

Analiza și Prelucrarea Secvențelor de Imagini

Ș.I.dr.ing. Bogdan IONESCU

București, 2010

> M4. Structura temporală

- 4.1. [Introducere]
- 4.2. [Detecția de “cut”]
- 4.3. [Detecția de “fades”]
- 4.4. [Detecția de “dissolves”]
- 4.5. [Activitate inter-plan]

Analiza și Prelucrarea Secvențelor de Imagini, Ș.I. Bogdan IONESCU

1

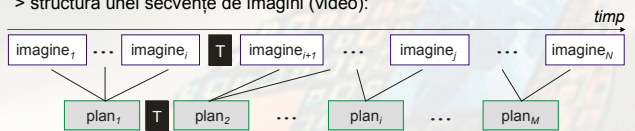
> M4. Structura temporală [Introducere]

Analiza și Prelucrarea Secvențelor de Imagini, Ș.I. Bogdan IONESCU

2

Planele video

> structura unei secvențe de imagini (video):



- **image** (frame) = “atomul” secvenței;
- **plan video** (video shot) = secvența continuă de imagini înregistrată de la pornirea camerei până la oprirea acesteia;
- **tranziție video** (video transition, T) = un efect vizual de scurtă durată ce face legătura între planele video vecine temporale;

segmentare temporală = descompunerea secvenței în plane video;
= detecția tranzițiilor video;

Analiza și Prelucrarea Secvențelor de Imagini, Ș.I. Bogdan IONESCU

3

Tranziții video

A. tranziții abrupte (sharp) = “cut”

> un cut reprezintă concatenarea directă a două plane video vecine temporale:



Analiza și Prelucrarea Secvențelor de Imagini, Ș.I. Bogdan IONESCU

4

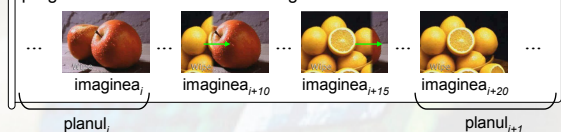
Tranziții video

[Lienhart 01]

B. tranziții graduale (gradual transitions)

- **clasa spațială**: aplică imaginilor celor două plane transformări spațiale;
- > exemple: wipes, mattes, page turns, ...

> wipe reprezintă efectul prin care o imagine este înlocuită progresiv de o alta folosind o margine de o anumită formă:



Analiza și Prelucrarea Secvențelor de Imagini, Ș.I. Bogdan IONESCU

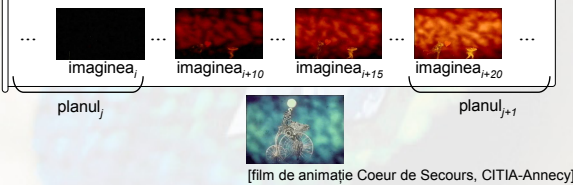
5

Tranziții video

B. tranziții graduale (continuare)

- **clasa cromatică**: imaginile planelor video sunt modificate prin transformări de culoare;
- > exemple: fades, ...

> fade este o tranziție ce permite, fie dizolvarea progresivă a unei anumite imagini într-o imagine constantă (fade-out), fie apariția progresivă a unei imagini dintr-o imagine constantă (fade-in);

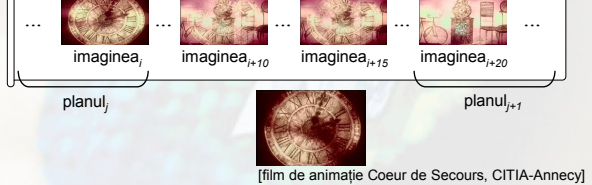


Tranziții video

B. tranziții graduale (continuare)

- **clasa cromatică**: imaginile planelor video sunt modificate prin transformări de culoare (continuare);
- > exemple: dissolves, ...

> dissolve reprezintă superpoziția unui efect fade-out peste un efect fade-in, suprapunere ce are ca efect vizual dizolvarea unei imagini în alta;



Tranziții video

B. tranziții graduale (continuare)

- **clasa spațio-cromatică**: o combinație a clasei spațiale și cromatice, imaginile planelor fiind modificate atât prin transformări spațiale, cât și cromatice;
- > exemple: morphing, ...

> morphing reprezintă transformarea unui obiect în altul, atât din punct de vedere geometric cât și al culorilor;



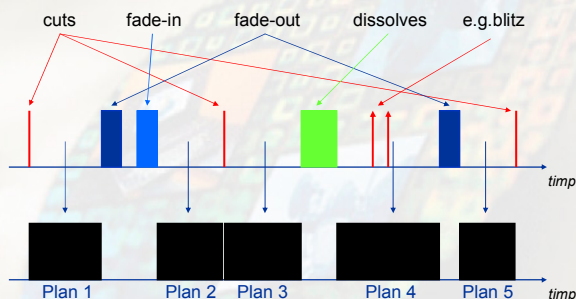
Tranziții video

> Observații:

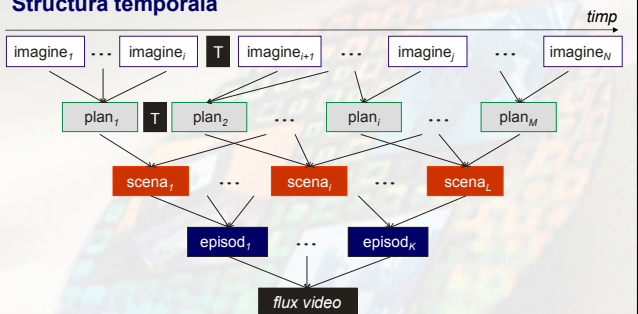
- tranzițiile de tip cut sunt cel mai des întâlnite, de regulă 30min de film conțin ~ 300 de cut;
- tranzițiile de tip fade (fade-in, fade-out) și dissolve sunt cele mai frecvente tranziții graduale;
- frecvența de apariție este cel puțin cu un ordin de măsură mai redusă decât la cut, 30min ~ 30 tranziții graduale;
- folosirea tranzițiilor nu este întâmplătoare:
 - frecvență \nearrow de cut \approx acțiune, ritm
 - fade și dissolve \approx schimbă timpul sau locul acțiunii;
 - fade-out + fade-in = pauză (plan video de imagini negre);

Determinarea planelor video

> exemplu:



Structura temporală



- **scena** = grupuri de plane video corelate din punct de vedere al conținutului semantic (unitate de loc, de timp și acțiune);
- **episod** = grupurilor de scene ce sunt similare din punct de vedere al acțiunii globale;

> M4. Structura temporală [Detecția de "cut"]

Detecția tranzițiilor de tip cut

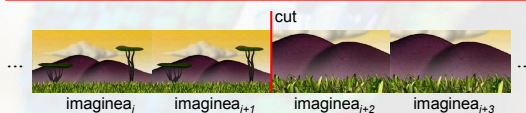
> formal un cut este definit astfel:

$$S(x, y, t) = [1 - u(t - t_{cut})] \cdot P_1(x, y, t) + u(t - t_{cut}) \cdot P_2(x, y, t)$$

unde (x, y) sunt coordonate spațiale, t coordonata temporală, P_1 și P_2 sunt cele două plane concatenate, t_{cut} reprezintă momentul de apariție al primei imagini de după cut iar $u(t)$ este:

$$u(t) = \begin{cases} 1 & t \geq 0 \\ 0 & \text{altfel} \end{cases}$$

→ un cut introduce o discontinuitate vizuală importantă !



Detecția tranzițiilor de tip cut

> etape generale de detecție:

→ **parametrizarea:** [puterea discriminatorie];

> la nivel de imagine sunt extrași o serie de parametri care evidențiază această discontinuitate, ex. histograme, statistici, etc.;

→ **variația temporală:** [cut 7, restul 7];

> se evaluează variația acestora între imaginile la momentele k și $k+l$, unde $k=1, \dots, N_{secv}$ (indicele temporal) iar l este pasul de analiză ($l \geq 1$);

> se folosesc măsuri de distanță (măsuri de similaritate).

→ **compararea cu un prag** (thresholding): [~ optimal];

> discontinuitate > prag atunci un cut are loc între momentele k și $k+l$ (precizie ?)

Detecția tranzițiilor de tip cut

> principala sursă de erori este disimilaritatea imaginilor din cadrul aceluiasi plan video (\neq definiție teoretică), cauze principale:

- mișcarea camerei video,
- mișcarea obiectelor din scenă,
- fluctuații ale intensității luminoase.

> soluții: folosirea altor surse de informație pe lângă valorile de discontinuitate, exemplu:

- compensarea mișcării [Marichal 98],
- divizarea imaginii în mai multe regiuni [Ionescu 06],
- detecția blițului aparatului foto [Heng 99].

Detecția tranzițiilor de tip cut

> metodele de detecție existente:

- ↳ bazate pe analiza intensității pixelilor,
- ↳ bazate pe analiza contururilor,
- ↳ bazate pe analiza de mișcare,
- ↳ realizează detecția în domeniul comprimat (MPEG),
- ↳ bazate pe alte surse de informație,
- ↳ mixte.

Detecția tranzițiilor de tip cut

metode bazate pe analiza intensității pixelilor

> cea mai simplă metodă constă în calculul diferenței dintre imagini consecutive la momentele k și $k+l$, de exemplu:

- [Otsuji 91] dacă numărul de pixeli N_{pixeli} ce se schimbă de la o imagine la alta $> T \Rightarrow$ cut

$$N_{pixeli} = \frac{1}{X \cdot Y} \sum_{x=1}^X \sum_{y=1}^Y D_{k,k+l}(x, y)$$

unde $X \times Y$ este dimensiunea imaginii iar $D_{k,k+l}$ funcția de discontinuitate:

$$D_{k,k+l} = \begin{cases} 1 & |I_k(x, y) - I_{k+l}(x, y)| > T_1 \\ 0 & \text{altfel} \end{cases}$$

unde I_k este imaginea la momentul k iar T_1 un prag.



Detecția tranzițiilor de tip cut

metode bazate pe analiza intensității pixelilor (continuare)

> calculul diferenței dintre imagini pe baza histogramelor color sau pe niveluri de gri:

- [Yeo 95] măsura de discontinuitate: distanța dintre binii histogramelor imaginilor la momentele k și $k+l$,
- [Lienhart 01] [Kim 02] măsura de discontinuitate calculată în diverse spații de culoare HSV, YUV, Lab, etc.
- [Nagasaka 92] divizarea imaginii în blocuri de pixeli:

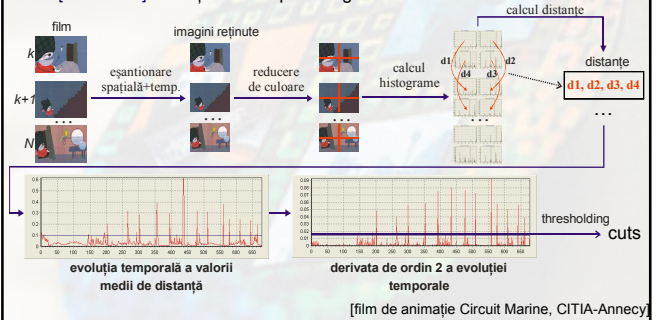
$$D(i) = \sum_{j=0}^{63} \frac{[H_{k,i}(j) - H_{k+l,i}(j)]^2}{H_{k+l,i}(j)}$$

unde i este indicele blocului de pixeli, j este binul histogramei iar $H_{k,i}$ reprezintă histograma blocului i la momentul k .

Detecția tranzițiilor de tip cut

metode bazate pe analiza intensității pixelilor (continuare)

- [Ionescu 07] detecție bazată pe histogramă:

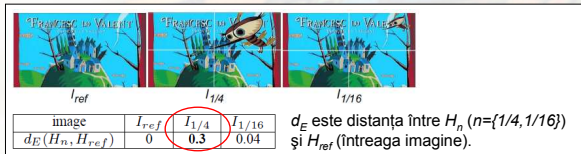


Detecția tranzițiilor de tip cut

metode bazate pe analiza intensității pixelilor (continuare)

- [Ionescu 07] detecție bazată pe histogramă (continuare):

> imaginea este împărțită în 4 regiuni:



> obiectele de dimensiune $> 1/4$ din imagine pot schimba semnificativ valoarea histogramei și astfel pot conduce la detecții false (valori importante ale distanței),

Detecția tranzițiilor de tip cut

metode bazate pe analiza intensității pixelilor (continuare)

- [Ionescu 07] detecție bazată pe histogramă (continuare):

> sunt calculate 4 distanțe (între histogramele corespunzătoare fiecărui bloc) și apoi reținută distanța medie:

$$D_k(i) = \sqrt{\sum_{j=1}^{N_k} [H_{k,i}(j) - H_{k+l,i}(j)]^2}$$

unde i este indicele blocului de pixeli, $i=1, \dots, 4$, j este binul histogramei iar $H_{k,i}$ reprezintă histograma blocului i din imaginea la momentul k .

= funcția de discontinuitate;

- > dacă doar o regiune se modifică semnificativ, prin mediere D_k este puțin afectat;
- > dacă marea parte a imaginii se modifică atunci se modifică semnificativ și D_k ;

Detecția tranzițiilor de tip cut

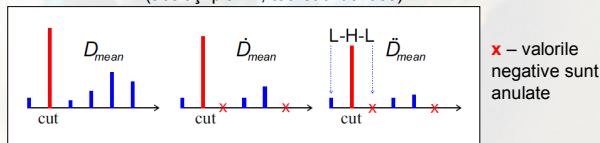
metode bazate pe analiza intensității pixelilor (continuare)

- [Ionescu 07] detecție bazată pe histogramă (continuare):

> derivarea temporală a funcției de discontinuitate:

→ un cut = succesiune de valori "Low-High-Low":

- Low value: distanța dintre ultimele două imagini înainte de cut (același plan 1, imagini teoretic identice),
- High value: distanța între imaginile tranziției cut,
- Low value: distanța dintre următoarele două imagini după cut (același plan 2, teoretic identice).

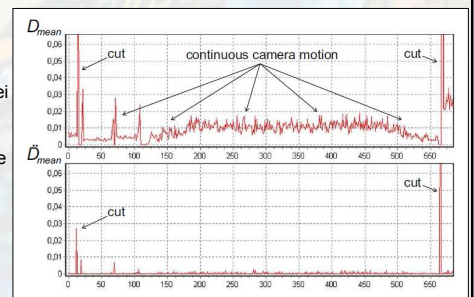


Detecția tranzițiilor de tip cut

metode bazate pe analiza intensității pixelilor (continuare)

- [Ionescu 07] detecție bazată pe histogramă (continuare):

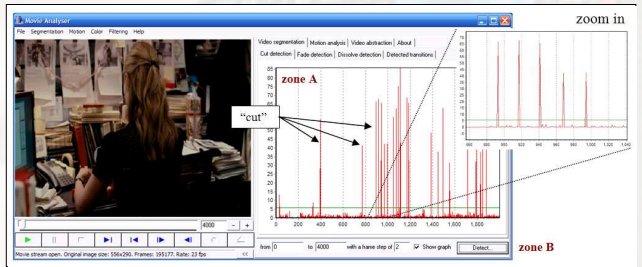
> derivarea temporală a funcției de discontinuitate → diminuarea variațiilor continue temporale generate de mișcare, efecte vizuale, etc.



Detecția tranzițiilor de tip cut

metode bazate pe analiza intensității pixelilor (continuare)

> demonstrație:



[http://alpha.imag.pub.ro/VideoIndexingRP2/]

Detecția tranzițiilor de tip cut

metode bazate pe analiza contururilor

> *ipoteză*: un cut produce o discontinuitate structurală la nivel de imagine (contururile din imaginea dinaintea tranziției nu se vor mai regăsi în imaginea precedentă tranziției)

• [Zabih 95] propune calculul unui raport de schimbare a contururilor din imagine ("Edge Change Ratio"):

$$D_{k,k+l} = ECR_{k+l} = \max \left(\frac{X_k^{out}}{\sigma_k}, \frac{X_{k+l}^{in}}{\sigma_{k+l}} \right)$$

unde σ_k reprezintă numărul de puncte de contur existente în imaginea la momentul k iar X_k^{out} și X_{k+l}^{in} reprezintă numărul de puncte de contur ce au dispărut din imaginea la momentul k și respectiv ce au apărut în imaginea la momentul $k+l$.

Detecția tranzițiilor de tip cut

metode bazate pe analiza contururilor (continuare)

• [Kim 02] folosește raportul de potrivire al contururilor (EMR – Edge Matching Rate);

> metode mai puțin eficiente datorită dependenței de metodele de estimare a **contururilor** ~ sensibile la zgomot;

> informația de contur este folosită de regulă împreună cu alte informații, precum intensitate sau mișcare;

> un avantaj al acestei abordări este dat de faptul că informația de contur poate fi folosită în același timp și pentru detecția tranzițiilor video graduale (fade sau dissolve);

Detecția tranzițiilor de tip cut

metode bazate pe analiza de mișcare

> *ipoteză*: un cut produce o discontinuitate a mișcării, metodele din această categorie bazându-se pe **compensarea** sau **estimarea mișcării**. (detalii M1 – Informația video și M5 – Informația de mișcare)

• [Shahraray 95] imaginile sunt divizate în blocuri disjuncte pentru care se realizează compensarea mișcării folosind ca funcție de cost diferența între intensitățile pixelilor, iar funcția de discontinuitate este calculată astfel:

$$D_{k,k+l} = \sum_{i=1}^{N_b} c_i \cdot d_{k,k+l}(i)$$

unde $d_{k,k+l}$ reprezintă valorile normalizate între 0 și 1 ale funcției de cost minime între imaginile la momentele k și $k+l$, i este indicele blocului de pixeli iar c_i sunt ponderi.

Detecția tranzițiilor de tip cut

metode bazate pe analiza de mișcare (continuare)

• [Porter 00] folosește ca funcție de cost corelația dintre blocurile de pixeli calculată în domeniul frecvențial.

• [Zhong 96] [Lupatini 98] folosesc estimarea fluxului optic iar măsurile de similaritate între imagini sunt calculate pe baza vectorilor de mișcare și a deplasărilor survenite în imagine.

> metode cu o complexitate de calcul importantă datorată estimării mișcării;

> mai puțin precise decât metodele bazate pe histogramme, dar mai eficiente pentru secvențele cu un conținut predominant de mișcare;

Detecția tranzițiilor de tip cut

metode de detecție în domeniul comprimat

> *ipoteză*: discontinuitatea informației vizuale trebuie să se reflecte și la nivelul coeficienților MPEG (DCT).

• [Arman 93] folosește blocuri de pixeli codate DCT din care selectează 64 de coeficienți din componentele alternative $\Rightarrow V_k = (c_1, \dots, c_{64})$:

$$D_{k,k+l} = \frac{V_k \cdot V_{k+l}}{|V_k| \cdot |V_{k+l}|} \quad \text{produs scalar normalizat.}$$

\rightarrow un cut este detectat dacă: $1 - |D_{k,k+l}| > T$ (prag)

> metode ce pot fi implementate cu ușurință în **timp real** deoarece nu necesită decompresia datelor.

Detecția tranzițiilor de tip cut

alte metode

- > folosesc alte surse de informație decât cele descrise anterior:
 - [Boreczky 98] transformă problema detecției într-o problemă de clasificare (modele Markov ascunse HMM);
 - [Hanjalic 02] abordare generică a segmentării (statistică: minimizarea probabilității erorii medii de detecție);
 - [Guimaraes 03] folosește noțiunea de **ritm vizual** al secvenței;
- > sunt abordări în general particulare ce nu au fost studiate suficient.



Detecția tranzițiilor de tip cut

estimarea pragului de detecție

- > similaritatea a două imagini se reduce în cele din urmă la compararea unei măsuri de distanță cu un anumit prag τ :
 - $> \tau \Rightarrow$ sunt disimilare,
 - $< \tau \Rightarrow$ sunt similare.
- > alegerea pragului:
 - ↳ A. **metode euristice** ([Otsuji 91] [Nagasaka 92] [Arman 93])
 - > pragul este ales ca rezultat al expertizei manuale a datelor folosite. **[nu este valabil pentru orice secvență!]**
 - ↳ B. **metode statistice** ([Zhang 93])
 - > funcția de discontinuitate este modelată cu o distribuție Gaussiană
 $\rightarrow \tau = \mu + r \cdot \sigma$ **[adaptat global!]**

Detecția tranzițiilor de tip cut

estimarea pragului de detecție (continuare)

- > alegerea pragului (continuare):
 - ↳ C. **metode adaptive** ([Yeo 95])
 - > valoarea pragului este adaptată la conținutul local al secvenței (fereastră de analiză), exemplu un cut este detectat dacă:

$$D_{k,k+1} = \max_{i \in W} \{D_{k+i,k+i+1}\} \text{ și } D_{k,k+1} \geq \alpha \cdot D_{s \max}$$
 unde $D_{k,k+1}$ este funcția de discontinuitate, w este fereastra de analiză, α este un parametru iar $D_{s \max}$ reprezintă a doua valoare maximă a funcției de discontinuitate.
 [care este dimensiunea ferestrei pentru a fi relevantă?]

Detecția tranzițiilor de tip cut

estimarea pragului de detecție (continuare)

- > alegerea pragului (continuare):
 - ↳ D. **metode mixte** ([Hanjalic 97])
 - > combină abordările de mai sus, de exemplu alegerea parametrului α folosit la metoda adaptivă pe baza modelării statistice cu funcții Gausiene.
 - ↳ E. **metode optimale** ([Vasconcelos 00])
 - > inspirate de teoria detecției statistice, folosesc informații statistice despre distribuția tranzițiilor cut obținute în urma experimentărilor pe un număr semnificativ de secvențe. **[setul de experimentare trebuie să fie suficient de diversificat!]**

> M4. Structura temporală [Detecția de “fades”]

Detecția tranzițiilor de tip fade

[Lienhart 01]

> formal un fade este definit ca fiind transformarea intensităților pixelilor secvenței $S_i(x,y,t)$ printr-o funcție monotonă $f(t)$:

$$F(x,y,t) = f(t) \cdot S_i(x,y,t), \quad 0 \leq t \leq T$$

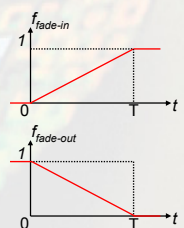
unde (x,y) sunt coordonate spațiale, t coordonata temporală, $F(x,y,t)$ reprezintă secvența de fade, T reprezintă durata tranziției iar $f(t)$ este definită astfel:

• fade-in:

$$f_{\text{fade-in}}(t) = \frac{t}{T}, \quad f(0) = 0 \text{ si } f(T) = 1$$

• fade-out:

$$f_{\text{fade-out}}(t) = 1 - \frac{t}{T}, \quad f(0) = 1 \text{ si } f(T) = 0$$



Detecția tranzițiilor de tip fade

→ un fade conduce la modificarea intensității luminoase a imaginii !



> metodele de detecție existente:

- ↳ bazate pe analiza intensității pixelilor,
- ↳ bazate pe analiza contururilor,
- ↳ bazate pe alte surse de informație,
- ↳ abordări mixte.

Detecția tranzițiilor de tip fade

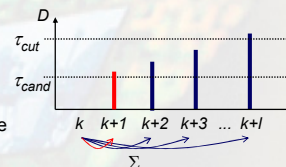
metode bazate pe analiza intensității pixelilor

> o primă abordare pornește de la detecția tranzițiilor de tip cut și folosește două praguri:

- [Zhang 93] dacă $D_{k,k+1} > \tau_{cand}$ și $D_{k,k+1} < \tau_{cut}$ ($\tau_{cand} < \tau_{cut}$)
⇒ imaginea la momentul k este o potențială imagine de început a unei tranziții graduale;

> această imagine este comparată cu imaginile următoare, iar distanța între acestea acumulată;

> în momentul în care nu mai crește monoton, dacă cumulat $> \tau_{cut}$
⇒ tranziție graduală.



Detecția tranzițiilor de tip fade

metode bazate pe analiza intensității pixelilor (continuare)

> o abordare uzuală constă în analiza schimbării intensității luminoase, exemplu analiza dispersiei acesteia:

$$\sigma\{F(x, y, t)\} = f(t) \cdot \sigma\{S_1(x, y)\}$$

(în condiții de ergodicitate: momentele statistice = momentele temporale) unde σ reprezintă dispersia, $F(x, y, t)$ este secvența de fade iar $S_1(x, y)$ reprezintă secvența inițială.

- [Lienhart 99] sunt localizate imaginile monocromatice și este analizată creșterea pozitivă a intensității luminoase și a dispersiei acesteia în sensul pozitiv al axei temporale;

- [Alattar 97] determină punctele de extrem negativ ale derivatei a doua a valorilor varianței intensității luminoase,
→ fade dacă valoarea derivatei întâi a intensității luminoase medii este constantă între două puncte de extrem negativ;

Detecția tranzițiilor de tip fade

metode bazate pe analiza intensității pixelilor (continuare)

> folosirea informației de cromatică:

- [Fernando 99] definește următorul parametru în spațiul YC_bC_r :

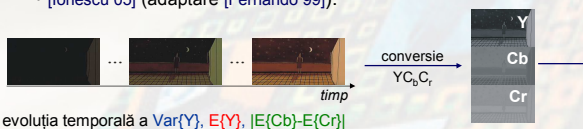
$$R_k = \begin{cases} \frac{\Delta_k^Y}{\Delta_k^C} & k < L_1 \text{ sau } k \geq (L_1 + T) \\ \left| \frac{C_0 - m_{k+1}^Y + (L_1 - k) \cdot \Delta_k^Y}{C_0 - m_{k+1}^C + (L_1 - k) \cdot \Delta_k^C} \right| & L_1 \leq k < (L_1 + T) \end{cases}$$

unde Δ_k^Y și Δ_k^C reprezintă schimbările incrementale ale mediei semnalului Y și respectiv ale semnalului $C = (C_b + C_r)/2$, m_{k+1}^Y și m_{k+1}^C sunt valorile medii ale lui Y și C calculate la momentul $k+1$, L_1 reprezintă momentul de început al tranziției, T este durata totală a acesteia, iar C_0 reprezintă nivelul semnalului video pentru începutul tranziției → fade = valori ~ constante ale R_k

Detecția tranzițiilor de tip fade

metode bazate pe analiza intensității pixelilor (continuare)

- [Ionescu 05] (adaptare [Fernando 99]):



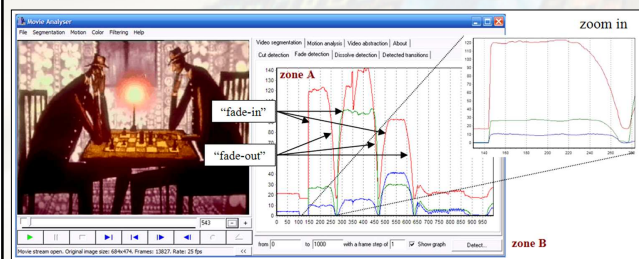
decizie (fade-in):
- $\text{Var}\{Y\} \approx 0$ (start)
- $E\{Y\} \nearrow$ sau
- $|E\{Cb\} - E\{Cr\}| \nearrow$
- durată $\in [t_{\min}, t_{\max}]$
(fade-out sens invers)

[film de animație Casa, CITIA-Anneey]

Detecția tranzițiilor de tip fade

metode bazate pe analiza intensității pixelilor (continuare)

> demonstrație:



[http://alpha.imag.pub.ro/VideoIndexingRP2/]

Detecția tranzițiilor de tip fade

metode bazate pe analiza contururilor

> *ipoteză*: pe durata unei tranziții de tip fade contururile obiectelor din imagine fie dispar (fade-out), fie apar gradual (fade-in).

- [Zabih 99] folosește mărimea ECR (vezi detecția de cut), pe durata unei tranziții de tip fade-in, numărul de pixeli de contur ce apar în imagine, $ECR_{in} >$ numărul pixelilor de contur ce dispar din imagine, ECR_{out} ($ECR_{in} > ECR_{out}$);

> pe durata unui fade-out este invers $ECR_{in} < ECR_{out}$.

- [Yu 97] [Lupatini 98] abordări similare, contabilizarea punctelor de contur ce apar sau dispar.

> ca și în cazul detecției de cut, aceste metode sunt mai sensibile la mișcare și schimbări ale intensității din imagine.

Detecția tranzițiilor de tip fade

metode bazate pe alte surse de informație

> de exemplu:

- [Porter 01] se folosește de informația de mișcare pentru a reduce falsele detecții;
- [Guimaraes 03] se folosește de ritmul vizual bazat pe histogramă pentru a reduce influența zgomotului;
- [Miene 01] parametri sunt calculați în domeniul frecvențial al coeficienților FFT;
- [Heng 01] o abordare generică statistică a detecției tranzițiilor graduale.



> valabilă concluzia de la detecția de cut: sunt abordări particulare nu suficient de robuste.

> M4. Structura temporală [Detecția de “dissolves”]

Detecția tranzițiilor de tip dissolve

[Lienhart 01]

> formal, secvența unui dissolve $D(x, y, t)$ de durată T este definită pe baza secvențelor $S_1(x, y, t)$ și $S_2(x, y, t)$ astfel:

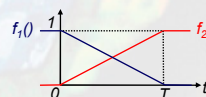
$$D(x, y, t) = f_1(t) \cdot S_1(x, y, t) + f_2(t) \cdot S_2(x, y, t)$$

unde (x, y) sunt coordonate spațiale și t coordonata temporală, $t=0, \dots, T$ iar f_1 și f_2 sunt două funcții.

> în funcție de $f_1(t)$ și respectiv $f_2(t)$ întâlnim:

- **cross-dissolve** (suprapunere fade-out cu fade-in):

$$f_1(t) = 1 - \frac{t}{T}, \quad f_2(t) = \frac{t}{T} \quad 0 \leq t \leq T$$



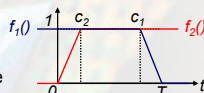
Detecția tranzițiilor de tip dissolve

> în funcție de $f_1(t)$ și respectiv $f_2(t)$ întâlnim (continuare):

- **additive-dissolve** (suma unui fade-out cu fade-in):

$$f_1(t) = \begin{cases} 1 & t \leq c_1 \\ \frac{T-t}{T-c_1} & \text{altfel} \end{cases} \quad f_2(t) = \begin{cases} \frac{t}{c_2} & t \leq c_2 \\ 1 & \text{altfel} \end{cases}$$

unde $c_1, c_2 \in (0; T)$, $c_2 < c_1$ și $0 \leq t \leq T$.



> d.p.d.v. al diferenței vizuale dintre secvențele $S_1(x, y, t)$ și $S_2(x, y, t)$, avem următoarele situații:

- S_1 și S_2 au distribuții de culoare suficient de diferite
→ ușor de confundat cu o tranziție cut;

Detecția tranzițiilor de tip dissolve

> d.p.d.v. al diferenței vizuale dintre secvențele $S_1(x, y, t)$ și $S_2(x, y, t)$, avem următoarele situații (continuare):

- S_1 și S_2 au distribuții de culoare similare
→ greu de detectat pe baza discontinuității vizuale dintre histograme, + diferențe structurale importante;
- S_1 și S_2 au distribuții de culoare + structură similare (caz particular de morphing);

> metodele de detecție existente:

- ↳ bazate pe analiza intensității pixelilor,
- ↳ bazate pe trăsături (feature-based),
- ↳ bazate pe transformări,
- ↳ bazate pe alte surse de informație.

Detecția tranzițiilor de tip dissolve

metode bazate pe analiza intensității pixelilor

> dacă mișcarea este neglijabilă, atunci pentru un cross-dissolve:

$$\frac{\partial D(x, y, t)}{\partial t} = \frac{S_2(x, y) - S_1(x, y)}{T}, \quad 0 \leq t \leq T$$

• [Hampapur 95] localizarea în secvență a tuturor schimbărilor liniare ale intensității pixelilor.

> pornind de la definiția matematică putem scrie altfel:

$$Var\{D(x, y, t)\} = f_1^2(t) \cdot Var\{S_1(x, y)\} + f_2^2(t) \cdot Var\{S_2(x, y)\}$$

unde S_1 și S_2 sunt independente de timp.

$$\Rightarrow Var\{D(x, y, t)\} = \frac{(T-t)^2}{T^2} \cdot Var\{S_1(x, y)\} + \frac{t^2}{T^2} \cdot Var\{S_2(x, y)\}$$

Detecția tranzițiilor de tip dissolve

metode bazate pe analiza intensității pixelilor (continuare)

$$Var\{D(x, y, t)\} = \frac{(T-t)^2}{T^2} \cdot Var\{S_1(x, y)\} + \frac{t^2}{T^2} \cdot Var\{S_2(x, y)\}$$

$$\Rightarrow Var\{D(x, y, t)\} = c \cdot (t-a)^2 - b$$

unde a , b , și c sunt constante:

$$a = \frac{T \cdot Var\{S_1(x, y)\}}{Var\{S_1(x, y)\} + Var\{S_2(x, y)\}}$$

$$b = \frac{Var\{S_1(x, y)\} \cdot Var\{S_2(x, y)\}}{Var\{S_1(x, y)\} + Var\{S_2(x, y)\}}$$

$$c = \frac{Var\{S_1(x, y)\} + Var\{S_2(x, y)\}}{T^2}$$

Detecția tranzițiilor de tip dissolve

metode bazate pe analiza intensității pixelilor (continuare)

$$Var\{D(x, y, t)\} = c \cdot (t-a)^2 - b$$

$$\Rightarrow \frac{\partial Var\{D(x, y, t)\}}{\partial t} = 2 \cdot c \cdot (t-a) \Rightarrow \frac{\partial^2 Var\{D(x, y, t)\}}{\partial t^2} = 2 \cdot c$$

> pe parcursul unei tranziții de tip dissolve evoluția temporală a varianței intensității pixelilor are un comportament parabolic:

→ derivata secundă ~ 0 (înainte și după dissolve) + valori constante pozitive pe durata tranziției.

• [Alattar 93] exploata prezența a două puncte de extrem negativ în derivata secundă (corespund începutului și sfârșitului tranziției dissolve).

Detecția tranzițiilor de tip dissolve

metode bazate pe analiza intensității pixelilor (continuare)

• [Truong 00] folosește următoarele considerații:

- derivata întâi a varianței intensității pixelilor pe parcursul unui cross-dissolve trebuie să fie monoton crescătoare (valoare negativă → valoare pozitivă),

- varianța intensității pixelilor pentru cele două secvențe S_1 și S_2 trebuie să fie superioară unui anumit prag,

- durata unui dissolve nu trebuie să depășească un anumit interval de valori $[T_{min}; T_{max}]$.

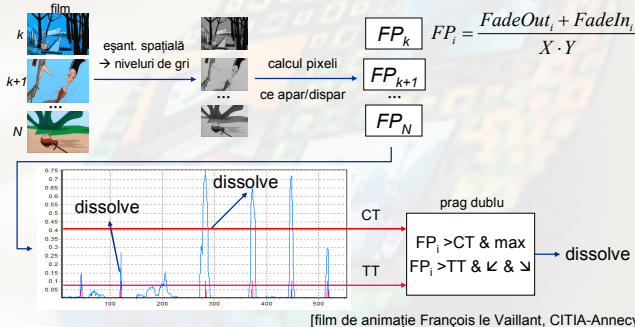
> metode ce nu pornesc de la definiția matematică:

• [Su 05] analizează pixelii ce prezintă o creștere / descreștere monotonă a intensității luminoase pentru o anumită fereastră de analiză, ∇ dissolve.

Detecția tranzițiilor de tip dissolve

metode bazate pe analiza intensității pixelilor (continuare)

• [Ionescu 10] (adaptare [Su 05] [Zhang 93]):



Detecția tranzițiilor de tip dissolve

metode bazate pe analiza intensității pixelilor (continuare)

• [Ionescu 10] (adaptare [Su 05] [Zhang 93]):

> detecție folosind un prag dublu:

A. dacă $FP_k > \tau_{CT}$ (Certain Threshold) și FP_k este maxim local, → se declară dissolve în $[k - t_{max}/2; k + t_{max}/2]$ (t_{max} reprezintă durata maximă a unui dissolve).

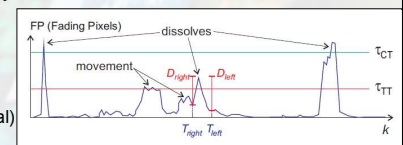
B. dacă $FP_k > \tau_{TT}$ (Tolerant Threshold) dar $FP_k < \tau_{CT}$, atunci k este potențial imaginea de mijloc a unui dissolve,

→ dacă $D_{left} > 0.5FP_k$

$$\text{și}$$

$$D_{left} > 0.5FP_k$$

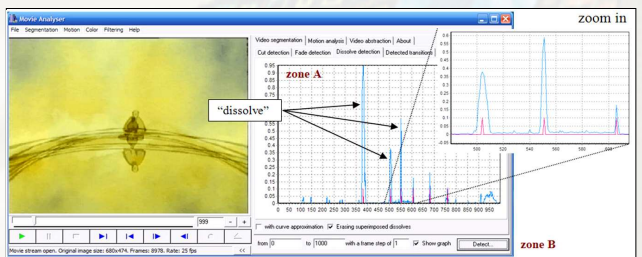
→ dissolve (adaptare la formă semnal)



Detecția tranzițiilor de tip dissolve

metode bazate pe analiza intensității pixelilor (continuare)

> demonstrație:



[http://alpha.imag.pub.ro/VideoIndexingRP2/]

Detecția tranzițiilor de tip dissolve

metode bazate pe trăsături

> ipoteză: pe durata tranziției contururile obiectelor din imaginea de început dispar progresiv în timp ce contururile obiectelor din imaginea finală apar progresiv.

- [Zabih 99] folosește raportul ECR (pe parcursul tranziției acesta are valori semnificative);
- [Lienhart 99] definește contrastul conturilor, EC (edge contrast):

$$EC(I_E) = 1 + \frac{s(I_E) - w(I_E) - 1}{s(I_E) + w(I_E) + 1}$$

unde $I_E(x, y, k)$ reprezintă harta de contururi a imaginii curente la momentul k , $s(I_E)$ și $w(I_E)$ reprezintă numărul total de puncte de contur semnificative și respectiv "slabe".

→ dissolve = minim local EC mărginit de variații abrupte.

Detecția tranzițiilor de tip dissolve

metode bazate pe trăsături (continuare)

- SIFT - Scale-Invariant Feature Transform, SURF - Speeded Up Robust Features, etc.

metode bazate pe transformări

> detecția este realizată într-un domeniu transformat, de regulă în spațiul FFT sau DCT:

- [Gu 97] calculează procentul de blocuri de pixeli pentru care diferența absolută între coeficienții DCT rămâne superioară unui anumit prag pe durata a 10 până la 60 de imagini (durata medie max. dissolve).
- [Boccignone 00] dissolve când numărul de blocuri de pixeli ce prezintă o modificare semnificativă a coeficienților DCT precum și gradul stocastic (aleator) al vectorilor de mișcare > prag.

> avantaje implementare timp real.



Detecția tranzițiilor de tip dissolve

metode bazate pe alte surse de informație

> similar detecției de cut și fade:

- [Boreczky 98] structura secvenței este modelată cu rețele Markov ascunse (HMN);
- [Park 05] folosește spectrul ritmului vizual (Visual Rhythm Spectrum);
- [Das 08] folosește histograme fuzzy și matrice de co-ocurență fuzzy;

> metode mai puțin specializate care de cele mai multe ori abordează detecția mai multor tipuri de tranziții graduale simultan.

concluzii:

- metodele de detecție de cut destul de robuste, rate uzuale ~ 95%;
- detecția de fade și în special de dissolves mult mai complexă, rate de detecție medii [70-80]%;
- tendință dezvoltare metode de detecție generică a pasajelor de tranziție.

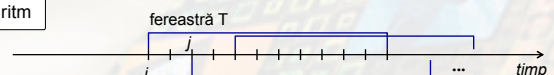


> M4. Structura temporală
[Activitate inter-plan]

Activitate inter-plan: ritm

> ipoteză: conținutul de acțiune este în general corelat cu o rată semnificativă de schimbări de plan,

ritm



$\xi_T(i)$ numărul de schimbări de plan ce au loc în fereastra temporală de dimensiune T (ex. 5s) începând cu cadrul la momentul i ;

> evaluând valorile lui $\xi_T(i)$ pentru întreaga secvență (ex. ferestre suprapuse cu un pas 1), determinăm ritmul vizual mediu al secvenței:

$$\bar{v}_T = E\{\xi_T(i)\} \sim \text{tempo mediu.}$$

Activitate inter-plan: acțiune

$\bar{v}_T = E\{\xi_T(i)\}$, exemple:



documentare

știri

muzică

$$\bar{v}_T = 0.2^*$$

$$\bar{v}_T = 0.24^*$$

$$\bar{v}_T = 0.56^*$$

[* valori medii pentru 30 de exemple de secvențe.]

acțiune : corelație $\xi_T(i)$ - conținut real de acțiune → teste subiective,

- "hot action": conținut de acțiune ridicat (schimbări rapide, mișcare rapidă, efecte vizuale, ...),
- "regular action": acțiune medie (scene de dialog, evenimente, ...),
- "low action": scene statice.

Activitate inter-plan: acțiune

> corelație $\xi_T(i)$ - conținut de acțiune → teste subiective, caz particular filme de animație (Festivalul Internațional al Filmului de Animație, Annecy):

Movie	Segment (frames)	Length (s)	\bar{v}_T
Low action	633-1574	38	0.31
	6945-8091	46	0.37
François le Vaillant	4257-6523	91	0.18
	6898-7683	31	0.38

Action Type	Hot Action	Regular Action	Low Action
$E\{\xi_T\}$	3.65	1.97	0.48
σ_{ξ_T}	0.85	0.29	0.23
Interval	2.8-∞	1.67-2.26	0.25-0.71

ground truth

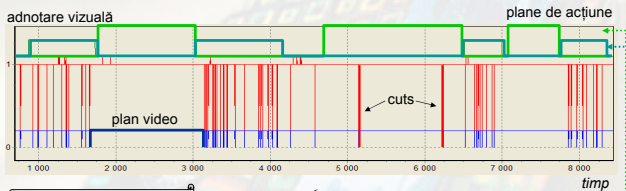
[B. Ionescu, L. Ott, P. Lambert, D. Coquin, A. Pacureanu, V. Buzuloiu, SPIE - Journal of Electronic Imaging, 2010]

→ localizarea segmentelor pentru care numărul de schimbări într-o fereastră de durată T se încadrează în intervalele de mai sus,

$$[E\{\bar{v}_T\} - \sigma_{\bar{v}_T}; E\{\bar{v}_T\} + \sigma_{\bar{v}_T}]$$

Activitate inter-plan: acțiune

> determinare plane de acțiune:



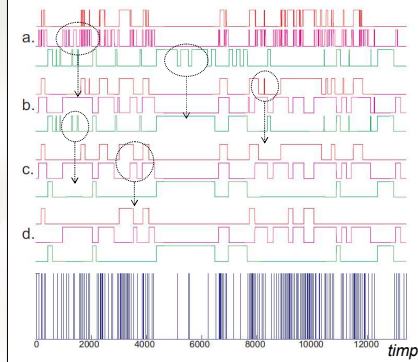
a. thresholding: marcarea cadrelor (val. 1) care fac parte dintr-un anumit conținut de acțiune.

$$HA_{shot}(i) = \begin{cases} 1 & \text{if } \xi_{T=5s}(i) > 2.8 \\ 0 & \text{otherwise} \end{cases}$$

$$LA_{shot}(i) = \begin{cases} 1 & \text{if } \xi_{T=5s}(i) < 0.7 \\ 0 & \text{otherwise} \end{cases}$$

Activitate inter-plan: acțiune

> determinare plane de acțiune (continuare):



b. merging: segmente vecine la distanță $< T$ s sunt fuzionate (eliminarea golurilor, ~ eliminare supra-segmentare).

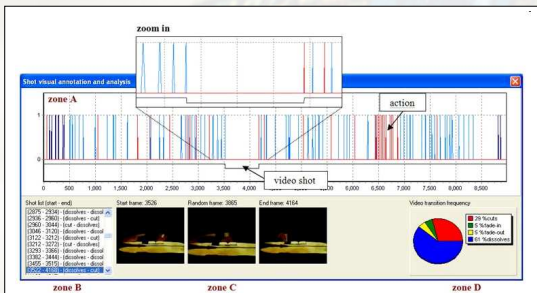
c. pruning: segmente mici cu durată $< T$ s sunt eliminate (eliminarea segmentelor irelevante și ne-perceptibile).

d. restoring compl.: segmentele de acțiune pentru diverse clase se suprapun → și logic.

$$action = \frac{T_{action}}{T_{total}}$$

Activitate inter-plan

> demonstrație: detecție plane video, adnotare vizuală structură temporală, activitate inter-plan, activitate intra-plan, conținut de culoare,



[http://alpha.imag.pub.ro/VideoIndexingRP2/]

> Sfârșit M4. Structura temporală