

Porozumenie obrazu

Sonka, Hlavac, Boyle: Image Processing, Analysis and Machine vision,
kapitola: Image understanding

Porozumenie obrazu

Počítačové videnie pozostáva z nižšej a vyššej úrovne, porozumenie obrazu je tou najvyššou úrovňou v tejto klasifikácii.

Hlavným cieľom počítačového videnia je dosiahnut' správanie počítača podobné biologickým systémom. Vytvára sa pri ňom v počítači **vnútorný model sveta**, na ktorý sa pozerá.

Človek je veľmi dobre pripravený na úlohy porozumenia, ale tiež niekedy nepozná riešenie (mikroskopické snímky, letecké snímky apod.).

Často má ohraničený pohľad na scénu.

Riadiace stratégie porozumenia obrazu

- **paralelné**

Takmer všetky operácie nižšej úrovne spracovania obrazu možno vykonávať paralelne.

- **sekvenčné**

Vyššie úrovne spracovania, ktoré používajú vyššiu úroveň abstrakcie, sú obvykle sekvenčné.

Hierarchické riadenie porozumenia obrazu

- Stratégiu riadenia volíme aj podľa toho akú informáciu z obrazu potrebujeme dostat'
- Ak potrebujeme ucelený obraz sveta tak zvolíme stratégiu zdola nahor - riadené údajmi ,
- ak však máme vnútorný model a chceme overiť, či náš vstup zodpovedá tomuto modelu použijeme stratégiu zhora nadol – riadené modelom.
- Taktiež existuje kombinovaná metóda týchto dvoch stratégií.

Riadenie zdola nahor

- Riadiaca stratégia riadená údajmi
- postup od ikonických obrazových dát, cez predspracovanie, segmentáciu až po porozumenie, ktoré pozostáva z porovnania rozpoznaných objektov s reálnymi objektmi.
- Táto stratégia je úspešná vtedy, keď sú k dispozícii kvalitné údaje, ktoré poskytujú dobrý vstup do vyššej fázy – napr. dobre osvetlené objekty na priemyselnom páse.
- Ďalším príkladom je Marrov postup od 2D ku 3D

Riadenie zhora nadol

- Riadiaca stratégia riadená modelom
- nemá takú ustálenú formu, akú je možné popísat' pre predchádzajúcu strategiu.
- Hlavným princípom tejto stratégie je vytvorenie vnútorného modelu a jeho verifikácia, čiže ide o cieľovo-orientovaný prístup.
- Ciele na vyššej úrovni sa rozložia na podciele v nižšej úrovni spracovania atď>.
- Základný mechanizmus je generovanie hypotéz a ich overovanie.

Kombinované riadiace stratégie

- kombinujú riadenie cez údaje aj cez model a obvykle dávajú lepšie výsledky ako tie základné stratégie.
- Kombinujú sa informácie na vyššej úrovni spracovania, ktoré uľahčujú spracovanie na nižšej úrovni.
- Používajú sa aj rozličné rozlíšenia obrazu.

Príklad kombinovaného riadenia - rozpoznanie priebehu cievy

- 1) prístup založený na modely – model favorizuje ľavé a pravé hranice ako typické pre cievne zobrazovanie
- 2) prístup založený na tvorbe hypotéz a ich verifikovaní – približná hranica cievy je detegovaná na nižšej úrovni rozlíšenia a presnosť sa zvyšuje pri plnom rozlíšení
- 3) využitie apriórnej znalosti – znalosť o smeroch hrán tvoriacich hranicu cievy môže modifikovať cenovú funkciu použitú pri algoritme hľadania hranice
- 4) viacstupňový prístup – počas spracovania sa uplatňujú modely rôznej sily.

- Hľadanie hranice sa vykonáva v 3D grafe, hľadá sa ľavá aj pravá hranica súčasne a ich symetria je zakomponovaná do procesu hľadania.
- Hľadanie sa riadi
 - modelom tam, kde sú slabé obrazové dátá a
 - obrazovými dátam tam, kde ich kvalita je dobrá.
- Otázka sily modelu sa uplatňuje tak, že
 - silný model sa použije pri nízkom rozlíšení a
 - slabší model pri vyššom rozlíšení ponechajúc priestor pre hľadanie na základe obrazových dát pri plnom rozlíšení, pričom výsledky na nižšej úrovni sú modelovou hranicou pre vyššiu úroveň.

Algoritmus na hľadanie hranice cievky – kombinovaná riadiaca stratégia

1. (zhora nadol) – deteguj stredovú čiaru cievky spolu s operátorom (ukáž cievu, ktorá sa má spracovať) a narovnaj obraz cievky do vodorovnej osi (10.3. a,b)
2. (zdola nahor) – deteguj obrazové hrany pri plnom rozlíšení (10.3. c)
3. (zdola nahor) –
deteguj smery
lokálnych hrán
v narovnanom obraze
(10.3. d)

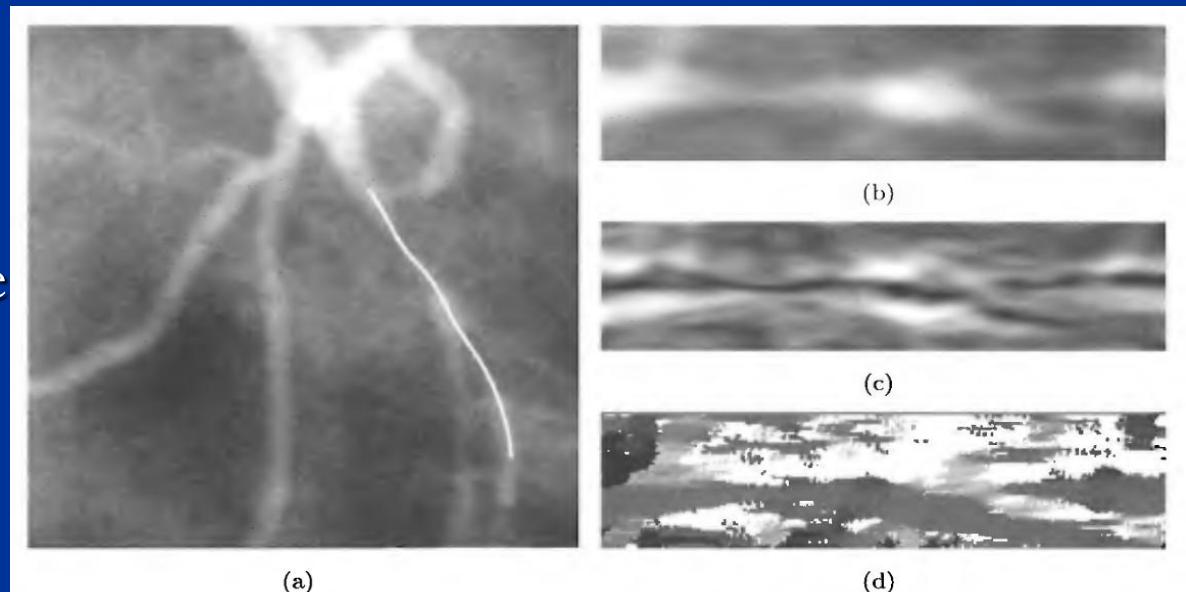


Figure 10.3: Steps of coronary border detection (I). (a) Centerline definition. (b) Straightened image data. (c) Edge detection. (d) Edge direction detection.

4. (zhora nadol) – modifikuj cenovú maticu s použitím apriórnej znalosti o smeroch hrán (10.4. a)
5. (zdola nahor) – zostroj obraz v nízkom rozlíšení a cenovú maticu pre takéto rozlíšenie
6. (zhora nadol) – hľadaj na nízkej úrovni rozlíšenia dvojicu hraníc s využitím informácie o symetrii (10.4. b)

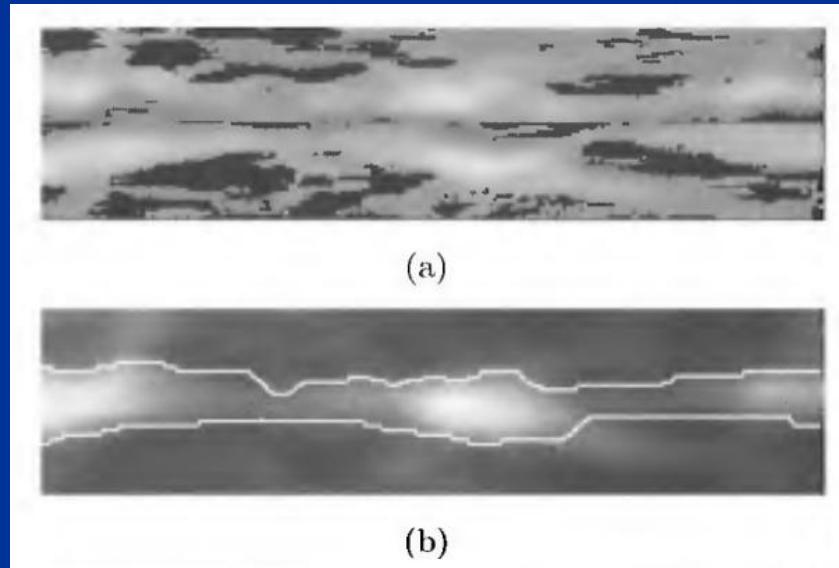
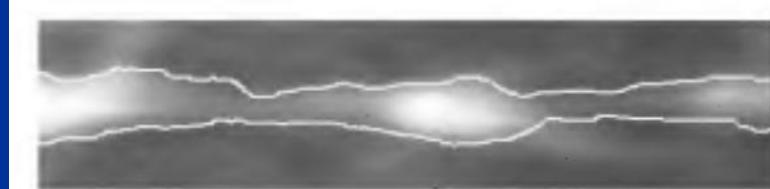


Figure 10.4: Steps of coronary border detection (II). (a) Modified cost function—note that the cost increases in non-probable border locations image areas, where edge direction does not support location of the border. (b) Approximate coronary borders acquired in low resolution.

7. (zhora nadol) – nájdi správnu polohu pre hranicu v plnom rozlíšení s použitím hranice na nízkej úrovni modelu. (10.4. c)
Symetrický model je teraz slabší ako pri nízkej úrovni rozlíšenia.
8. (zdola nahor) – transformuj výsledky z narovnaného obrazu do pôvodného obrazu. (10.4. d)
9. (zhora nadol) – ohodnot' vážnosť poškodenia cievky.



(c)



(d)

(c) Precise full-resolution border in straightened image. (d) Full-resolution coronary borders in original image data.

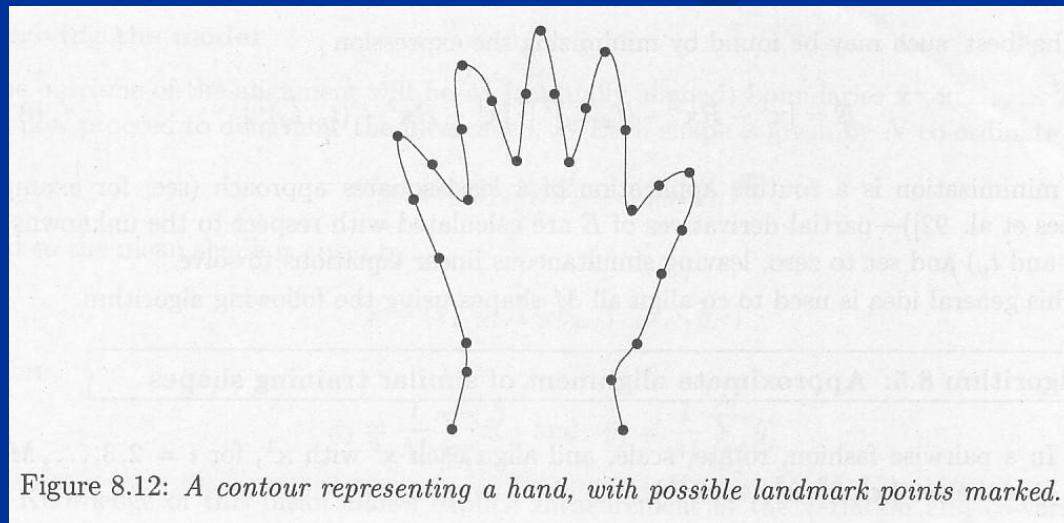
Nehierarchické riadenie

- nerozlišuje medzi nižšou a vyššou úrovňou spracovania.
- spolupráca konkujúcich expertov a princíp tabule na zdielanie údajov medzi expertmi.
- Používa sa na problémy, ktoré sa dajú rozdeliť na menšie podproblémy, z ktorých každý si vyžaduje určitú expertnú znalosť.
- Použije sa expert, ktorý môže najviac prispiet' ku riešeniu, vykoná sa akcia a zhodnotí sa jej prínos pre zvýšenie informácie o probléme.

- Tabuľa slúži na zdieľanie výsledkov jednotlivých expertov.
- Obsahuje aj mechanizmus na ovplyvnenie riadenia procesu cez tzv. démonov, pričom musí obsahovať aj ich synchronizáciu.
- tabuľa reprezentuje
 - stále dopĺňanú časť interného modelu odpovedajúceho obrazovým dátam a súčasne
 - obsahuje mechanizmus, špecifikujúci, ktorý démon sa použije pre ten ktorý prípad.
- Tabuľa sa nazýva aj krátkodobá pamäť, pričom dlhodobá pamäť reprezentuje bázu znalostí, ktorá pozostáva zo všeobecnejších znalostí.

Modely rozdelujúce body (Point distribution models)

- PDM je technika na popis tvaru, ktorá sa dá použiť na lokalizovanie podobných tvarov v iných obrazoch.
- Je veľmi užitočná na popis príznakov, ktoré majú známy „všeobecný“ tvar, ale ktoré nemôžu byť popísané presným modelom kvôli variáciám tvaru.



- PDM prístup vyžaduje existenciu trénovacej množiny príkladov (často značiek tvaru), z ktorých sa odvodzuje štatistický popis tvaru a jeho variácie, ktoré sa dajú vyjadriť ako vlastné hodnoty kovariančnej matice.
- PDM popisuje modely variácií – smery maximálnych variácií sú usporiadané, takže sa vie, kde sa variácie najskôr objavia.
- Väčšina variácií tvaru sa dá vyjadriť malým počtom spôsobov variácie.
- Zarovnanie trénovacích dát (data aligning)

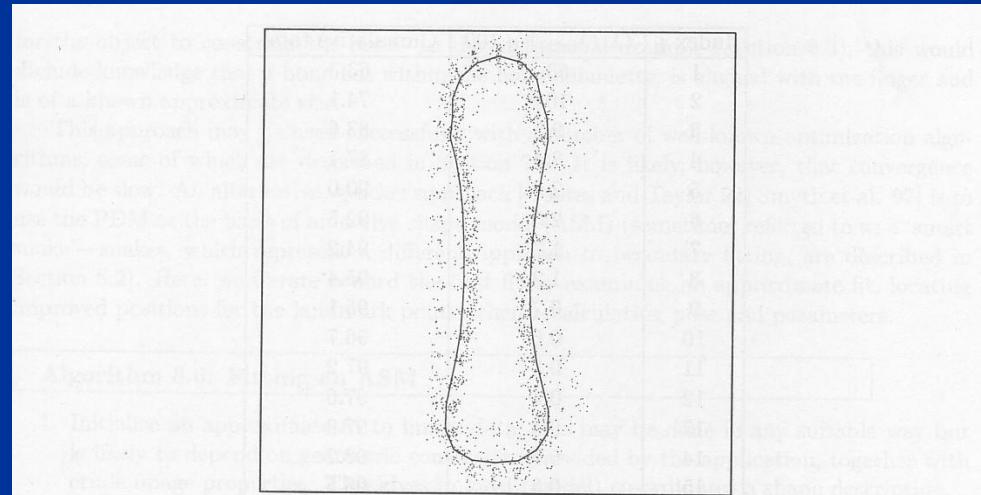


Figure 8.13: PDM of a metacarpal. Dots mark the possible positions of landmarks, and the line denotes the mean shape. Courtesy N.D. Efford, School of Computer Studies, University of Leeds.

Algorithm: Approximate alignment of similar training shapes

1. In a pairwise fashion, rotate, scale, and align each x^i with x^1 , for $i = 2, 3, \dots, M$ to give the set $\{x^1, \tilde{x}^2, \tilde{x}^3, \dots, \tilde{x}^M\}$.
2. Calculate the mean of the transformed shapes
3. Rotate, scale, and align the mean shape to align to x^1 .
4. Rotate, scale, and align $\tilde{x}^2, \tilde{x}^3, \dots, \tilde{x}^M$ to match to the adjusted mean.
5. If the mean has not converged, go to step 2.

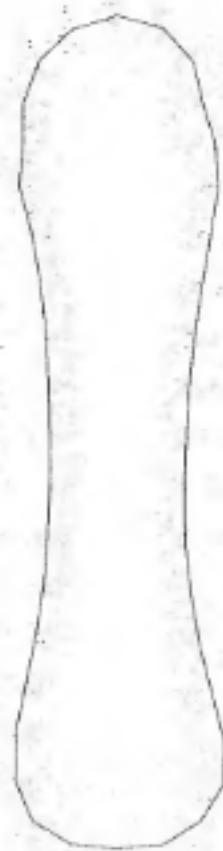


Figure 10.8: PDM of a metacarpal. Dots mark the possible positions of landmarks, and the line denotes the mean shape. *Courtesy of N. D. Efford, School of Computer Studies, University of Leeds.*

Index i	$\lambda_i/\lambda_{\text{total}} [\%]$	Cumulative total
1	63.3	63.3
2	10.8	74.1
3	9.5	83.6
4	3.4	87.1
5	2.9	90.0
6	2.5	92.5
7	1.7	94.2
8	1.2	95.4
9	0.7	96.1
10	0.6	96.7
11	0.5	97.2
12	0.4	97.6
13	0.3	97.9
14	0.3	98.2
15	0.3	98.5
16	0.2	98.7

Table 10.1: Relative contributions to total data variance for the first 16 principal components.

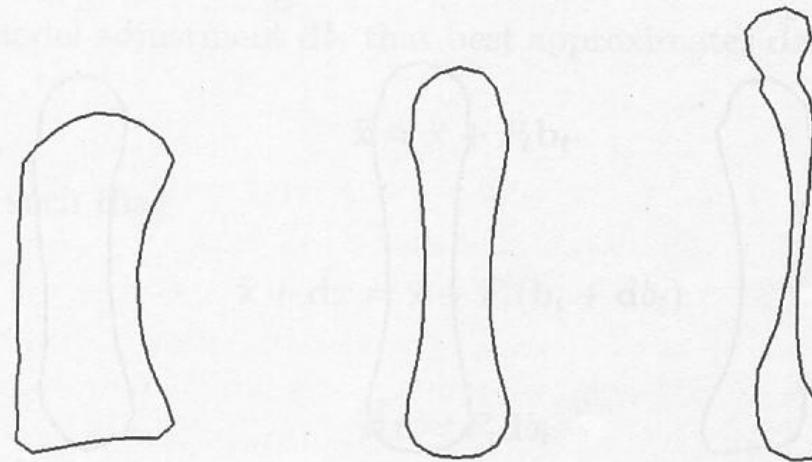


Figure 8.14: *The first mode of variation; $-2.5\lambda_1$, mean shape, $2.5\lambda_1$.* Courtesy N.D. Efford, School of Computer Studies, University of Leeds.

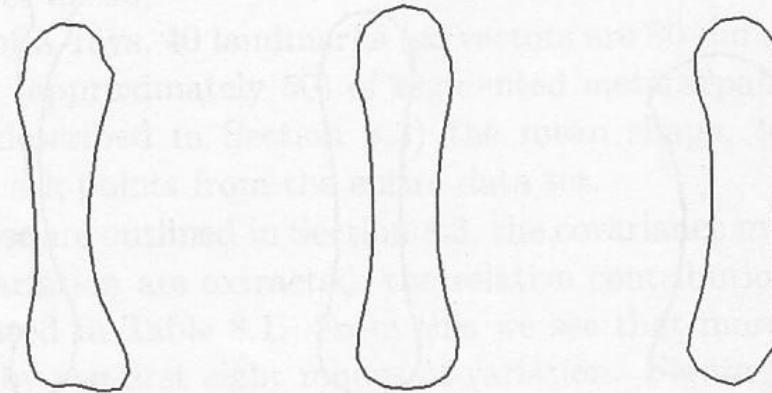


Figure 8.15: *The third mode of variation; $-2.5\lambda_3$, mean shape, $2.5\lambda_3$.* Courtesy N.D. Efford, School of Computer Studies, University of Leeds.

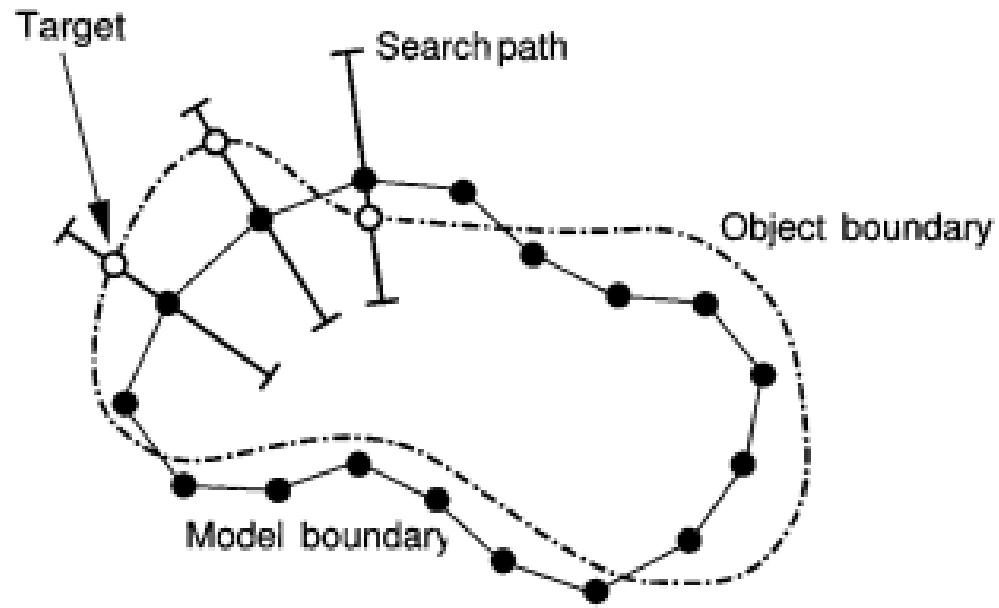


Figure 10.11: Searching an approximate model fit for target points to which landmarks may move. *Courtesy of N.D. Efford, School of Computer Studies, University of Leeds.*

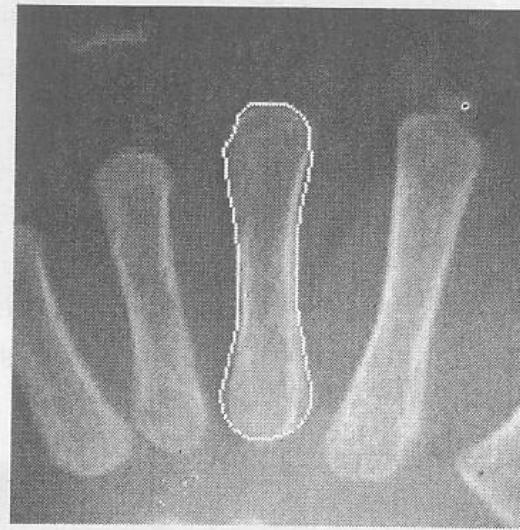
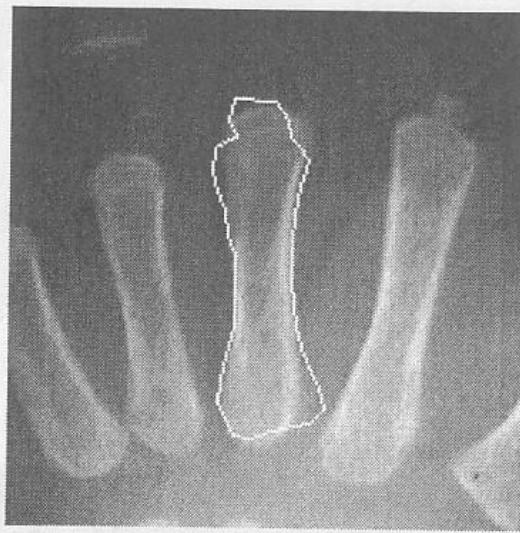
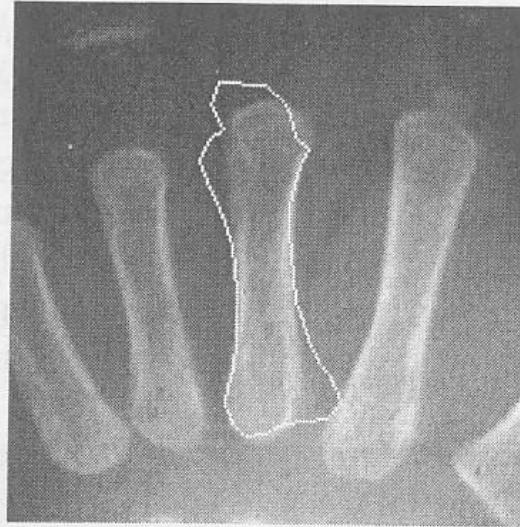
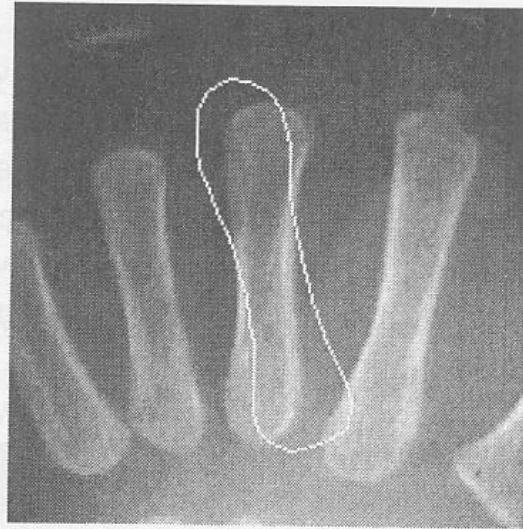


Figure 8.17: Fitting an ASM to a metacarpal; various stages of convergence—initialization, 3, 6, and 10 iterations. Courtesy N.D. Efford, School of Computer Studies, University of Leeds.

Fitovanie modelu pomocou náhodného výberu vzoriek - RANSAC

- RANSAC – random sample consensus
- Tradičné prístupy k fitovaniu modelu sú založené na princípe najmenších štvorcov – minimalizovanie chyby
- Ak dáta nie su kvalitné, odľahlé body (outliers) môžu negatívne ovplyvniť model
- RANSAC začína s jednoduchým modelom založeným na malej vzorke dát, potom použije zvyšné dáta na potvrdenie modelu a identifikovanie odľahlých bodov - outliers
- Odľahlé body sa odstránia a model sa prepočíta.
- RAMSAC reprezentuje zmenu paradigmy vo fitovaní modelu: „začať s malou vzorkou a pridávať“, čo je opačný prístup ako metóda najmenších štvorcov

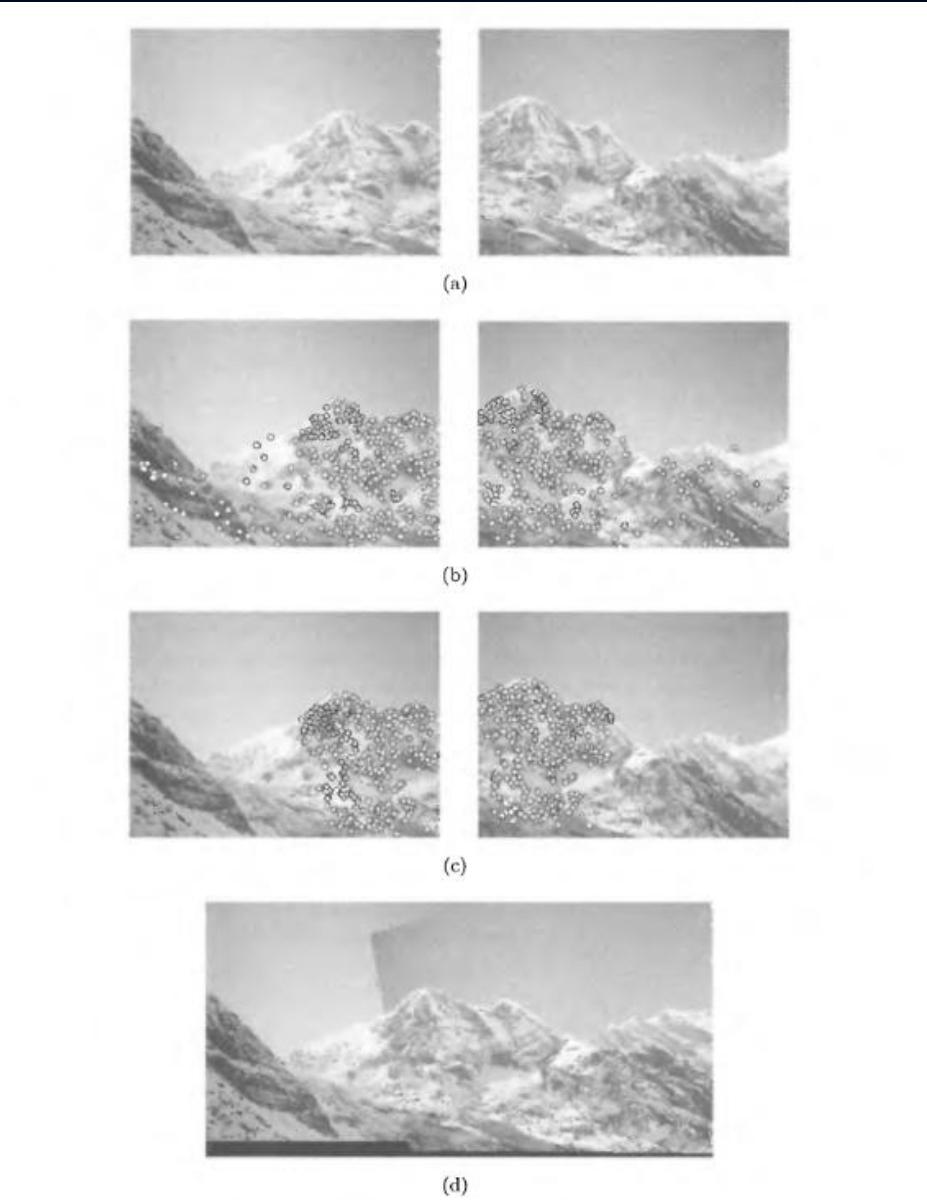


Figure 10.6: Use of RANSAC in panoramic stitching: (a) An image pair that overlaps. (b) Each dot represents an ‘interest point’, whose vector matches a point in the other image. (c) The illustrated dots are the RANSAC ‘inliers’; they conform to the hypothesis of the best-performing candidate homography that is tried. (d) Prior to smoothing, the resulting overlap. *Courtesy of D. Lowe and M. Brown, University of British Columbia. The original images are, of course, in color.*

Rozpoznávanie obrazcov v porozumení obrazu

- Kontrolované alebo nekontrolované metódy rozpoznávania obrazcov možno použiť pre klasifikáciu pixlov.
- Pri porozumení obrazu sa príznakové vektory získané z lokálnych multispektrálnych hodnôt susedných pixlov predkladajú klasifikátoru, ktorý prideluje označenie každému pixlu v obraze, na základe čoho sa potom skúmanému pixlu pridelí konečné označenie.
- Tým sa dosahuje kontextuálna klasifikácia.
- Výsledný označený obraz môže mať veľa malých oblastí, ktoré sa zle klasifikujú. Na to je možné použiť kontextovo závislý postup dodatočného spracovania, aby sme sa vyhli zlej klasifikácii.

Označovanie scén

- Pravdepodobnostné označovanie umožňuje existenciu viacerých označení pre jeden objekt.
- Značky sú vážené pravdepodobnostne, s dôverou pre každú značku v každom objekte.
- Tento proces vždy dá výsledok interpretácie spolu s mierou dôvery pre takúto interpretáciu.

- A window is rectangular.
- A table is rectangular.
- A drawer is rectangular.

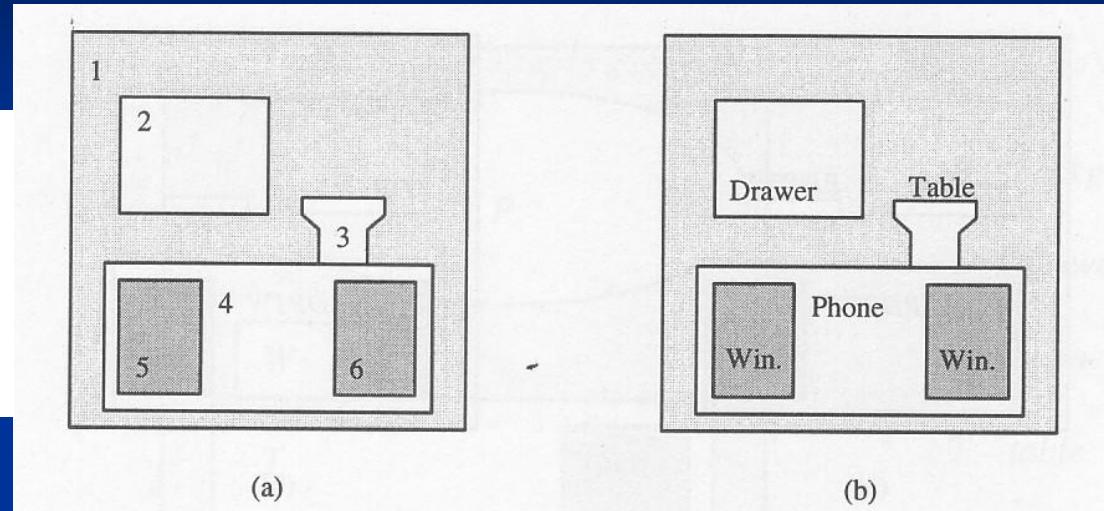
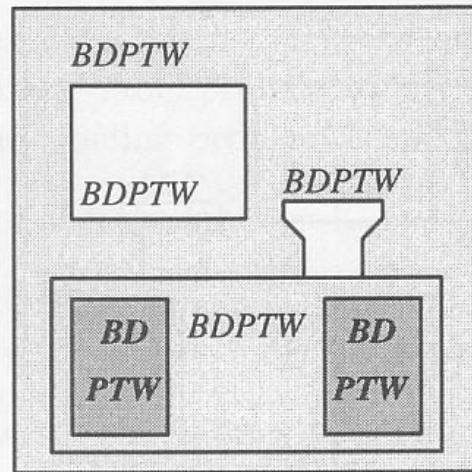


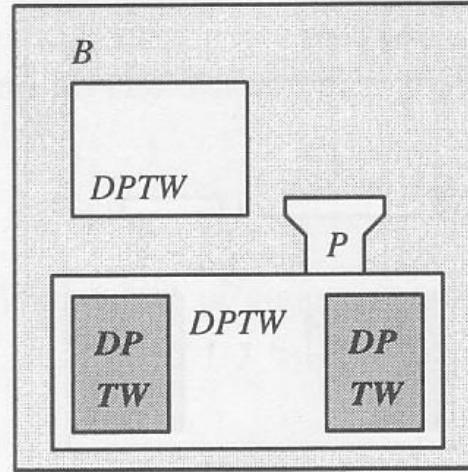
Figure 8.24: *Scene labeling*: (a) scene example; (b) inconsistent labeling.

Let the binary constraints be:

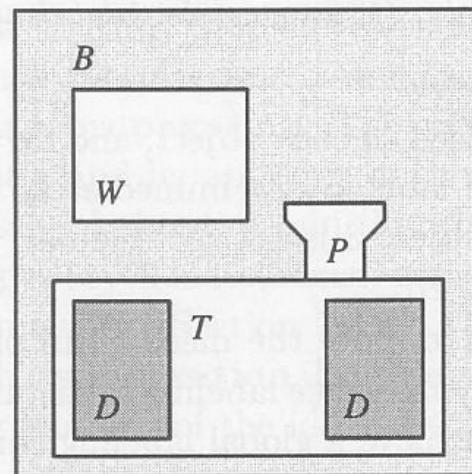
- A window is located above a table.
- A phone is above a table.
- A drawer is inside a table.
- Background is adjacent to the image border.



(a)



(b)



(c)

B background
D drawer
P phone
T table
W window

Figure 8.25: Discrete relaxation: (a) all labels assigned to each object; (b) locally inconsistent labels are removed; (c) final consistent labeling.

Princíp posúvania obmedzení

- umožňuje prispôsobovať lokálne obmedzenia globálnym obmedzeniam v celom obraze.
- Označovanie objektov závisí od vlastností objektu a miery kompatibility potenciálnych značiek objektov so značkami ostatných súvisiacich objektov.
- Aj vzdialené objekty môžu spolu súvisieť vzhľadom na princíp posúvania obmedzení.
- Ked' prehľadávame strom interpretácií, vrcholom stromu sú priradené všetky možné značky a používa sa hľadanie do hĺbky s návratom (back-tracking). Testujú sa všetky možné označenia.

Sémantická segmentácia obrazu a porozumenie

- Sémantické narastanie oblastí zahŕňa kontext do spájania oblastí s využitím apriórnej informácie o vzťahoch medzi susednými oblastami a tiež posúvanie obmedzení na dosiahnutie globálne optimálnej segmentácie a interpretácie obrazu.
- Pri odstraňovaní slabej hranice sa berie do úvahy aj kontext, čo predstavujú oblasti, ktoré sa majú spájať.

Sémantická segmentácia obrazu a porozumenie

- Genetická interpretácia obrazu je založená na princípe overovania hypotéz.
- Objektívna funkcia, ktorá ohodnocuje kvalitu segmentácie a interpretácie, sa optimalizuje genetickým algoritmom, ktorý generuje nové generácie hypotéz o segmentácii a interpretácii obrazu, ktoré treba otestovať.

$$C_{\text{image}} = \frac{\sum_{i=1}^{N_R} C(\theta_i)}{N_R}$$

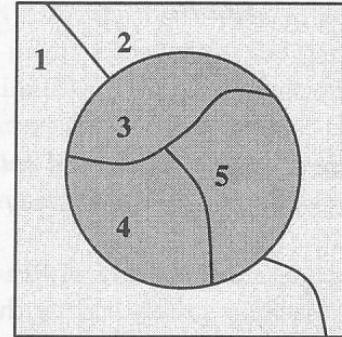


Figure 8.29: A simulated scene ‘ball on the lawn’.

- Unary condition: Let the confidence that a region is a *ball* be based on its compactness (see Section 8.3.1),

$$C(\theta_i = B | X_i) = \text{compactness}(R_i) \quad (10.49)$$

and let the confidence that a region is *lawn* be based on its greenness,

$$C(\theta_i = L | X_i) = \text{greenness}(R_i) . \quad (10.50)$$

Let the confidences for regions forming a perfect ball and perfect lawn be equal to one

$$C(B|\text{circular}) = 1 \quad C(L|\text{green}) = 1 .$$

- Binary condition: Let the confidence that one region is positioned inside the other be given by a compatibility function

$$r(B \text{ is inside } L) = 1 \quad (10.51)$$

and let the confidences of all other positional combinations be equal to zero.

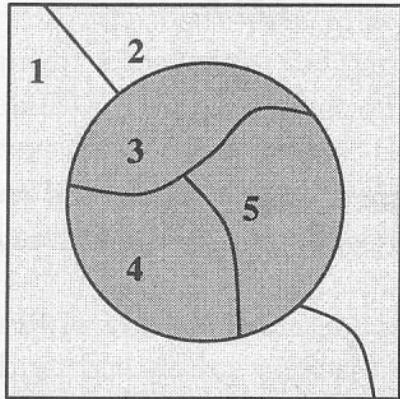
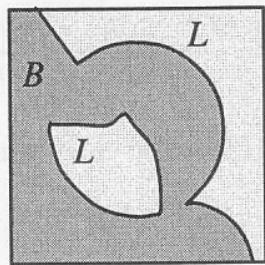
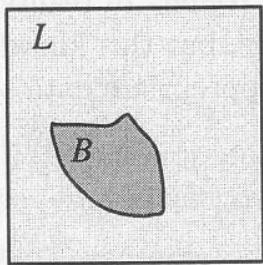


Figure 8.29: A simulated scene ‘ball on the lawn’.

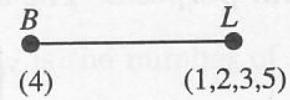
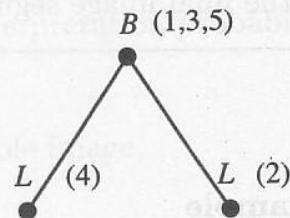
$BL BLB$	$C_{\text{image}} = 0.00$
$LL LBL$	$C_{\text{image}} = 0.12$
$LLBLB$	$C_{\text{image}} = 0.20$
$BLLBL$	$C_{\text{image}} = 0.00$



$BLBLB$

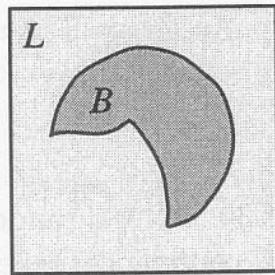


$LLLBL$

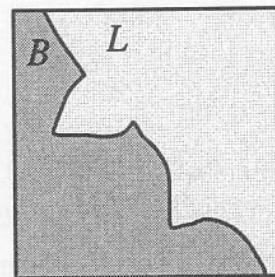
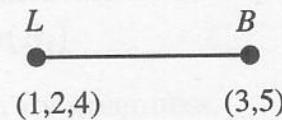


$LLL BL$	$C_{\text{image}} = 0.12$
$LLB LB$	$C_{\text{image}} = 0.20$
$LLLLB$	$C_{\text{image}} = 0.14$
$LLBBL$	$C_{\text{image}} = 0.18$

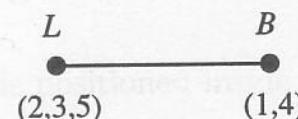
Figure 8.30: Starting hypotheses about segmentation and interpretation: interpretation, corresponding code strings, and corresponding region adjacency graphs.



LLBLB

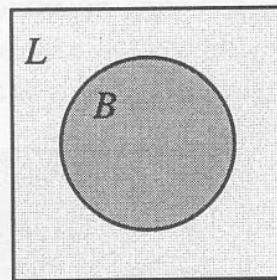


BLLBL



<i>LLBL B</i>	$C_{\text{image}} = 0.20$
<i>LLBB L</i>	$C_{\text{image}} = 0.18$
<i>LLBLL</i>	$C_{\text{image}} = 0.10$
<i>LLBBB</i>	$C_{\text{image}} = 1.00$

Figure 8.31: *Hypotheses about segmentation and interpretation: interpretations, corresponding code strings, and corresponding region adjacency graphs.*



LLBBB

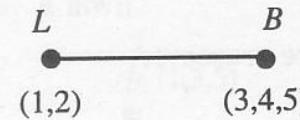


Figure 8.32: *Optimal segmentation and interpretation: interpretation, corresponding code string, and region adjacency graph.*

Označovanie scén

- Diskrétné označovanie umožňuje iba jednu značku pre každú objekt vo výslednom označení.
- Je snaha dosiahnuť konzistentné označenie v celom obrazze.
- Tento proces bud' skončí konzistentným označením alebo nemožnosťou konzistentného označenia scény.

Marr

- Zlom vo výskume 3D videnia statickej scény
- Navrhol teóriu 3D videnia
- Svoj prístup považoval za zvláštny prípad všeobecného stoja na spracovanie informácií, ktorý chápal v 3 úrovniach

Marrova teória

- **1. Výpočtová teória:** opisuje čo zariadenie má robiť akú informáciu má poskytnúť zo vstupnej informácie. Tiež by mala opisovať logiku stratégie, ktorá vykonáva túto úlohu.
- **2. Reprezentácia a algoritmus:** opisuje ako má byť teória realizovaná, konkrétnie reprezentácia informácie a algoritmus na manipuláciu s nimi
- **3. Implementácia:** opisuje konkrétnie programy pre konkrétny hardvér

3 úrovne reprezentácie 3D scény

- Prvotný náčrt (primal sketch),
 - 2.5 rozmerný náčrt (2.5D sketch),
 - Plná 3D reprezentácia (full 3D representation).
-
- Postupne sa zvyšuje množstvo informácie o 3D tvare

Prvotný náčrt

- Nájst' dôležité zmeny intenzity v obraze – hrany
- Zmeny intenzity sa určujú v rôznych škálovaniach – bluring filters - Gaussov filter
- Hrany sa určia pomocou zero-crossing – Laplacian operátor
- Zero-crossing sa zoskupia podľa polohy a orientácie aby sme získali informáciu o tokenoch (hrany, pruhy, bloby) ktoré poskytnú informáciu o orientácii plôch v scéne

2.5D náčrt

- Rekonštruuje relatívnu vzdialenosť od pozorovateľa k povrchom – hĺbková mapa
- Určujú sa normály plôch
- Hlavný prístup sa nazýva Tvar z X, kde X môže byť pohyb, jas, stereovidenie, textúra

3D reprezentácia

- Prechod k objektovo centrovanej súradnicovej sústave, nezávislosť na pozorovateľovi
- Reprezentácia pomocou volumetrických primitív na rozdiel od reprezentácie pomocou povrchov
- Marrova teória je dobrý teoretický rámec ale nevedie k dobrým výsledkom pri vision aplikáciách

Ďalšie paradigmá videnia: aktívne a účelové videnie

2 školy sa snažia vysvetliť mechanizmus videnia

1. Staršia

- sa snaží použiť explicitnú metriku na začiatku úlohy tj. na čiary, krivky, normály.
- Geometria sa získava spôsobom bottom-up
- Výstupom je geometrický model

2. Mladšia

- nezískava metriku z vizuálnych dát pokiaľ si to úloha nevyžiada
- Dáta sa zberajú systematicky
- Databáza obrazov je model

Aktívne videnie

- Veľa tradičných CV systémov a teórií snímajú dátá kamerami s pevnými charakteristikami – aktívne a účelové videnie môžu byť vhodnejšie
- V aktívnom vision systéme sú vlastnosti snímania dát dynamicky kontrolované interpretáciou scény.
 - Veľa vizuálnych úloh je jednoduchších ak je pozorovateľ je aktívny a kontroluje vizuálne senzory (pohyb kamery)
 - Ak nemáme dost dát interpretovať scénu, kamera sa môže pozerat z iného bodu.
 - Aktívne videnie je inteligentné získavanie dát kontrolované meranými, čiastočne interpretovanými parametrami scény a ich chybami

<i>Task</i>	<i>Passive observer</i>	<i>Active observer</i>
Shape from shading	Ill-posed. Regularization helps but a unique solution is not guaranteed due to non-linearities.	Well-posed. Stable. Unique solution. Linear equations.
Shape from contour	Ill-posed. Regularization solution not formulated yet. Solution exists only for very special cases.	Well-posed. Unique solution for monocular or binocular observer.
Shape from texture	Ill-posed. Assumptions about texture needed.	Well-posed without assumptions.
Structure from motion	Well-posed but unstable.	Well-posed and stable. Quadratic constraints. simple solution.

Table 11.1: Active vision makes vision tasks well-posed.

Kvalitatívne videnie

- Hľadá kvalitatívny opis objektov alebo scén
- Motivácia je nereprezentovať geometriu, ktorá nie je potrebná pre kvalitatívne (negeometrické) úlohy alebo rozhodnutia
- Kvalitatívna informácia je viac invariantná voči rôznym nechceným transformáciám (trochu odlišný bod pohľadu) alebo šumu ako kvantitatívne.
- Kvalitatívnosť umožňuje interpretáciu pozorovanie udalosti na niekoľkých úrovniach komplexity

Účelové videnie

- Dôležité je identifikovať cieľ úlohy a zjednodušiť ju určením iba tej informácie, ktorá je potrebná
- Prístup môže byť heterogénny a kvalitatíva odpoved' je postačujúca
- Účelové videnie zatiaľ nemá solídny teoretický základ, ale štúdium biologického videnia je bohatý zdroj inšpirácie