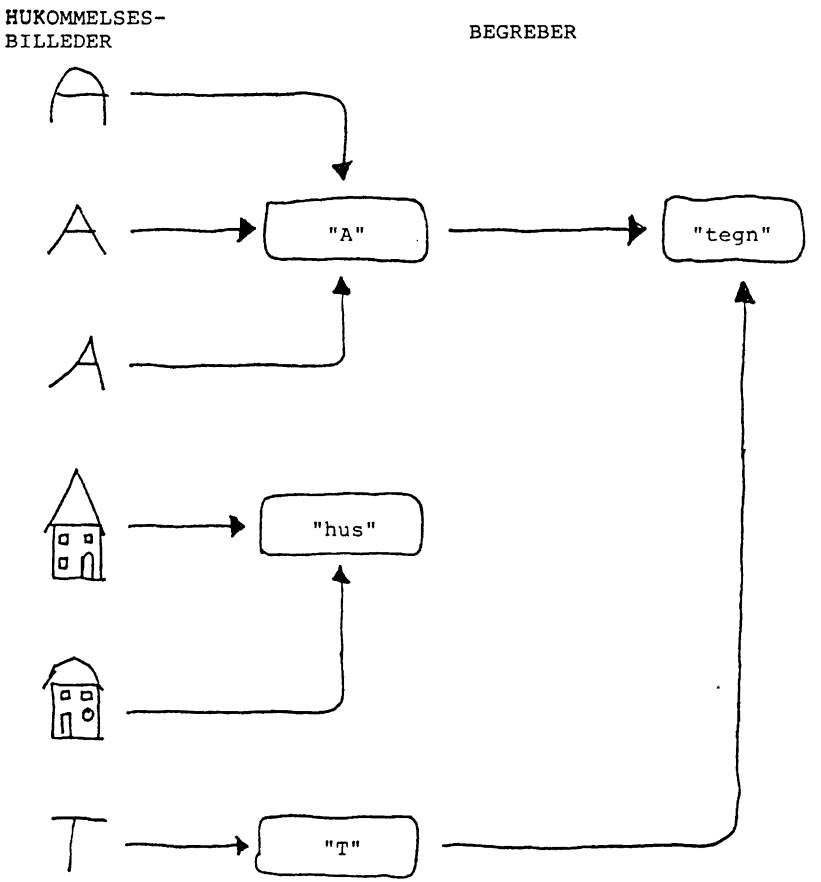
Figur 6. Gaussisk filtrering. Billedet Im *(x, y)* filtreres med en cirkulært symmetrisk todimensional Gaussfunktion G_i *(x, y)*, hvor *i* angiver spredningsparameteren δ (målt i pixels). Foruden det rå billede ses det filtrerede billede for $i = 2, 4, 6$ og 8.

Model for genkendelse ved sammenligning af synsindtryk med hukommelsesbilleder

I det følgende fremstilles en egentlig model for visuel genkendelse ved sammenligning af synsindtryk med hukommelsesbilleder. I modellen antages genkendelse at være baseret på template matching i form af krydskorrelation. Positions-, størrelses- og orienteringsinvarians opnås ved aksetransformationer, og støjproblemet tackles ved gaussisk filtrering. Modellen er for øjeblikket implementeret som et computerprogram, hvis anvendelsesområde er begrænset til genkendelse af enkeltvis præsenterede todimensionale mønstre (såsom håndskrevne bogstaver).

Efter først normalisering ved aksetransformationer (omparametrisering) og dernæst gaussisk filtrering antages et inputbillede at kunne lagres perma-

nent i langtidshukommelsen med en pil (»pointer«) hen til en symbolsk eller begrebsmæssig repræsentation, der også er lagret permanent. Den begrebsmæssige repræsentation angiver, at det korresponderende hukommelsesbillede repræsenterer et »A«, en »flaske«, en »giraf«, eller hvad der nu kan være tale om.



Figur 7. Et udsnit af langtidshukommelsens struktur. Billedmæssige repræsentationer peger på begrebsmæssige repræsentationer.

Figur 7 viser den foreslåede hukommelsesstruktur grafisk. Den rummer to fundamentalt forskellige typer af indre repræsentationer: billedmæssige repræsentationer og begrebsmæssige repræsentationer. Som vist på figuren antages disse som hovedregel at være organiseret således, at flere billedmæssige repræsentationer peger på en og samme begrebsmæssige repræsenta-

tion. Strukturen kunne udvides ved at lade begrebsmæssige repræsentationer pege på andre begrebsmæssige repræsentationer i et semantisk netværk, men en sådan udvidelse falder uden for rammerne af denne artikel.

Genkendelse antages at finde sted ved, at en visuel stimulus efter de beskrevne normaliseringsoperationer sammenlignes parallelt med alle hukommelsesbilleder i langtidshukommelsen. Det hukommelsesbillede, som giver den højeste grad af match, bestemmer, hvorledes den præsenterede stimulus indordnes begrebsmæssigt, d.v.s. genkendes.

Algoritme

Den generelle genkendelsesmodel må specificeres. I dette afsnit beskriver vi en selvlærende genkendelsesalgoritme i kvasi-programmeringsmæssige termer. Algoritmen er for øjeblikket implementeret på en VAX datamat, men den skal betragtes som et bud på den type af algoritmer, der er realiseret i vor hjerne.

1. Sensorisk registrering af visuel stimulus

Stimulus aktiverer cellerne i en $N \times N$ receptorflade, hvorved billedfunktionen Im (x, y) frembringes (jvf. figur 8, panel 1).

2. Beregning af tyngdepunkt

Lad Im' (x, y) være en billedfunktion, der har værdien 1, hvis Im (x, y) overstiger en vis tærskelværdi, og som ellers har værdien 0. Tyngdepunktets koordinater (C_x, C_y) er da givet ved

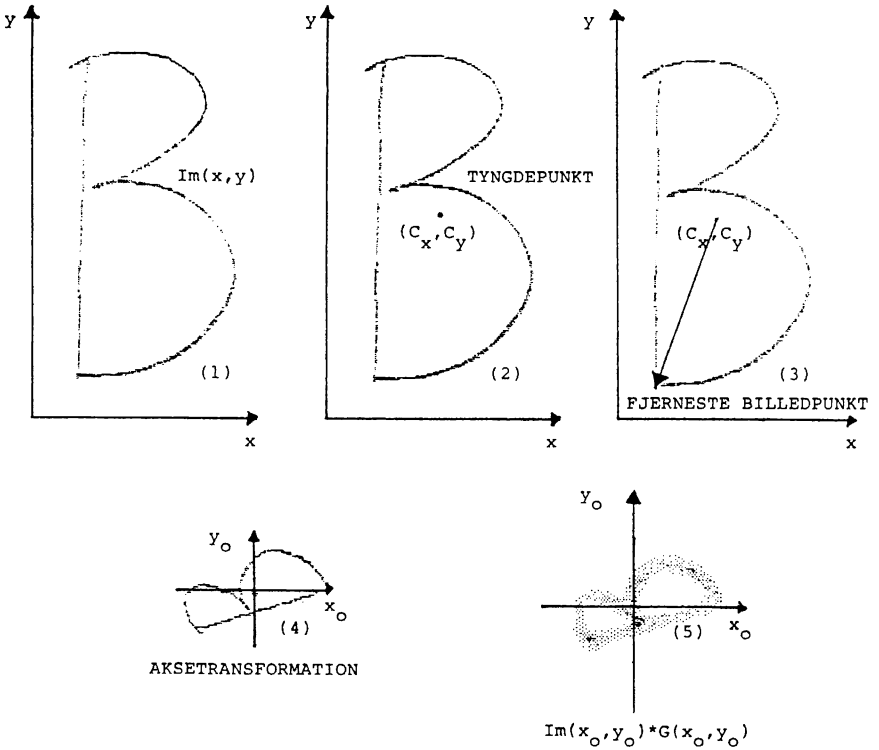
$$C_x = \Sigma\Sigma\, [x * \text{Im}'(x, y)] / \Sigma\Sigma\, \text{Im}'\, (x, y)$$

$$C_y = \Sigma\Sigma\, [y * \text{Im}'(x, y)] / \Sigma\Sigma\, \text{Im}'\, (x, y),$$

hvor $\Sigma\Sigma$ angiver summation over hele receptorfladen (jvf. figur 8, panel 2).

3. Beregning af størrelse

Størrelsen beregnes som afstanden fra tyngdepunktet til det fjerneste af de billedpunkter Im' (x, y), der har værdien 1 (jvf. figur 8, panel 3).



Figur 8. Forbehandling af inputbillede. Panel 1: Billedfunktionen Im *(x, y)* af et håndskrevet »B«. Panel 2: Markering af det beregnede tyngdepunkt (C_x, C_y). Panel 3: Beregnet indre retning og størrelse angivet ved en vektor med udgangspunkt i tyngdepunktet. Panel 4: Omparametrisering af billedfunktionen ved aksetransformation. Panel 5: Filtrering med Gaussfunktion.

5. *Aksetransformation*

Billedfunktionen Im *(x, y)* omparametriseres ved parallelforskydning, drejning og skalatransformation af *xy*-koordinatsystemet, så at origo placeres i det beregnede tyngdepunkt, *x*-aksen bliver parallel med den beregnede indre retning, og måleenheden på akserne (skalaen) bliver lig med den beregnede størrelse (jvf. figur 8, panel 4).

4. Beregning af indre retning

Den indre retning (figur 8, panel 3) beregnes som retningen af den vektor, der går fra tyngdepunktet til det fjerneste af de billedpunkter Im' *(x, y)*, som har værdien 1. (Hvis mængden *M* af de billedpunkter, der ligger længst væk fra tyngdepunktet, har mere end eet medlem, beregnes den indre retning som summen af de vektorer, der går fra tyngdepunktet til et medlem af *M*. Hvis sumvektoren er nul, tilordnes en indre retning ved tilfældigt træk fra mængden af de vektorer, der går fra tyngdepunktet til et medlem af *M*.)

6. Filtrering

Det omparametriserede billede Im *(x, y)* filtreres (figur 8, panel 5) ved foldning med en cirkulært symmetrisk todimensional Gaussfunktion af formen

$$G(x, y) = (1/(2\pi\delta^2)) \exp(-(x^2 + y^2)/2\delta^2).$$

7. Tabula rasa test

Gå til 11, hvis langtidshukommelsen LTM er tom.

8. Match

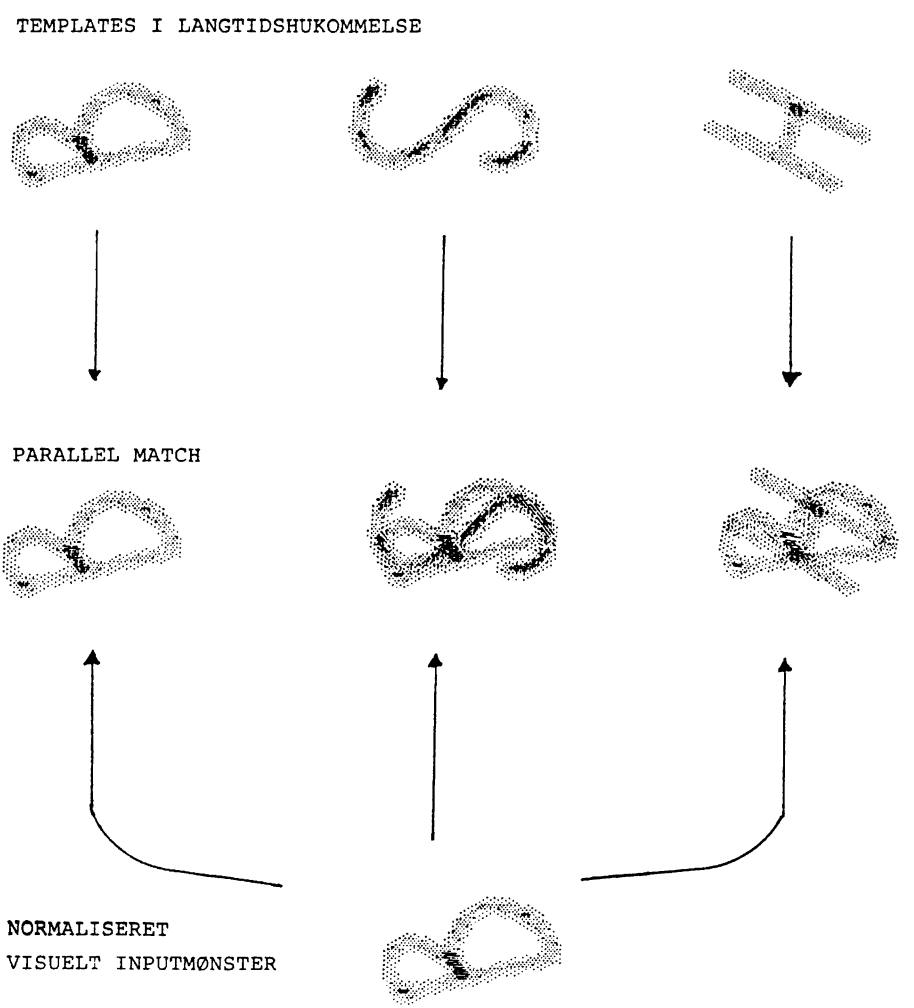
For hvert hukommelsesbillede H *(x, y)* i LTM beregnes korrelationskoefficienten mellem H *(x, y)* og det filtrerede billede Im $(x, y) * G(x, y)$ (se figur 9).

9. Genkendelse

Den givne stimulus klassificeres under den begrebsmæssige kode, der hører til det hukommelsesbillede, for hvilket den beregnede korrelationskoefficient er størst.

10. Feedback

Gå til 1, med mindre en ekstern vejleder korrigerer den foretagne klassificering og angiver en anden begrebsmæssig repræsentation for inputbilledet.



Figur 9. Sammenligning (korrelation) mellem forbehandlet inputbillede og lagrede hukommelsesbilleder. I modellen antages sammenligningerne at foregå parallelt.

11. Indlæring

Indlæg det filtrerede billede Im $(x, y) * G(x, y)$ som hukommelsesbillede i LTM med en pil til den af vejlederen angivne korrekte begrebsmæssige repræsentation. Gå dernæst til 1.

Afsluttende bemærkninger

Den beskrevne model for genkendelse ved sammenligning af synsindtryk med hukommelsesbilleder er for øjeblikket under afprøvning. Foreløbige resultater synes at vise, at modellen faktisk fungerer: Systemet er tilsyneladende i stand til at lære at genkende todimensionale mønstre såsom håndskrevne bogstaver, uden at langtidshukommelsen overbelastes, og de fejl, systemet begår i løbet af indlæringsfasen, ligner de fejl, som mennesker begår (f.eks. forveksling af »C« og »G«). En mere præcis, kvantitativ redegørelse for systemets effektivitet og en mere formel sammenligning mellem de fejl, systemet begår, og de fejl, mennesker begår, kan snart forventes.

LITTERATUR

BUNDESEN, C. (1986). *Studier af visuel informationsbehandling: Sammenfattende redegørelse.* (Psykologisk Forskningsrapport nr. 5). København: Københavns Universitet.

BUNDESEN, C. & LARSEN, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 214-220.

BUNDESEN, C., LARSEN, A. & FARRELL, J.E. (1981). Mental transformations of size and orientation. I J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 279-294). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

CORCORAN, D.W.J. (1971). *Pattern recognition.* Harmondsworth, England: Penguin.

DUDA, R.O. & HART, P.E. (1973). *Pattern classification and scene analysis.* New York: Wiley.

HØFFDING, H. (1889). Umiddelbar Genkendelse. *Videnskabelig Selskabs Skrifter, 6. række, III, 1.* København.

KOSSLYN, S.M. (1980). *Image and mind.* Cambridge, Massachusetts: Harvard University Press.

LARSEN, A. (1985). Pattern matching: Effects of size ratio, angular difference in orientation, and familiarity. *Perception & Psychophysics, 38,* 63-68.

LARSEN, A. (1988). *Eksperimentelle undersøgelser over visuel skinbevægelse, identifikation, klassifikation og selektion.* (Psykologisk Forskningsrapport nr. 8). København: Københavns Universitet.

LARSEN, A. & BUNDESEN, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 1-20.

LINDSAY, P.H. & NORMAN, D.A. (1972). *Human information processing.* New York: Academic Press.

MARR, D. (1982). *Vision.* San Francisco: Freeman.

NEISSER, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

REED, S.K. (1973). *Psychological processes in pattern recognition.* New York: Academic Press.

SHEPARD, R.N. & COOPER, L.A. (1982). *Mental images and their transformations.* Cambridge, Massachusetts: MIT Press.

SHEPARD, R.N. & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701-703.

UHR, L. (1973). *Pattern recognition, learning & thought.* Englewood Cliffs, New Jersey: Prentice-Hall.

ULLMAN, J.R. (1973). *Pattern recognition techniques.* London: Butterworths.