

RAČUNALNIŠKI VID: OBRAVNAVA VIZUALNIH INFORMACIJ IN ANALIZA SLIK

G. KANDUS
M. ŠPEGEL

UDK: 681.3 : 007.52

INSTITUT JOŽEF STEFAN, LJUBLJANA

V članku je omenjenih nekaj algoritmov in metod ter podan pregled teorije računalniškega vida. Poudarek je na procesiranju vizualne informacije in na analizi scene. Še prav je računalniški vid samostojno področje, so v zadnjih letih pospešeno razvijajo predvsem kot najsplošnejši senzor inteligentnih robotskih sistemov.

COMPUTER VISION: In the article some algorithms and methods are mentioned and an overview of computer vision theory is given. The emphasis is on visual information processing and on scene analysis. Though the computer vision is an independent field, its fast progress in last few years is made as most general sensor for intelligent robot systems.

1. Uvod

Pri človeku je gledanje neposredna iskušnja, ki ji ni mogoče priti do konca s še tako skrbno introspekcijo. Med informacije, ki jo sprejemajo naša oči in podoba, ki si jo na podlagi nje izoblikujejo možgani, je še precej korakov zavrtih v možgi. Nevrofiziologija ve marsikaj o individualni živčni celici, a le malo o tem, kako celice kontaktirajo med seboj pri prenašanju, spreminjanju in zapisovanju informacije, kar je bistveno za razumevanje problemov gledanja. Biološki sistem gledanja je prekompleksen, da bi ga bilo mogoče simulirati.

Računalniški vid je trd oreh. V literaturi je mogoče zaslediti vrsto člankov, ki so problemsko orientirani in opisujejo različne algoritme, metode in tehnike, ki so uporabne ob določenih predpostavkah, ni pa mogoče najti dela, ki bi nudil celovit pregled in analizo še znanih postopkov ali pa skušal utemeljiti teorijo računalniškega vida.

Naloga sistema za računalniško gledanje je, da opiše tisto kar vidi. Vhod je ona ali več slik, ki jih določajo dvodimenzionalna polja vrednosti (običajno intenziteta svetlobe, včasih barva), ki nam jih posredujejo senzorji (običajno TV kamera) iz območja zornega kota (scene). Izhod je običajno v simbolni obliki. Održati mora neki aspekt tri dimenzionalne realnosti in biti pripravljen za izvajanje določenih nalog.

Področja, ki se deloma prekrivajo s področjem računalniškega gledanja so:

- procesiranje slik
- prepoznavanje vzorcev
- analiza scene.

Kratek opis teh področij:

Procesiranje slik je spreminjanje in bogatenje slikovne vsebine. Je odino od zgoraj omenjenih področij, ki ima sliko za vhod. Zal pa ima sliko tudi za izhod. Metode, ki jih procesiranje slik uporablja kot svoja glavna orodja (linearno, prostorno invariantno metodo in transformacijsko

tehnike), igrajo pri računalniškem gledanju le omejeno vlogo.

Bistvena metoda, ki jo uporablja prepoznavanje vzorcev je klasifikacija vzorcev, ki prestikujejo vektorje v celo števila. Komponente vektorjev predstavlja množico vrednosti, celo števila pa označujejo razrede, katero množico produkti pripadajo. Za računalniško gledanje so zanimive tehnike, ki se uporabljajo za izračun numeričnih vrednosti značilnosti, izločenih iz slikovnih podatkov.

Analiza scene spominja in dopolnjuje opis slik v bolj zapletene in uporabne opise. Ena od prvih nalog tega področja je bilo iz prvotne slike dobiti risbo, sestavljeno iz črt mojih ploskev predmetov. Da bi poenostavili komplicirano analizo prvotne slike so raziskovalci pričeli uporabljati metode, ki izkoriščajo pridobeno znanje o sceni. Na ta način so jo mogoče dokopali do ustrezni opisov tudi na podlagi raznorodno nepopolnih ali motenih podatkov. Paziti pa je treba, da ne opisujemo tistega, kar želimo videti ali kar pričakujemo, namisto dejanske scene.

2. Procesiranje vizualne informacije

Pri vseh treh prej omenjenih področjih pogoščemo metode, ki bi podatke o intenzitoti prvotne slike prevedle v simbolni opis. Problem je tudi, v kateri fazi procesa naj se zgodi ta transformacija. Nevarje raziskavo sugerirajo prehod na simbolni opis v začetnem delu procesiranja vizualne informacije. Opis sestavlja relativno velike število osnovnih podatkov, iz katerih izhaja vsa poznejša analiza.

2.1 Izračun osnovne slike

Število skoraj tako dobro zaznava predmete, ki jih vidi na fotografiji, kot so jih vidi narisano, šep prav je informacije, ki v obliki intenzitete svetlobe pade na očesno mrežnico

v obeh primerih bistveno različna. To nas napeljuje na misel, da risarjevi simboli odgovarjajo simbolom, ki si jih človek izračuna pri tem, ko si skuša razložiti sliko. Torej ne izgubimo mnogo informacije, če vrednosti, ki ustrezajo intenziteti slike transformiramo v opis, kako se te vrednosti spreminjajo na robu približno homogenih območij. S tem zelo zmanjšamo količino podatkov potrebnih pri nadaljni analizi slike ob ne bistvenem zmanjšanju množine informacije.

Nekaj osnovnih pojmov, ki sestavljajo simbolni opis:

- rob
- linija
- okroglina

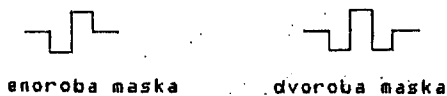
Njihovi parametri pa so:

- orientacija
- velikost
- kontrast
- pozicija
- zabrisanost

če odkrijemo na sliki rob predmeta, ga zapišemo v simbolni obliki na primer takole:

(ROB(POZICIJA(34 48)(73 48))
(KONTRAST 34)
(NEJASNOST 17)
(ORIENTACIJA 0))

Herskovitz in Binford (1970) sta empirično ugotovila, da so stopnice, izbokline in strehasti profili najpogostejše oblike spremembe intenzitete svetlobe, ki prihaja s scene, ki je sestavljena iz poliedrov. Znano je, da je človeški vid občutljiv ne le na spremembo, ampak tudi na hitrost spremembe intenzitete. Približke prvemu in drugemu odvodu intenzitete slike lahko dobimo s konvolucijo med lokalnimi vrednosmi intenzitete ter enorobo masko in dvorobo masko v obliki palice. Enoroba maska meri gradient intenzitetnega polja v določeni smeri; dvoroba maska pa lokalne spremembe gradienta intenzitete (Slika 1).



Slika 1

Algoritem za izračun simbolnega opisa lahko strnemo v naslednje štiri korake:

- poišči in primerjaj med seboj vrhove dobljene s konvolucijo in maskami različnih širin
- določi značilne vrhove
- grupiraj bližnje vrhove
- iz konfiguracije vrhov, ki so posledica sprememb intenzitete, določi tipe osnovnih pojmov (rob, linija, okroglina) predstavljajo ter njihove parametre (kontrast, pozicija, zabrisanost, itd).

Tako dobljeno skico imenujemo izvirno osnovno skico. Popolno osnovno skico dobimo z grupiranjem pojmov. Grupiramo ne le enake pojme, ampak tudi pojme, ki se po legi in orientaciji ujemajo s sosednjimi elementi. Kratke segmente tretiramo kot šum, če so šibkejši in jih upoštevamo, če so močnejši od daljših segmentov, ki jih sekajo. Zelo kratke samostojne segmente ignoriramo. Včasih nam pri presoji ali naj vključimo segment v skico ali ne, pomagajo abstraktno zanižljene linije. Pričakovati je namreč, da tudi kontne točke linij, čeprav nepovezane, leže na gladkih krivuljah.

2.2 Ločevanje predmetov od ozadja

Naslednji računalniški problem pri procesiranju vizualne informacije je segmentacija. Na tej stopnji se pojavi problem, kako ločiti predmet od ozadja ali predmete med seboj. Ločevanje se običajno izvede z diferenciacijo prvega reda ali z grupirnimi operacijami elementov osnovne skice.

Linijski zbir pomeni grupiranje površinskih elementov, tako da se orientacija grupe ujema z orientacijo elementov. To je v bistvu eno dimenzionalno grupiranje, ki je zelo važno, kajti tudi večja območja se najlažje poiščejo z določenjem njihovih meja.

Theta zbir se razlikuje od linijskega po tem, da se notranja orientacija elementov, ki jih grupiramo, razlikuje od orientacije grupe kot celote.

Elemente osnovne skice lahko grupiramo tudi na podlagi statističnih zakonitosti, če zberemo naslednje podatke:

- celotna dolžina črt, število okroglin, število različnih kontrastov
- celotno število elementov in celotna dolžina črt za posamezne orientacije
- porazdelitev parametrov velikosti
- porazdelitev kontrastnih parametrov.

če se izkaže, da neka lastnost prevladuje, je to dodatna informacija o obliki in orientaciji grupe, če predno ugotovimo točne obrise območja.

Nekateri avtorji (Bacjy (1972)) so uporabljali za ločevanje Fourierjevo analizo. Nekajkrat je bilo mogoče s pomočjo močnostnega spektra ločiti območja z različnimi strukturami. Drugi avtorji so uporabljali razne diferencialne operatorje. Eden najstarejših je Robertsov gradient (Roberts (1963)).

2.3 Uporaba rezultatov začetne faze procesiranja

a) Stereo tehnika:

Na razpolago imamo dve sliki scene, posneti iz dveh bližnjih točk na isti vodoravni višini. Sliki se rahlo razlikujeta, toda vsaki točki na eni sliki odgovarja točka na drugi sliki. Iz relativnih razlik, ki jih izračunamo, je mogoče uoločiti globino, to je razdaljo opazovane točke od točke opazovanja.

Stereo problem se tako prevede na primerjavo dveh prvotnih opisov. Pogoja, katerima mora biti zadoščeno, sta:

- enkratnost: vsaki točki na eni sliki ustreza natanko ena točka na drugi sliki
- zveznost: globina predmetov se zvezno spreminja.

Marr, Poggijo, Palm (1977) razvijajo metodo za izračunavanje lokalnih razlik para stereogramov, znano pod imenom kooperativni algoritem.

b) Ugotavljanje oblike predmetov s pomočjo gibanja

če se predmet giblje glede na opazovalca, lahko iz razlik na zaporednih posnetkih dobimo informacijo o njegovi obliki. Ullman (1979) razdeli problem na:

- primerjavo elementov na časovno zaporednih posnetkih

- ugotavljanje oblik predmetov s pomočjo neritov o spremembi njihovega položaja

Ullman postavi teoreme, ki pravi, da lahko določimo lego togega telesa v prostoru, če poznamo lego štirih njegovih točk, ki ne leže na isti ravnini, če jih opazujemo iz treh različnih smeri. Iz teorema izpelje metodo za izračun oblike gibajočih se predmetov, ki daje rezultate za katere trdi, da prekašajo človeško zmogljivost določanja oblike s pomočjo podatkov o gibanju teles.

2.4 Nivojni opisov

Za območno orientirane predstavitve ločimo naslednje nivoje opisov:

- pixel: slikovni element
- krpa: skupina bližnjih slikovnih elementov, ki imajo podobne lastnosti
- območje: smiselna skupina krp, ki odgovarja istemu predmetu
- podslika: del slike, ki odgovarja predmetu ali množici predmetov
- slika: predmeti in njihovo ozadje

Linjsko orientirane opise dobimo, če zamenjamo

- krpo z delom linije
- območje z linijo
- podsliko z množico linij, ki odgovarjajo predmetu

Območno orientirani opisi so primernejši za opis zaobljenih predmetov, medtem ko so linjsko orientirani opisi primernejši za opis planarnih predmetov.

Grobo lahko trdimo, da se nižji nivoji ukvarjajo s konkretnimi podatki, medtem ko se višji nivoji ukvarjajo s simboličnimi abstrakcijami. Podatkov na spodnjih nivojih je ogromna množina, če pri sceni, ki jo analiziramo, uporabimo določeno znanje, ki je bodisi splošno in opisuje splošno znana dejstva bodisi posebno in opisuje, kar vemo o sceni, lahko občutno zmanjšamo količino podatkov za opis predmetov pri prehajanju iz nižjih na višje nivoje. S tem zelo zmanjšamo potrebne računalniške kapacitete in pridobimo na času računanja, kar je pomembno, ker morajo biti rezultati pri računalniškem gledanju čisto dobljeni v realnem času.

3. Analiza scene

Raziskovanje analize scene je potekalo v dveh smereh. Zaradi nerazvitih matematičnih orodij so se najprej pojavile analize onostavnih scen, sestavljenih iz pravih geometrijskih teles, v večini primerov iz poliedrov. Pionirji na tem področju so bili Guzman (1968) in Clowes (1971), Huffman (1971) in kasneje Waltz (1976).

Simbolni opis scene je bil podan kot črna risba (projekcija) teles v splošni legi. Rezultat analize scene pa je bila lista, iz katere je bila vidna pripadnost površin črne slike posameznim predmetom scene.

Iz želje po analizi realnih scen, za katere poliedrska aproksimacija ne zadošča, se je razvil opis s tako imenovanimi posplošenimi valji, na katere je prvi opozoril Binford (1971).

Marr (1979) v svojem preglednem članku ne omenja analize scene, ampak govori o končni fazi procesiranja vizualne informacije.

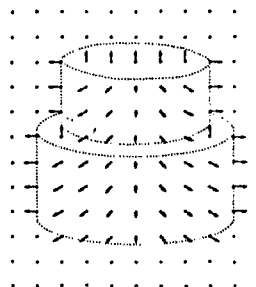
Črtni sliki odgovarja dvo in pol dimenzionalna skica; opisu teles pa tri dimenzionalna predstavitev.

Končna faza procesiranja vizualne informacije ima nalogo, da s pomočjo informacije o legi in položaju površin, katerih opis je odvisen od mesta opazovanja, pridemo na opis, ki ni odvisen od opazovališča in je z njegovo pomočjo mogoče prepoznati oblike predmetov iz različnih oddaljenosti in pod različnimi koti. Središče koordinatnega sistema se pri tem premakne od opazovalca k predmetu.

Lahko je opaziti, da vsebujejo slike velik del informacije o tridimenzionalni naravi njenih površin. Horn (1975) pokaže, kako je mogoče razviti diferencialne enačbe o osvetljenosti slike tako, da lahko iz gradienta osenčenja ugotovimo obliko predmetovih površin, če poznamo njihovo odbojno funkcijo.

3.1 Dvo in pol dimenzionalna skica

Velikokrat je koristno upoštevati podatke dobljene iz več virov hkrati: iz stereo slik, iz analize gibanja, barve in osenčenja ter jih kombinirati. Predstavitev, ki eksplicitno kaže obliko in orientacijo površin imenujemo dvo in pol dimenzionalna skica (Slika 2). Orientacijo površin predstavljajo puščice, ki pokrivajo sliko. Dolžina puščice je sorazmerna nagljenosti ravnine v tisti točki glede na os opazovanja.



Dvo in pol dimenzionalna skica

Slika 2

Cilj te faze procesiranja vizualne informacije je odkriti, kakšne so orientacije površin na sliki, kateri obrisi v osnovni skici odgovarjajo nezveznostim površin v dvo in pol dimenzionalni skici in kateri obrisi manjkajo v osnovni skici in jih je treba vstaviti v dvo in pol dimenzionalno skico tako, da je zadoščeno zakonom tri dimenzionalnega prostora.

3.2 Tri dimenzionalna predstavitev

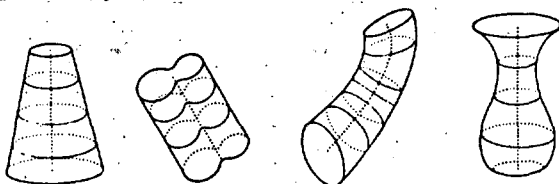
Pri tri dimenzionalni predstavitvi predmetov nas zanima predvsem:

- masno težišče
- površina, volumen in razsežnost
- rotacijske in zrcalne simetrijske osi.

Zato je ugodno, če je opis predmetov prostorski ter v koordinatnem sistemu, ki jo za opisa naraven. Tem zahtevam se približamo z opisom s posplošenimi valji (Slika 3).

Posplošeni valj (Binford 1971) je ploskev, ki jo opiše zaključena ravninska krivulja pri tem, ko ravnina te krivulje potuje vzdolž neke osi. Krivulja ni nujno, da je krog in os ni nujno, da je pravica. Krivulja

se lahko vzdolž osi spreminja; vendar tako da ohranja podobnost.



Nekaj primerov posplošenih valjev.

Slika 3

Zahtevnejši predmeti so lahko sestavljeni iz več različnih posplošenih valjev.

3.3 Klasifikacija spojišč

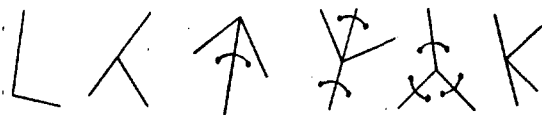
Guzman (1968) je napisal program, ki ugotavlja, kateri predmeti v obliki poliedrov sestavljajo sceno. Predmeti se lahko dotikajo drug drugega ter deloma prekrivajo.

Simbolni opis scene predstavljajo spojišča, linije in ploskve. Simbolni opis odgovarja dvo dimenzionalni črtni sliki.

Razmerja med dvo dimenzionalno črtno sliko in tri dimenzionalno sceno so naslednja:

- ločimo dve vrsti spojišč: tista, ki pripadajo ogliščem predmetov na sceni in tista, ki so posledica delnega zakrivanja robov
- črte pripadajo vidnim ali deloma vidnim robovom teles
- površine pripadajo vidnim ali deloma vidnim ploskvam teles

Največ informacije o temi, katere ploskve pripadajo istim in katere ploskve pripadajo sosednjim telesom, dobimo iz spojišč.

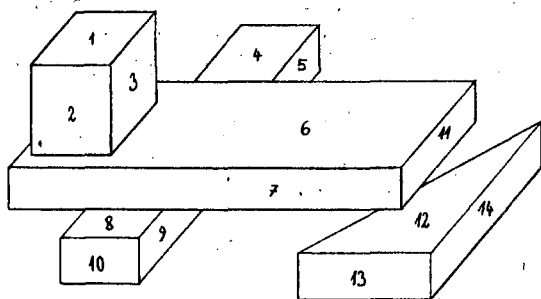


L T PUSZICA PSI VILICE K

Znak označuje pripadnost dveh ravnin istemu telesu.

Nekaj osnovnih tipov spojišč

Slika 4



Primer tridrske scene

Slika 5

Rezultati, ki ga da program kot opis scene na sliki 5 jet

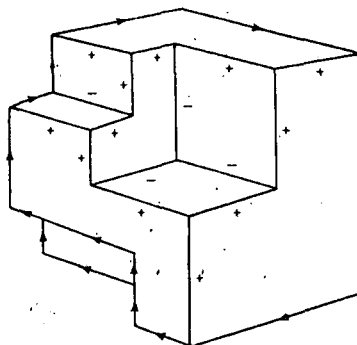
- (TELO 1 JE: 1:2:3)
- (TELO 2 JE: 4:5:8:9:10)
- (TELO 3 JE: 6:7:11)
- (TELO 4 JE: 12:13:14)

Guzmanov program lahko analizira precej zapletene scene; ni pa sposoben analizirati konkavnih predmetov in prepoznavati nesmiselnih predmetov.

3.4 Označevanje robov

Razen spojišč nam lahko tudi črte dvo dimenzionalne slike, ki predstavljajo robove teles, nudijo vir informacij pri analizi scene.

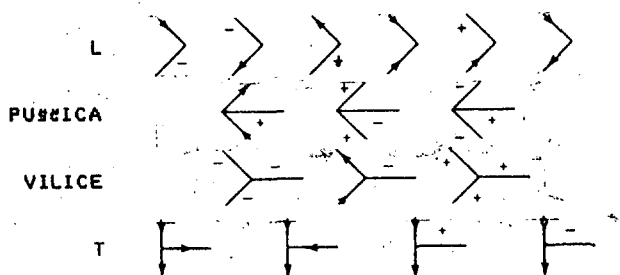
Ločimo dve vrsti robov: konkavne in konveksne robove. Konkavni robovi so na sliki vidni le, če sta vidni obe stični ploskvi. Označimo jih z znakom (-) minus. Konveksne robove označimo z znakom (+) plus; če sta vidni obe stični ploskvi, če pa je vidna le ena ploskva, pa govorimo o prikritem robu in ga označimo z znakom (→) puščica. Puščica je usmerjena tako, da je na desni vidna, na levi pa prikrita stična ploskva; če se premikamo v smeri puščice. Primer označevanja v skladu z omenjenimi pravili je na sliki 6.



Označevanje robov

Slika 6

Število vseh kombinacij je precej večje od števila dejansko možnih kombinacij različno označenih črt. Primeri možnih kombinacij za nekatera spojišča so na sliki 7. Vidimo da so na primer pri vilicah možne le 3 od skupno 81 kombinacij.



Možne kombinacije označitev za nekatera spojišča

Slika 7

S pomočjo informacije o tipih robov, ki se stekajo v spojišča je mogoče rešiti primere, v katerih je bil Guzmanov program nemočen. Analizirati je namreč mogoče tudi telesa z odprtini in vdolbinami.

Možno so še nadaljnje razširitve in izboljšave označevalnih metod in izločevanje namogočih kombinacij označenih črt v spojitvah.

Waltz (1976) razvije algoritem, ki upošteva osenčenje scene, s čimer pa se precej poveča število možnih kombinacij različnih označitev.

4. Uporaba računalniškega vida

Pri računalniškem vidu imamo opravka s prenosom in obdelavo velike množine informacije. Informacijo sprejmemo s pomočjo TV kamere ter pošljemo preko A/D vmesnika v računalnik, kjer jo obdelamo tako, da iz nje izluščimo opis, ki ga lahko potem koristno uporabimo.

Najpogosteje uporabljamo računalniški vid v industriji bodisi samostojno, bodisi kot dopolnilo inteligentnemu robotskemu sistemu.

Nekaj primerov uporabe robotov opremljenih s TV kamero:

- pobiranje različnih predmetov s tekočega traku in polaganje na določeno mesto (sortiranje, pakiranje, itd.)
- zajemanje različnih predmetov s kupa in polaganje na določeno mesto (sortiranje, pakiranje, podajanje stroju, itd.)
- barvanje na tekočem traku obešenih predmetov
- oblačno varjenje vpetih predmetov
- vrtnanje ploščic tiskanih vezij

Sistem za gledanje sestavlja niz programov za izračun ploskovno ali linijsko orientiranih opisov predmetov na podlagi slike, sprejete preko TV kamere. S pomočjo značilnosti kot so: površina, obseg, simetrijske osi in ukrivljenost predmeta, sistem identificira predmete, ki jih vidi. Naslednji korak je določitev lege predmetov in njihove orientacije. Lego določa običajno masno središče, orientacijo pa zasuk predmeta glede na osnovno orientacijo, ki jo določimo v učni fazi, ko predmet predstavljamo sistemu. Podatki o identitoti, legi in orientaciji predmeta sprožijo pri robotu ustrezno akcijo.

Uporaba računalniškega vida v industriji ni nujno vezana na robotski sistem. Nekatore samostojne aplikacije so:

- analiza in kontrola velikosti predmetov;
- analiza in kontrola površinskih lastnosti predmetov: barve, gladkosti, homogenosti, itd.

Računalniški vid uporablja tudi medicina, vojska in uprava javne varnosti.

Uporaba v medicini:

- analiza očiščnega tkiva
- analiza krvne slike
- analiza roentgonske slike

Uporaba v vojaške namene:

- analiza zračnih posnetkov
- odkrivanje maskiranih objektov
- orientiranje na nepoznanem terenu

Uporaba pri UJV:

- analiza prstnih odtisov
- analiza in kontrola prometa

Zanimivo in koristne pa so tudi druge aplikacije računalniškega vida:

- v meteorologiji pri napovedovanju vremena
- v astronomiji pri vodenju vesoljskih vozil

- v ekologiji pri določanju onesnaženosti voda
- v optični analizi tekstov, itd.

Računalniški vid je v zadnjih letih skokovito napredoval predvsem zaradi razvoja tehnologije mikroprocesorjev in TV kamer, ter pocenitve digitalnega spomina na tržišču. Vse kaže, da se bo dinamični razvoj nadaljeval tudi v prihodnosti.

5. Literatura

1. Bacjy R. (1972) Computer Identification of the Textured Visual Scenes. Stanford AI Lab Memo 180.
2. Binford T. O. (1971) Visual Perception by Computer. Presented to the IEEE Conference of Systems and Control, Miami.
3. Clowes M. B. (1971) On Seeing Things, Artificial Intelligence Journal 2, pp 79-116
4. Guzman A. (1968) Decomposition of Visual Scene into Three Dimensional Bodies, AFIPS proceedings of the Fall Joint Computer Conference 33, pp 291-304
5. Gerskowitx A., Binford T. (1970) On Boundary Detection, MIT AI Lab Memo 183
6. Horn B. K. P. (1975) Obtaining Shape from Shading Information, Psychology of Computer Vision, Ed. P. H. Winston; McGraw-Hill, New York pp 115-155
7. Horn B. K. P. (1978) Vision, Proceedings of AISB/GI Conference on Artificial Intelligence, Hamburg, pp 147-163
8. Huffman D. (1971) Impossible Objects as Nonsense Sentences, Machine Intelligence 6, Edinburgh University Press, Edinburgh
9. Marr D., Poggio T., Palm G. (1977) Analysis of the a Cooperative Stereo Algorithm, Biological Cybernetics 28, pp 223-239
10. Marr D. (1975) Early Processing of Visual Information, MIT AI Memo 340
11. Marr D. (1979) Visual Information Processing, IJCAI 79 pp 1108-1126
12. Roberts L. (1963) Machine Perception of Three Dimensional Solids, Tech. Report 315, MIT Lincoln Lab
13. Ullman S. (1979) The Interpretation of Structure from Motion, Proceedings of Royal Society London
14. Waltz D. (1976) Automata Theoretical Approach to the Visual Information Processing, Ed. Yeh R. T. in Applied Computation Theory, Prentice-Hall Inc., Englewood Cliffs, pp 468-529
15. Ward M. R., Rossol L., Holland S. W. (1979) CONSIGHT: A Practical Vision-Based Robot Guidance System, 9-th International Conference on Industrial Robot Technology, Washington D.C., pp 195-211
16. Winston P. H. (1977) Artificial Intelligence, Addison-Wesley Publishing Company, Reading