

Algoritmi pentru indexarea multimedia

Lect. Georgeta DRULA

Universitatea Bucuresti

Alegerea unui model de date fundamental permite automatizarea proceselor în sistemele multimedia, folosind diferite categorii de algoritmi. Se poate constata ca acesti algoritmi sunt în principal recursivi, datorita structurii arborescente oferite de catre modele si se bazeaza pe continut, pe structura sau pe context. Luarea în considerare a contextului în procesele de cautare a informatiei multimedia a determinat necesitatea ca algoritmii de regasire bazati pe continut sa fie însoțiti si de procese de navigare.

Datorita complexitatii naturii datelor multimedia, algoritmii necesari analizei mediilor se aplica pe mai multe niveluri corespunzatoare gradelor de înțelegere a informatiei. Acestea determina de asemenea mai multe niveluri de indexare care se vor folosi apoi în procesul de regasire.

Cuvinte cheie: *analiza bazata pe continut pentru video, indexarea video, analiza si indexare a mediilor pe structura, analiza si indexare a mediilor bazata pe continut, indexare multimedia, M-tree.*

Indexarea multimedia bazata pe continut pentru imagini si video

Descrierea bazata pe continut a imaginii se realizeaza în principal prin recunoasterea si interpretarea unor caracteristici esentiale ale tipului de mediu, cum ar fi: culoarea, textura, forma, locatia spatiala, regiunile de interes, caracteristicile faciale, cadrele cheie, detectarea modificarilor de scena. Recunoasterea automata a caracteristicilor importante ale unui mediu necesita un algoritm corespunzator, ale caror criterii de optimizare sunt atât micsorarea costurilor, dar si asigurarea unei indexari consistente. Algoritmii bazati pe continut desi împart câteva trasaturi comune, au fiecare un aspect propriu dependent de mediul tratat. Recunoasterea bazata pe continut, complet automatizata este departe de a se realiza. Tendinta actuala care ar furniza rezultate mai bune ar fi combinarea metodelor baza-

te pe text cu cele bazate pe continut pentru descrierea imaginii si a miscarii.

Algoritmul Griffioen propune pentru analiza si indexarea secventelor video digitale compresate tehnici statistice. El asigura extragerea si codificarea automata a informatiei semantice din acestea si se bazeaza pe metoda de analiza în componente principale pentru a extrage caracteristicile secventei video. Se face analiza cadru de cadru si se extrag caracteristici (obiecte vizuale din imagine) prin identificarea de forme dominante, cum ar fi: fatetele, formele, texturile pe baza unor sabloane deja existente. Aceste caracteristici extrase se clasifica pe baza criteriilor de similaritate. Algoritmul este capabil sa controleze fidelitatea redarii secventelor.

Modalitatile clasice de analiza semantica si indexare a secventelor video se bazeaza pe decodificarea fiecarui cadru video, apoi cautarea de pixeli corespunzatori unui obiect

imagine pentru clasificarea si apoi recunoasterea lui. Metoda Griffioen poate clasifica obiectele dintr-un video direct în forma compresata, nefiind nevoie de decompresarea video la nivel de pixel. Evitarea acestei operatii mareste eficienta de calcul fara a se pierde din capacitatea de clasificare. Alte metode de analiza si indexare video se orienteaza spre identificarea cadrelor cheie din planuri si segmente video. Sunt necesare cel putin doua astfel de cadre cheie (algoritmii Nagasaka; Zhang, Low si Smoliar).

Metoda lui Griffioen si Seales [GrSe95] propune extragerea automata a informatiei de continut din imagini fixe si din secvente video în formate standard JPEG, M-JPEG, MPEG, folosite pentru captarea, stocarea si transportul acestora. Identificarea obiectelor vizuale din imagini se bazeaza pe spatiul propriu, determinat prin analiza în componente principale care a fost propusa ca solutie în diferite contexte. Extragerea de obiecte se realizeaza pe mai multe niveluri prin trecerea imaginii sau a secvenței video printr-o serie de transformari care sa identifice informatia semantica importanta a unui obiect. Aceste transformari pornesc de la considerentul ca semantica asociata cu datele din obiect se modifica pe durata de viata a acestuia. De aceea, este important sa se modifice setul de functii (operatii) asociate cu un obiect, dupa ce el a fost instantiat. În acest fel, setul de functii asociate obiectului depinde de înțelesul semantic curent al obiectului.

De exemplu, pentru extragerea caracteristicilor pe mai multe nivele ale unei imagini fixe se considera initial, înțelesul de “ima-

gine în ansamblu” pe care se pot executa functiile de *marire*, *segmentare* si *detectarean laturilor*. Dupa executia functiei de *detectare_laturi*, înțelesul semantic al imaginii se modifica în “**imagine din laturi**”. Pentru acest înțeles sunt necesare alte functii de prelucrare corespunzatoare, care sa identifice acele laturi, care pot reprezenta de exemplu trasee în imagine. Cu cât semantica imaginii evolueaza, cu atât setul de functii asociate obiectului vor deveni mai specifice si mai puternice. De asemenea, functiile care nu mai sunt necesare pot fi eliminate. Pentru exemplul cu functiile ce determina laturi în imagine exista numeroase tehnici din grafica interactiva ce pot realiza acest lucru. Fiecare tehnica implementeaza un algoritm nou, diferit, care este potrivit în anumite circumstante. Din acest punct de vedere, este mai avantajos sa se lege datele imagine de o clasa de functii abstracta, din care va fi aleasa o functie în mod dinamic, la momentul executiei.

Imaginea trece printr-o serie de transformari pentru a i se extrage caracteristicile principale si a fi reprezentata într-o forma codificata, de obiecte imagine. Obiectul imagine codificat contine nu numai datele curente si semantica lor asociata, dar si secventa de operatii ce este aplicata obiectului original pâna ajunge în starea curenta. Algoritmul amintit a fost implementat de catre autori într-un sistem prototip, ce permite utilizatorului sa defineasca *grupuri de functii abstracte*, *clase de date semantice* si un *graf de procesare semantica* ce este utilizat pentru extragerea interactiva a informatiei incluse în imagine. Sistemul con-

tine de asemenea si instrumente ce permit recunoasterea caracteristicilor imagine, a muzicii, identificarea documentului dupa autor si titlu.

Aplicarea analizei în componente principale pentru video se concretizeaza prin reprezentarea unui numar mare de imagini cheie "sablon" pe un sub-spatiu, mai mic dimensional. O imagine noua este clasificata prin similaritatea sau disimilaritatea cu o imagine cheie "sablon", prin câteva operatii în subspatiul precalculat. Cum imaginile cheie "sablon" sunt extrase folosind transformarea Karhunen-Loeve, orice imagine noua poate fi clasificata dupa un anumit criteriu.

Pasii algoritmului sunt:

I. Datele de intrare sunt un set de imagini reprezentate ca vectori, unde fiecare imagine este reprezentata la rândul ei prin vectori de pixeli. Vectorul de pixeli contine valorile pixelilor în ordinea furnizata de linia de scanare.

$$\mathbf{f} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k\} \quad (1)$$

unde:

f = setul de imagini, fiecare cu câte $n = x * y$ pixeli.

Pentru fiecare imagine-cadru analizata, transformarea Karhunen-Loeve extrage valoarea medie a fiecarui vector de intrare, $\mathbf{m}$.

$$\mathbf{m} = (1/k) * \sum \mathbf{f}_i \quad (2)$$

Din acest pas rezulta un nou set vector $\mathbf{f}' = \{\mathbf{f}'_1, \mathbf{f}'_2, \dots, \mathbf{f}'_k\}$, unde $\mathbf{f}'_i = \mathbf{f}_i - \mathbf{m}$.

II. Se calculeaza matricea de covarianța a vectorilor de intrare ajustati.

$$\mathbf{C}_{n \times n} = \mathbf{U} \mathbf{U}^T \quad (3)$$

III. Transformarea Karhunen-Loeve se obtine în spatiul si din valorile proprii ale

imaginii, prin descompunerea structurii ei. Vectorii imagine sunt complet ortogonali si formeaza o baza ce împarte spatiul N -dimensional. Astfel, folosindu-se vectorii proprii ca baza, se descriu vectorii de intrare initiali $\mathbf{f}_i$:

$$\mathbf{f}_i = \sum \mathbf{p}_{ij} \mathbf{A}_j + \mathbf{m} \quad (4)$$

IV. O proprietate importanta a spatiului propriu este ca vectorii pot fi ordonati conform valorilor lor asociate. Vectorul cel mai mare este si cel mai semnificativ, codificând cea mai mare variatie a setului de vectori de intrare. Aceasta ordonare permite vectorilor originali de intrare sa fie aproximati de primii w vectori $\{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_w\}$ cu $w < k$:

$$\mathbf{f}_i \approx \sum \mathbf{p}_{ij} \mathbf{A}_j + \mathbf{m} \quad (5)$$

Fiind dati doi vectori care se proiecteaza pe spatiul propriu, cu cât mai apropiate sunt proiectiile lor pe acest spatiu cu atât vectorii initiali sunt mai puternic corelati. Distanțele în spatiul propriu sunt o aproximare a corelatiilor în domeniul imagine. Rezultatul este ca spatiul propriu furnizeaza o modalitate puternica de clasificare a unei imagini în raport cu un set de imagini cunoscute. Pentru aceasta este suficient sa se proiecteze imaginea în spatiul propriu si sa se masoare distanta euclidiană la alte puncte din spatiu.

Indexarea multimedia bazata pe structura

Natura multidimensională a caracteristicilor obiectelor multimedia a sugerat pentru indexarea multimedia bazata pe structura folosirea *metodelor de acces spatial* (SAMS – *Spatial Access Methods*) cum sunt: *R-Tree*, *B-Tree*, *K-Tree* si variantele

sale, potrivite pentru a formula cererile de similaritate în indexarea și regasirea datelor. Cu toate că eficacitatea arborilor metrici a fost deja demonstrată, proiectanții bazelor de date acuza aceste metode că sunt intrinsec statice ceea ce nu permite aplicarea lor în medii dinamice. Pentru a rezolva această limită se propune metoda de indexare **M-Tree** care este un arbore metric ce a fost special proiectat ca metoda dinamică de indexare și acces pentru bazele de date multimedia.

Metoda de indexare **M-Tree** este rezultatul combinării avantajelor metodelor de acces spațial (SAMs) cu capacitățile arborilor metrici statici și se aplică pentru indexarea obiectelor multimedia folosind caracteristicile și funcțiile distanță ce nu se încadrează într-un spațiu vector și sunt restricționate doar de postulatele metrice.

Conform acestei arborescente, obiectele sunt stocate în noduri de dimensiune fixă ce corespund regiunilor partitionate ale spațiului metric. Distanțele relative dintre obiecte se măsoară de către o funcție distanță specifică, d , arborescenta *M-Tree* fiind complet parametrizată pe distanță d .

Obiectele indexate se stochează în noduri de dimensiune fixă, noduri “frunza” ale arborescentei, și sunt reprezentate prin intrări cu formatul:

$$[v_j, oid(v_j)],$$

unde V_j sunt valorile caracteristice ale obiectului cu identificator $oid(v_j)$.

Nodurile interne ale structurii *M-Tree* stochează așa numitele obiecte “de rutare” cu format:

$$[v_r, r(v_r), ptr(N)],$$

unde:

v_r este valoarea caracteristică a obiectului de rutare,

$r(v_r) > 0$ este raza de acoperire și

$ptr(N)$ este pointerul la un nod copil N .

Semantica razei de acoperire pentru un obiect de rutare (intern) este aceea că fiecare obiect v din subarborele ce porneste din N , satisface restricția

$$d(v_r, v) \leq r(v_r).$$

În ideea de a crește performanța în timpul căutării, fiecare intrare în arbore stochează și distanța dintre valoarea caracteristicii acelei intrări și obiectul de rutare părinte.

Arborescenta *M-Tree* se dezvoltă de jos în sus (*bottom-up*). Ramificația unui nod N este gestionată prin alocarea unui nou nod N_t , la același nivel cu N , prin partitionarea intrărilor dintre cele două noduri și apoi prin promovarea informației relevante către nodul părinte N_p . Când rădăcina se divizează, o nouă rădăcină este creată și *M-Tree* crește cu un nivel.

Spre deosebire de celelalte arborescente ce se bazează pe un singur criteriu specific de organizare a obiectelor, *M-Tree* oferă posibilitatea de implementare a politicilor de partitionare alternative în funcție de necesitățile aplicației.

Procesul de indexare și căutare prin arborescenta *M-Tree* se bazează pe două tipuri de cereri de similaritate, pe domeniu și pe vecinătăți.

- Similaritate pe domeniu: Fiind dat un obiect cerere v_q și o distanță de căutare maximă $r(v_q)$ se selectează toate obiectele indexate O_j astfel încât:

$$d(O_j, v_q) \leq r(v_q);$$

Se considera o cerere domeniu $(v_q, r(v_q))$ ce definește o regiune cerere centrata în v_q . În ideea de a regăsi toate obiectele indexate din interiorul regiunii cerere se accesează doar nodurile N care intersectează această regiune. Această condiție este exprimată prin inegalitatea următoare:

$$d(v_r, v_q) \leq r(v_r) + r(v_q)$$

unde v_r este obiect de rutare al nodului N .

Se calculează distanța $d(v_r, v_q)$ pentru a se vedea dacă nodul N va fi accesat. Calculurile pot fi simplificate prin folosirea distanțelor precalculate și stocate. Astfel nodul N poate fi eliminat din căutare dacă se verifică inegalitatea:

$$\frac{1}{2}d(v_p, v_q) - d(v_r, v_p)^{\frac{1}{2}} > r(v_r) + r(v_q),$$

unde

$d(v_p, v_q)$ – distanța dintre obiectul cerere și obiectul părinte al nodului;

$d(v_r, v_p)$ – distanța calculată și stocată ca parte din intrarea v_r .

Această inegalitate este o consecință directă a inegalității triunghiului aplicate pentru $d(v_r, v_q)$.

- Similaritate pe vecinătăți: Fiind dat un obiect cerere v_q și un întreg $k \geq 1$ se selectează obiectele indexate k ce au cea mai scurtă distanță de la v_q .

Fata de similaritatea pe domeniu diferența este distanța prag $r(v_q)$, care este o distanță dinamică, deoarece este dată de distanța celui mai apropiat k vecin al valorii cerere v_q . Algoritmul care procesează cererile celor mai apropiați vecini implementează o politică de decizie a ordinii de vizitare a nodurilor care nu au fost încă eliminate.

Bibliografie:

[***]

Advances in Fuzzy Logic, Neural Networks and Genetic Algorithms, Ed. Takeshi Fusuhashi, IEEE/Nagoya-University World Wisepersons Workshop Nagoya, Japan, August 1994, Selected Papers

[GrSe95]

Griffioen, J., Seales, B., Yavatkar, R., Content-based Multimedia Data Management and Efficient Remote Access, <http://este.darmstadt.gmd.de>

[MaPa94]

Ma, K., Painter, J., Hansen, C., Parallel Volume Rendering Using Binary-Swap Compositing, IEEE Computer Graphics and Applications, vol. 4, no. 4, July 1994

[ZhLo97]

Zhang, H., J., Low, C., Y., Smoliar, S., W., Wu, J., H., Video Parsing, Retrieval and Browsing: An Integrated and Content-Based Solution,

<Http://www.iss.nus.sg/RND/MS/Projects/vc/vidorigin.html>

[ZhCh98]

Zhang, A., Chang, W., Shekholeslami, G., NetView: Integrating Large-Scale Distributed Visual Databases, IEEE Multimedia July – September 1998

[WoBl96]

Wold, E., Blum, T., Keislar, D., Wheaton, J., Content-Based Classification, Search, and Retrieval of Audio, IEEE Multimedia 1996, vol.3, no. 3