

Algoritm pentru recunoasterea miscarii într-o secventa video

Georgeta DRULA
Universitatea Bucuresti

Algoritmii de recunoastere a obiectelor, formelor, vorbirii, miscarii sau gesticii constituie o categorie specifica, ce caracterizeaza domeniul multimedia. Recunoasterea este o alta forma de clasificare, care nu se mai bazeaza pe masurarea dimensiunilor si formelor obiectelor. 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.

Cuvinte cheie: *object-recognition, pattern-recognition, speech-recognition, motion-recognition, gesture-recognition, matching, corelare spatiala.*

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

Metodele de recunoastere asigura clasificarea prin comparare a obiectelor. Tehnica de egalizare sau *matching* a imaginii compara portiuni de imagini unele cu altele. Ea se realizeaza printr-un proces cunoscut ca intercorelare spatiala. Tehnica implica compararea pixel cu pixel a unei imagini mici, de referinta, ce contine un obiect de interes cu o alta imagine în analiza. O masca a imaginii de referinta apare ca o imagine mica, ce înfatiseaza obiectul ce se

doreste a se gasi în imagine. Mecanismul consta în suprapunerea mastii imagine de referinta pe o alta imagine, pentru comparare pixel cu pixel.

Un algoritm deosebit din categoria celor de recunoastere este cel ce permite extragerea caracteristicilor de miscare dintr-o secventa video, luându-se în considerare atât operatiile efectuate cu camera video (înclinare, zoom) cât si continutul video prin caracteristicile sale temporale, *algoritmul de determinare a muchiiilor în miscare (moving edges)* [Qu96]. Un astfel de algoritm este foarte important pentru luarea în considerare a continutului video prin caracteristici temporale, folosite apoi la clasificarea si regasirea cadrelor în relatie cu un anumit eveniment. Miscarea este obtinuta prin operatiile camerei de luat vederi,

considerate la separarea si la reprezentarea continutului unui plan video.

Algoritmul de recunoastere a miscarii care determina muchiile în miscare între cadrele succesive se bazeaza pe sesizarea pozitiei si a miscarii obiectelor din acestea si parcurge trei etape:

1. Procesarea imaginilor pentru a determina *locatia muchiilor în miscare*;

Modelarea datelor video se face ca un sir de imagini ce variaza în timp, exprimate prin variabila $I(x, y, t)$. Derivata partiala în raport cu timpul a acesteia estimeaza diferentele dintre doua imagini cadru succesive din secventa video:

$$D(x, y, t) = I(x, y, t + \Delta t) - I(x, y, t) .$$

Aceste imagini se convertesc în imagini formate din muchii spatio-temporale si se normalizeaza cu valori cuprinse în intervalul $[0, 1]$. Aceasta operatie permite o multiplicare (M) echilibrata de pixeli pe imagini, în felul urmator:

$$M_x(x, y, t) = (\partial I(x, y, t) / \partial x) * \partial (\partial I(x, y, t) / \partial t)$$

$$M_y(x, y, t) = (\partial I(x, y, t) / \partial y) * \partial (\partial I(x, y, t) / \partial t)$$

Imaginile succesive cu muchii în miscare notate $E_t(x, y, t)$ si $E'_t(x, y, t)$ se obtin din imaginile $M(x, y, t)$ si respectiv din $M'_t(x, y, t)$ prin aplicarea unei suprimari nemaximale. Punctele muchiilor în miscare se calculeaza dupa suprimarea nemaximala.

Algoritmul a fost implementat de pe un sir de date video *full-frame* necomprimat la 4 cadre pe secunda. Punctele muchiilor în miscare se calculeaza dupa suprimarea ne-

maximala. Aceasta modalitate are câteva avantaje în comparatie cu modalitatile tipice de trasare a muchiilor, care mai întâi extrag muchiile din fiecare imagine si apoi determina miscarea prin corelatie: el nu ia în considerare punctele de pe muchiile care nu se misca.

2. Calculul succesiunii câmpurilor ce determina *viteza punctelor* muchiilor în miscare de-a lungul mai multor cadre;

Algoritmul implica selectarea punctelor dominante ale muchiilor, distribuite în fiecare imagine-cadru, calculând un câmp de derulare initial si apoi netezind acest câmp prin aplicarea unei restrictii. Pentru a asigura împrastierea punctelor dominante se divide imaginea cu muchii în miscare $E_t(x, y, t)$ în zone de 10×10 pixeli si se selecteaza un punct dominant în fiecare zona. Dintre punctele dominante se aleg numai acelea cu mai mult de doua puncte vecine si cu panta spatio-temporala cea mai mare. Setul de punctele dominante este indicat în imaginea $E_t(x, y, t)$ ca D_t . De la acestea se obtine un set initial de asemanari vector prin calculul unei corelatii de diferente absolute, pornind de la imaginea $I(x, y, t)$ pâna la imaginea $I(x, y, t + \Delta t)$ pentru toate punctele dominante D_t . Se produce o lista de asemanari posibile $V(x_e, y_e)$ pentru fiecare punct din setul de puncte dominante, $(x_e, y_e) \in D_t$.

Aceasta lista de asemanari este ordonata dupa resturile corelatiei diferentei absolute. Luând în considerare asemanarea cea mai mare dintre punctele $v(x_e, y_e) \in V(x_e, y_e)$ se obtine un câmp initial de zgomot. Se ne-

tezeste câmpul vector prin aplicarea restricției de variația, în felul următor:

$$\min \iint_A ((dv/dS)/\|v\|)dS = \sum_{(x_e, y_e)} \sum_{(x'_e, y'_e)} \{(\|v - v'\|)/(\|v - v'\|\partial v)\}$$

dacă $\|(x_e, y_e) - (x'_e, y'_e)\| \leq N$ și minimul $\neq 0$.

unde:

$$(x'_e, y'_e) \neq (x_e, y_e);$$

$v = v(x_e, y_e) \in V(x_e, y_e)$ și $v' = v'(x'_e, y'_e) \in V'(x_e, y_e)$ sunt vectorii selectați pentru (x_e, y_e) și respectiv (x'_e, y'_e) .

Scopul final este de a găsi un optim global (x_e, y_e) pentru fiecare punct muchie ales din candidații posibili $V(x_e, y_e)$.

Optimul se găsește folosind un alt algoritm, ce minimizează variația dintre puncte prin modificarea vectorilor viteza în puncte și prin eliminarea vectorilor viteza redundanți. Acest algoritm pornește cu setul de cele mai bune perechi locale care pot estima câmpul. Procesul continuă până când se creează o diferență mai mică decât o valoare prag sau se atinge numărul maxim de iterații. Rezultatul este supus procesului de rafinare. Sunt posibile trei rafinări într-o singură iterație, și anume:

- se schimbă un vector $v(x_e, y_e)$ cu un altul $(V_x, V_y) \in V(x_e, y_e)$, care provoacă o descreștere mare a expresiei anterioare, adică se depășește o valoare prag T_s ;
- se elimină punctele (x_e, y_e) redundante din D_t . Un punct se consideră redundant dacă prin eliminarea lui ar descrește suma cu mai mult decât un prag T_d ;
- adăugarea de noi puncte (x_e, y_e) în D_t . Se adaugă acele puncte a căror valoare ar crește suma cu mai puțin decât o valoare prag T_d .

3. Gruparea vectorilor într-o succesiune coerentă ce determină *calea de mișcare (direcția)* a obiectelor din imagine.

Se consideră ca vectorii apropiați din același cadru și punctul cu aceeași direcție generală aparțin aceluiași grup. Algoritmul determină iterativ grupurile corespunzătoare din fiecare cadru, considerând pe rând fiecare vector. Dacă nu există un grup compatibil se creează unul nou pentru a putea include vectorul. Odată ce vectorii din fiecare cadru sunt grupați se aplică un criteriu de trasare a căii de mișcare a grupurilor vector prin mai multe cadre. Direcția medie a fiecărui grup prezice locația grupului în următorul cadru prin proiecție liniară.

Acest algoritm se poate aplica cu succes în sisteme multimedia de sinteză, virtuale sau virtualizate. "Tragerea" corectă în *real-time* a secvențelor video mixate cu imaginile de sinteză în cadrul studioului virtual este dependentă de pozițiile camerei de luat vederi și de informațiile de *z-mixing* și de *chromakeing* ale acestora. Măsurarea parametrilor camerei de luat vederi se realizează printr-un astfel de algoritm de estimare a mișcării ce măsoară mișcarea globală a imaginii la un grad înalt de acuratețe.

Pentru captarea realității de virtualizat, sistemul solicită folosirea mai multor camere

de luat vederi asezate în puncte esențiale ce permit captarea detaliilor. Aceasta condiție impune sistemului de sinteză folosirea unor algoritmi care să poată manevra aceleași imagine cu grade de luminanță diferite. În plus, descrierea scenelor studioului virtualizat se bazează pe structura geometrică și fotometrică a tuturor suprafețelor vizibile dintr-o locație dată de camera de luat vederi. Pe acest considerent unghiul de transcriere trebuie să includă atât orientarea, cât și poziția camerei. Distribuția unghiurilor de transcriere și densitatea lor este importantă pentru calitatea reconstrucției virtualizate a evenimentelor. Și în această aplicație este necesar un algoritm ce determină mișcarea ca cel expus. O cerință impusă sistemului de sinteză multimedia și respectiv modelului său este aceea de menținere a sincronizării imaginilor scenei luate la același moment de către diferitele camere de luat vederi. Sincronizarea imaginilor este crucială pentru a virtualiza corect evenimentele dependente de timp, folosind algoritmi potriviți.

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

[BlBu96]

Blonde, L., Buck, M., Galli, R., A Virtual Studio for Live Broadcasting: The Mona Lisa Project, IEEE Multimedia summer 1996, vol. 3, no. 2

[GiAr98]

Gibbs, S., Arapis, C., Breiteneder, C., Lalioti, V., Virtual Studios: An Overview, IEEE Multimedia january - march 1998, vol. 5, no. 1

[KaRa97]

Kanade, T., Rander, P., Narayanan, P., J., Virtualized Reality: Constructing Virtual Worlds from Real Scenes, IEEE Multimedia january - march 1997, vol. 4, no. 1

[MoTa97]

Moezzi, S., Tai, L., Gerard, P., Virtual View Generation for 3D Digital Video, IEEE Multimedia january - march 1997, vol. 4, no. 1

[Qu96]

Quek, F., H., K., Unencumbered Gestural Interaction, IEEE Multimedia 1996, vol. 4, no. 3

[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](http://www.iss.nus.sg/RND/MS/Projects/vc/vidorigin.html)