ACADEMIA ROMÂNĂ

INSTITUTUL DE MATEMATICĂ SIMION STOILOW

Estimarea Robustă a Câmpurilor de Corespondență Vizuală și a Structurii Scenelor Dinamice

Autor,
Andrei ZANFIR

Conducător,
C.S. I Dr. Cristian SMINCHIȘESCU

REZUMAT TEZĂ DE DOCTORAT

București
2018

Capitolul 1

Introducere

Stabilirea de corespondențe între structuri similare (e.g. puncte) în imagini diferite reprezintă o problemă fundamentală de vedere artificială. Ea expune necesitatea pentru un raționament în imagine non-localizat, balansând geometria 3d care stă la bază scenei cu modelarea tuturor posibilităților de deformare la care multe obiecte – incluzând aici structuri deformabile și articulate precum oameni și animale – sunt supuse. Aplicațiile sunt diverse, pentru că toate tehnologiile actuale emergente se bazează pe variante de asociere de la o imagine la alta, precum fluxul optic, fluxul scenei, potrivirea rară, potrivirea semantică, etc. Mașinile autonome au nevoie să înțeleagă mișcarea vizuală, dronele trebuie să urmărească obiecte iar dispozitivele de realitate virtuală au nevoie să construiască modele geometrice 3d ale mediului înconjurător. Astfel, ele necesită soluții la întrebări fundamentale, dar complicate de tipul: Când sunt două primitive vizuale instanțele unei aceleiași structuri 3d în mișcare? Cum pot fi potrivite două structuri din aceeași categorie semantică, dar cu aparență diferită?

În această teză, nu numai că propunem atât modele cât și soluții pentru stabilirea de corespondențe în imagini, dar, totodată, oferim intuiții și direcții de cercetare de urmărit pe termen lung. Modelele noastre sunt construite să adreseze toate provocările actuale din potrivirea de imagini, incluzând aici deplasări mari și mișcări rapide de structuri, structuri repetitive, condiții complexe de iluminare, fundaluri încărcate, variabilitatea din interiorul aceleiași clase semantice, sau zone lărgi de ocluzie.

Pentru calculul fluxului optic, propunem o metodă care merge de la rar-la-dens, ce combină potrivirea rară cu un model afin nou și cu o metodă de interpolare geometrică, obținând rezultate competitive. Arătăm că urmărind aceste modele, alternative la abordarea clasică, variatională, reușim să aducem beneficii in materie de performanță.

Adresăm și problema fluxului scenei, corespondentul 3d al fluxului optic, prin exploatarea adițională a informației de adâncime, colectată cel mai adesea de la un senzor de adâncime. Prin incorporarea unui model de energie complex, ce balansează potrivirea rară, o ipoteză de model rigid 3d, termeni de constanță fotometrică și de adâncime, obținem rezultate *state-of-the-art*, ce pot surprinde mișcări rapide și cu granițe clare.

Totodată, noi propunem în continuare o formulare de învățare în profunzime a potrivirii, unde construim și antrenăm un model pentru a minimiza un obiectiv de potrivire de grafuri, ce combină relații unare și între perechi, din vecinătăți de noduri. Această abordare nouă implică o multitudine de provocări matematice, pentru că derivatele trebuie să fie calculate și transmise exact, prin straturi structurale complete (precum module de optimizare ce au soluții pe baza unor e.g. relaxări), urmărind regulile de propagare înapoi pentru matrice (*matrix backpropagation*). Construim un cadru complex și complet antrenabil, și arătăm cum obținem rezultate îmbunătățite peste seturi de date de testare atât pentru potrivirea geometrică, cât și pentru potrivirea semantică. Această ultimă muncă a obținut o mențiune la IEEE CVPR 2018, pentru titlul de cea mai bună lucrare.

Capitolul 2

Potrivire Local Afină de la Rar-la-Dens pentru Estimarea de Mișcare și de Ocluzie

Acest capitol este bazat pe "Locally Affine Sparse-to-Dense Matching for Motion and Occlusion Estimation" Marius Leordeanu, Andrei Zanfir and Cristian Sminchișescu, prezentat la IEEE International Conference on Computer Vision, Sydney, Australia, Decembrie 2013.

2.1 Introducere

Estimarea unui câmp dens de corespondențe între cadre succesive ale unui video este o problemă importantă pentru multe aplicații de recunoaștere sau de învățare vizuală. Aici, noi propunem o metodă nouă de potrivire de la rar la dens, ce este menită să estimeze un câmp de mișcare cu detecție de zone de ocluzie. Ca o alternativă la abordările actuale de rar-la-dens din literatura de flux optic, noi pornim de la un nivel mai înalt de potrivire rară, cu aparență bogată și constrângeri geometrice, folosind un model nou, local afin și sensibil la ocluzie. Apoi, ne mutăm spre modelul mai simplu, dar mai dens, de model de flux optic, cu o procedură de interpolare nouă, ce oferă o tranziție naturală de la câmpul rar de corespondențe la cel dens. Demonstrăm experimental că trăsăturile de aparență și constrângerile geometrice mai complexe permit estimarea corectă a mișcării, chiar și în

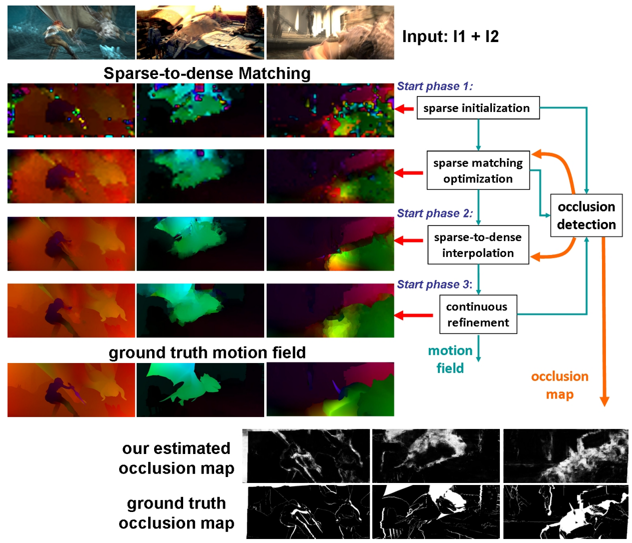

Figura 2-1: Stânga: câmpuri de mișcare estimate la diferite stagii ale metodei noastre. Dreapta: organigrama abordării noastre.

cazurile dificile de mișcări rapide și schimbări dramatice de aparență. Totodată, propunem o metodă de clasificare pentru detecția ocluziei, ce funcționează în conjuncție cu potrivirea rară-la-densă. Ne validăm abordarea pe setul de date Sintel, pe care obținem rezultate *state-of-the-art*.

Prezentare generală a abordării: Noi propunem o potrivire de imagini rară-la-dens ierarhică care integrează și generalizează idei din atât abordarea variațională tradițională de rar-la-dens când și din metodele mai recente ce folosesc potriviri rare de trăsături [1, 6, 4, 3, 13, 7]. Metoda noastră constă din trei stagii principale (Figura 2-1):

1. **Potrivire rară:** Inițializează un set discret de corespondențe candidat pentru fiecare punct de pe o grilă regulată, rară(în prima imagine), folosind o potrivire densă bazată pe kNN (în a doua imagine) cu descriptori de trăsături locale. Folosește rezultatul unui detector de muchii în imagine și de indicii din potrivirea rară inițială pentru a infera o prima hartă de probabilitate de ocluzie. Apoi, optimizează discret o funcție de cost pentru potrivire rară folosind atât termeni de date unari (de la trăsăturile locale) cât și relații geometrice bazate pe un model local afin cu constrângeri de ocluzie. Folosește aceste corespondențe îmbunătățite pentru a rafina harta de ocluzie.

2. **Interpolare de la rar-la-dens:** Fixează corespondențele rare de la pasul anterior, ca apoi să aplice același model geometric sensibil la ocluzie cu scopul de a obține o interpolare mai precisă de la rar-la-dens.

3. **Rafinare densă a corespondențelor:** Folosește un model de Variație Totală cu o optimizare continuă de flux pentru a obține câmpul final de corespondențe. Folosește indicii de potrivire, calculate de la toate stagiile, pentru a obține harta finală de ocluzie.

2.1.1 Model Spațial Local Afin

Contribuția cheie a lucrării prezentate în acest capitol este premiza spațială local afină pe care o propunem. Intuiția este că punctele care au o probabilitate mare să se regăsească pe o aceeași suprafață și care sunt apropiate între ele, au totodată o probabilitate mare să aibă o mișcare similară. Mișcările unor astfel de puncte, dintr-o vecinitate anume, sunt așteptate să urmeze îndeaproape o transformare afină. Pentru a modela această idee, trebuie să definim în primul rând un sistem geometric de vecinitați peste locațiile grilei (vezi Figura 2-2), și să conectăm puncte învecinate (i, j) folosind o funcție care măsoară puterea legăturii e_{ij} – i.e. probabilitatea ca două puncte să se afle pe o aceeași suprafață de obiect și să urmeze o transformare afină similară, de la I_1 la I_2.

Ne folosim de pozițiile vecinilor lui i, $\mathbf{p}_{N_i}^{(1)}$ în I_1, destinațiilor lor actuale în I_2, $\mathbf{p}_{N_i}^{(2)}$, și de ponderile lor de apartenență, pentru a estima un model de mișcare care *prezice* destinația $\mathbf{p}_i^{(2)}$ a punctului actual i, în I_2: $\tilde{\mathbf{p}}_i^{(2)} = T_{\mathcal{N}_i}(\mathbf{p}_i^{(1)})$. Atunci, eroarea de predicție dintre $\mathbf{p}_i^{(2)}(\mathbf{w}_i) = \mathbf{p}_i^{(1)} + \mathbf{w}_i$ și $\tilde{\mathbf{p}}_i^{(2)}$ va forma baza termenului nostru spațial:

$$E_S(\mathbf{w}) = \sum_i \|\mathbf{p}_i^{(2)}(\mathbf{w}) - T_{\mathcal{N}_i}(\mathbf{p}_i^{(1)})\|^2. \tag{2.1}$$

În cazul nostru particular, când $T_{\mathcal{N}_i}$ este o transfomare local afină, Eq. 2.1 se reduce la o energie de ordinul doi ce poate fi optimizată în mod eficient.

Pentru fiecare punct i îi estimăm transformarea afină $T_{\mathcal{N}_i} = (\mathbf{A_i}, \mathbf{t_i})$ din mișcarea vecinilor săi prin metoda ponderată a celor mai mici pătrate, cu ponderile $e_{ij},\ \forall j \in \mathcal{N}_i$. Fie

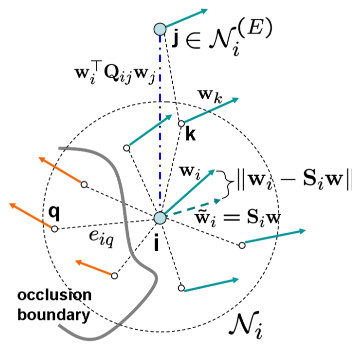

Figura 2-2: O trăsătură i este conectată la vecinul său q în $\mathcal{N}_i$ cu intensitatea e_{iq} – o funcție de distanță, și de conturul ce se interpune plus informația de ocluzie. Mișcările actuale ale vecinilor în $\mathcal{N}_i$ sunt folosite pentru a prezice mișcarea $\tilde{\mathbf{w}}_j$ într-un punct i, printr-o asociere afină $\mathbf{S}_i$. Termenul nostru cuadratic afin nou produce un sistem de vecinitate extinsă $\mathcal{N}_i^{(E)}$ în care perechile de puncte *conectate* (i, j) contribuie la eroarea totală cu cantitatea $\mathbf{w}_i^\top \mathbf{Q}_{ij} \mathbf{w}_j$.

$2N_i \times 1$ $\mathbf{p}_{N_i} = \mathbf{p}_{N_i}^{(2)}$ pozițiile estimate ale vecinilor săi in I_2 iar $\mathbf{M}$ pseudo-inversa Moore-Penrose a problemei celor mai mici pătrate, ce depinde doar de pozițiile $\mathbf{p}_{N_i}^{(1)}$ ale vecinilor și de ponderile acestora.

Soluția celor mai mici pătrate $\mathbf{M}\mathbf{p}_{N_i}$ ne oferă perechea estimată $(\mathbf{A_i}, \mathbf{t_i})$:

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{m}_1^\top \mathbf{p}_{N_i} & \mathbf{m}_2^\top \mathbf{p}_{N_i} \\ \mathbf{m}_3^\top \mathbf{p}_{N_i} & \mathbf{m}_4^\top \mathbf{p}_{N_i} \end{bmatrix} \qquad \mathbf{t}_i = \begin{bmatrix} \mathbf{m}_5^\top \mathbf{p}_{N_i} \\ \mathbf{m}_6^\top \mathbf{p}_{N_i} \end{bmatrix}. \tag{2.2}$$

Aici $\mathbf{m}_k^\top$ reprezintă al k-ulea rând al lui $\mathbf{M}$. Folosind $\mathbf{p}_i^{(1)} = [x_i, y_i]$, poziția de final prezisă $\tilde{\mathbf{p}}_i^{(2)}$ a lui i în imaginea I_2 este:

$$\begin{aligned} \tilde{\mathbf{p}}_i^{(2)} &= T_{\mathcal{N}_i}(\mathbf{p}_i^{(1)}) = \mathbf{A}_i \mathbf{p}_i^{(1)} + \mathbf{t}_i \qquad (2.3) \\ &= \overbrace{\left[\begin{array}{c|c|c} 0 & \begin{array}{c} x_i \mathbf{m}_1^\top + y_i \mathbf{m}_2^\top + \mathbf{m}_5^\top \\ x_i \mathbf{m}_3^\top + y_i \mathbf{m}_4^\top + \mathbf{m}_6^\top \end{array} & 0 \end{array} \right]}^{\mathbf{S}_i} \mathbf{p}^{(2)}. \end{aligned}$$

Observați că $\mathbf{S}_i$ nu depinde de dislocările necunoscute (sau de pozițiile finale $\mathbf{p}^{(2)}$) pe care încercăm să le rezolvăm. Ecuația 2.1 poate fi acum scrisă ca:

$$
\begin{aligned}
E_S(\mathbf{w}) &= \sum_i \|\mathbf{p}_i^{(2)} - \mathbf{S}_i \mathbf{p}^{(2)}\|^2 = \sum_i \|\mathbf{w}_i - \mathbf{S}_i \mathbf{w}\|^2 \\
&= \mathbf{w}^\top (\mathbf{I} - 2\mathbf{S}_{1\ldots n} + \sum_i \mathbf{S}_i^\top \mathbf{S}_i)\mathbf{w} \\
&= \mathbf{w}^\top \mathbf{S} \mathbf{w}, \qquad (2.4)
\end{aligned}
$$

Acest termen de energie $E_S(\mathbf{w})$ va fi folosit de către stagiile de potrivire rară și de interpolare de la rar-la-dens pentru a calcula sau pentru a actualiza soluția $\mathbf{w}^*$.

2.2 Rezultate experimentale

Arătăm rezultate calitative și cantitative în fig. 2-3 și în tabelul 2.1.

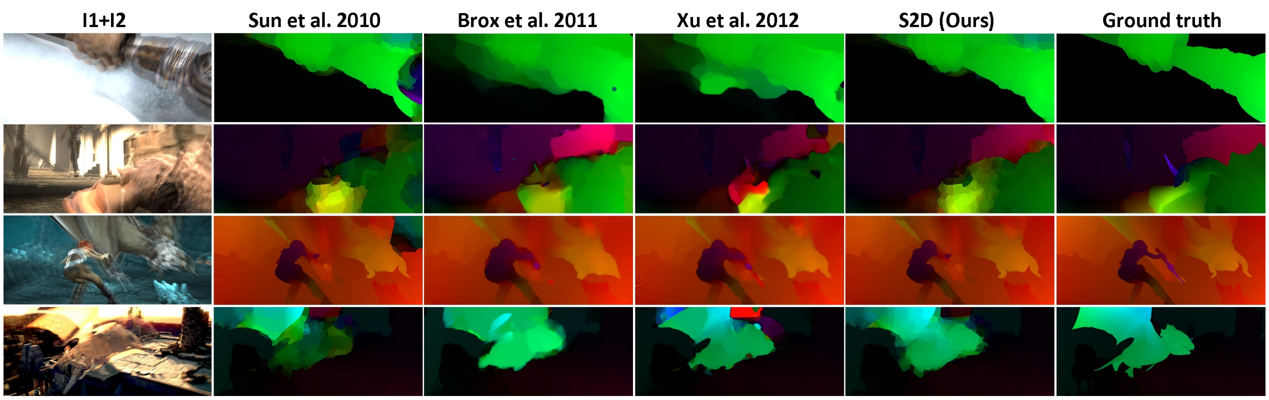

Figura 2-3: Rezultate de estimare a câmpului de mișcare. Algoritmul nostru poate să se descurce și cu deplasări mari de structură (în faza de potrivire rară), păstrând în același timp detaliile de mișcare prin rafinarea continuă de la ultimul stagiu.

Tabela 2.1: Rezultate finale pe setul de date pentru flux optic Sintel; epe reprezintă erorile în pixeli, epe_m sunt errorile pentru punctele vizibile, iar epe_o sunt erorile pentru punctele ocludate. Ultimele trei coloane arată erorile medii pentru diferite magnitudini ale mișcării (în pixeli).

Alg.	epe	epe_m	epe_o	0-10	10-40	40+
S2D	**7.88**	**3.91**	**40.16**	1.17	4.66	**48.90**
[13]	8.45	4.15	43.43	1.42	5.45	50.51
[3]	9.12	5.04	42.34	1.49	4.84	57.30
[11]-full	9.16	4.81	44.51	1.11	**4.50**	60.29
[5]	9.61	5.42	43.73	1.88	5.34	58.27
[11]++	9.96	5.41	47.00	1.40	5.10	64.14
[11]-fast	10.09	5.66	46.15	**1.09**	4.67	67.80
[12]	11.93	7.32	49.37	1.16	7.97	74.80

Capitolul 3

Flux de Scenă 3D pentru Deplasări Mari de Structură cu Raționament de Ocluzie

Acest capitol este bazat pe lucrarea "Large Displacement 3D Scene Flow with Occlusion Reasoning" Andrei Zanfir and Cristian Sminchișescu, publicată la IEEE International Conference on Computer Vision, Santiago de Chile, Chile, Decembrie 2015

3.1 Introducere

Apariția unor senzori RGB-D moderni, accesibili și preciși au intensificat nevoia pentru abordări de estimare a mișcării tridimensionale, dintr-un singur unghi de vedere, mișcare cunoscută și drept *flux de scenă*. În această lucrare noi propunem o formulare de *flux de scenă* de la rar-la-dens, bazată pe corespondențe care se bazează pe un raționament geometric explicit pentru a ține de cont de efectele dislocărilor mari de structură și pentru a modela ocluzia. Prin integrarea tuturor componentelor geometrice și fotometrice într-un singur model de energie, sensibil la ocluzie, definit peste vecinităti suprapuse și adaptabile la imagine, metoda noastră poate procesa mișcările rapide și zonele mari de ocluzie, așa cum sunt ele prezente în seturi de date complexe precum MPI Sintel Flow, ce a fost recentat augmentat cu informație de adâncime. Modelând explicit deplasările mari și ocluzia, suntem capabili să tratăm secvențe dificile ce nu pot fi actual procesate de către metodele existente *state-of-the-art*. Totodată, arătăm că, prin integrarea informației de adâncime în model, putem obține câmpuri de corespondențe cu suport spațial îmbunătățit și cu granițe mai ascuțite, comparativ cu *state-of-the-art*-ul.

3.2 Formularea Modelului

Noi proiectăm un model dens de energie care poate trata deplasările mari de structură și ocluzia. Modelul este exprimat în termeni de câteva sub-componente de energie. El se bazează pe ancore rare de corespondențe, pe asumpții de rigiditate locală definite peste vecinitati extinse construite folosind norul de puncte RGB-D, și pe termeni de constanță a aparenței și a adâncimii. Estimatele geometrice de ocluzie sunt folosite pentru a controla intensitatea conexiunilor din cadrul unei vecinităti și pentru a masca în mod automat regiuni pentru care nu pot fi stabilite corespondențe.

Câmpuri de parametri rigizi. Modelul nostru geometric leagă corespondențele dense $\mathbf{Y}$ de punctele inițiale $\mathbf{X}$ presupunând existența unui câmp $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, ..., \boldsymbol{\theta}_N^\top]^\top$, unde $\boldsymbol{\theta}_i = [\mathbf{v}_i^\top, \mathbf{t}_i^\top]^\top$ sunt parametrii rigizi ce constrâng fiecare vecinitate $\mathcal{N}(i)$ a unui punct $\mathbf{x}_i \in \mathbf{X}$. O vecinătate este reprezentată de cele mai apropiate *K* puncte de $\mathbf{x}_i$ în 3d, unde *K* este o valoare aleasă în mod corespunzător, în funcție de nivelul din piramidă.

Modelul complet de energie ce trebuie optimizat poate fi scris ca:

$$E(\mathbf{Y}, \boldsymbol{\Theta}) = E_A + \alpha E_Z + \beta E_M + \lambda E_G \tag{3.1}$$

unde α, β și λ sunt ponderi estimate prin validare.

Minimizare. Rezolvăm (3.1) peste câmpurile $\mathbf{Y}, \boldsymbol{\Theta}$ printr-o optimizare cu alternare de variabilă:

$$\boldsymbol{\Theta}^{k+1} \leftarrow \operatorname*{argmin}_{\boldsymbol{\Theta}} E(\mathbf{Y}^k, \boldsymbol{\Theta}^k) \tag{3.2}$$

$$\mathbf{Y}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{Y}} E(\mathbf{Y}^k, \boldsymbol{\Theta}^{k+1}) \tag{3.3}$$

unde k este indexul iterației. Rulăm optimizarea până la convergență.

3.3 Rezultate experimentale

Arătăm rezultate cantitative și calitative pe seturi de date de testare *state-of-the-art* în tabelul 3.1, în fig. 3-1 și în fig. 3-2.

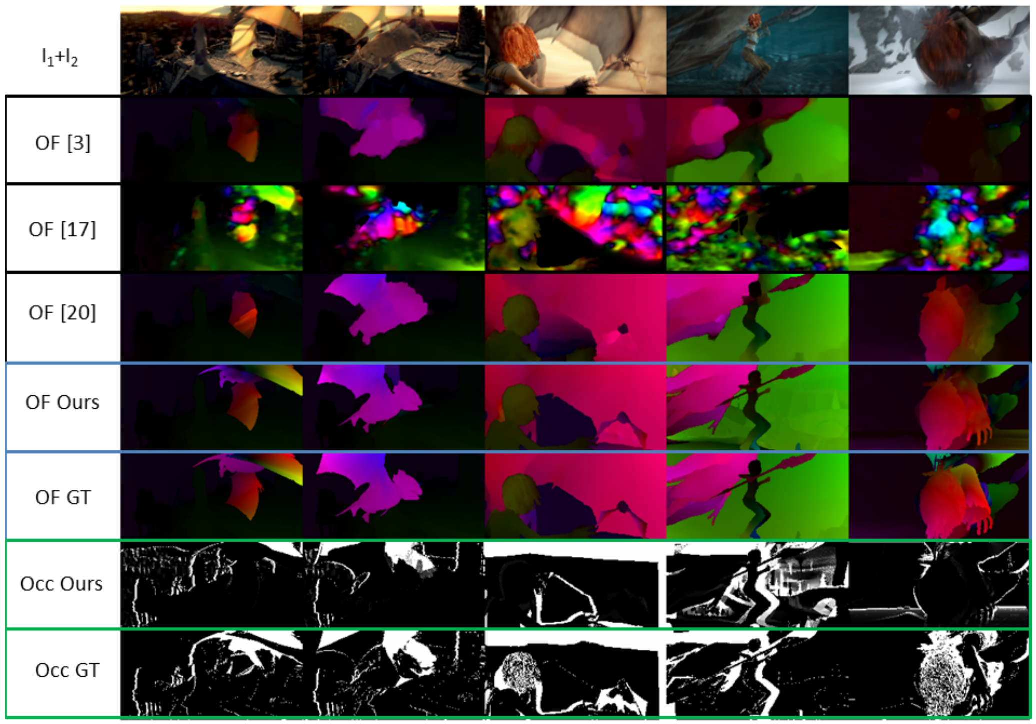

Figura 3-1: Cinci perechi de imagini și patru metode ilustrate pe setul de date Sintel: **primul rând:** imaginile de intrare suprapuse **al doilea rând:** [2], **al treilea rând:** flux de scena [8], **al patrulea rând** EpicFlow[9], **al cincilea rând:** metoda noastră propusă demonstrează granițe mai ascutie și estimate de suport spațial îmbunătățit, **al șaselea rând:** flux optic de referință. Pe ultimele două rânduri arătăm stările de ocluzie de referință și estimatele noastre continue.

Algoritmi	LDOF[2]	EpicFlow[9]	Noi
Eroare(în pixeli)	7.44	4.82	4.6

Tabela 3.1: Evaluare cantitativă a metodelor *state-of-the-art* 2d și 3d peste setul de date foarte competitiv (*Sintel-Test92*). Metoda noastră folosește informație 3d și oferă precizie îmbunătățită precum și estimate bune de ocluzie.

Figura 3-2: Mostre de estimare de flux al scenei cu rezultate în planul x-y pentru perechi de imagini din Sintel (primele trei coloane) și CAD120 (ultimele trei coloane). Deplasările de structură sunt *moderate*. Prezentăm rezultate pentru 3 metode diferite: **primul rând:** flux de scenă 3d [8]; **al doilea rând:** flux optic 2d [2]; **al treilea rând:** metoda noastră propusă pentru flux de scenă fără a se baza pe ancore de corespondență la distanță. A se observa nivelul de detaliu mai înalt capturat de către modelul nostru.

Capitolul 4

Învățarea Profundă a Potrivirii de Grafuri

Acest capitol este bazat pe lucrarea "Deep Learning of Graph Matching" Andrei Zanfir and Cristian Sminchișescu, publicată la IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake city, U.S.A, Iunie 2018

4.1 Introducere

Problema de potrivire a grafurilor sub constrângeri ale nodurilor și muchiilor este o problemă fundamentală în domenii diverse precum optimizarea combinatorială, învățarea automată sau vederea artificială, unde a reprezenta atât relațiile dintre noduri cât și structura lor din cadrul unei vecinătăți este esențială. Noi prezentăm un model complet, ce face posibil să învățăm toți parametrii procesului de potrivire de grafuri, incluzând vecinătile de noduri și de perechi de noduri, reprezentate ca ierarhii de extracție de trăsături profunde. Provocarea constă în formularea diferitelor straturi computationale ale modelului într-un mod ce permite propagarea eficientă și consistentă a gradienților, de la funcția de pierdere, prin stratul de optimizare combinatorială ce rezolvă problema de potrivire, până la ierarhie de extragere de trăsături. Experimentele noastre de vedere artificială și studiile de validare de componente realizate pe seturi de date competitive precum PASCAL VOC, Sintel și CUB demonstrează că modelele de potrivire rafinate cap-coadă sunt superioare variantelor bazate pe ierarhii de trăsături antrenate pentru alte probleme. Metodologic, contribuțiile noastre sunt asociate cu construcția diferitelor straturi matriceale ale grafului computațional, obținerea de derivate analitice de la funcția de pierdere până la straturile de trăsături în cadrul de *backpropagation* pentru matrice, accentul pus pe eficientă computațională pentru etapele inverse, precum și o funcție de pierdere cu votare. Modelul propus se aplică în mod general, nu doar pentru asocierea diferitelor imagini ale unei aceleași categorii (principala funcționalitate), dar și pentru imagini ale unei aceeași scene, sau luate dintr-un video.

4.2 Formularea problemei

Intrare. Ni se dau două grafuri de intrare $\mathcal{G}_1 = (V_1, E_1)$ și $\mathcal{G}_2 = (V_2, E_2)$, cu $|V_1| = n$ și $|V_2| = m$. Scopul nostru este de a stabili corespondențe între nodurile celor două grafuri, astfel încât să optimizăm un criteriu peste nodurile și muchiile puse astfel în corespondență (vezi mai jos).

Potrivirea Grafurilor. Fie $\mathbf{v} \in \{0,1\}^{nm\times 1}$ un vector indicator astfel încât $\mathbf{v}_{ia} = 1$ dacă $i \in V_1$ se potrivește la $a \in V_2$ și 0 altfel, respectând constrângerile de asociere unu-la-unu. Construim o matrice pătratică pozitivă $\mathbf{M} \in \mathbb{R}^{nm\times nm}$ astfel încât $\mathbf{M}_{ia;jb}$ măsoară cât de bine orice pereche $(i,j) \in E_1$ se potrivește cu $(a,b) \in E_2$. Pentru perechi ce nu formează muchii, întrările lor corespunzătoare din matrice se setează la 0. Intrările de pe diagonală conțin scoruri nod-la-nod, pe când întrările din afara ei conțin scoruri de muchie-la-muchie. Indicatorul optim $\mathbf{v}^*$ se poate formula ca

$$\mathbf{v}^* = \underset{\mathbf{v}}{\operatorname{argmax}}\, \mathbf{v}^\top \mathbf{M} \mathbf{v}, \text{ s.t. } \mathbf{C}\mathbf{v} = \mathbf{1}, \mathbf{v} \in \{0,1\}^{nm\times 1} \tag{4.1}$$

Matricea binară $\mathbf{C} \in \mathbb{R}^{nm\times nm}$ codează constrângerile de asociere unu-la-unu: $\forall a \sum_i \mathbf{v}_{ia} = 1$ și $\forall i \sum_a \mathbf{v}_{ia} = 1$. Această formulare este NP-hard, astfel încât relaxăm problema, scăpând de ambele constrângeri, binare și de asociere, și rezolvăm

$$\mathbf{v}^* = \underset{\mathbf{v}}{\operatorname{argmax}}\, \mathbf{v}^\top \mathbf{M} \mathbf{v}, \text{ s.t. } \|\mathbf{v}\|_2 = 1 \tag{4.2}$$

$\mathbf{v}^*$ optim este dat de către cel mai mare vector propriu al matricei $\mathbf{M}$. Pentru că $\mathbf{M}$ conține doar elemente ne-negative, prin folosirea teoremei Perron-Frobenius, știm că elementele lui $\mathbf{v}^*$ se regăsesc în intervalul $[0, 1]$, și interpretăm $\mathbf{v}^*_{ia}$ ca pe încrederea că i se potrivește cu a.

Învățare. Estimăm $\mathbf{M}$ parametrizată în termeni de trăsături unare și de perechi de puncte, calculate peste imaginile de intrare și reprezentate ca ierarhii de trăsături adânci. Învățăm ierarhiile de trăsături cap-coadă într-o funcție de pierdere ce integrează totodată și stratul de potrivire. Mai exact, dându-se un set de antrenament cu corespondențe între perechi de imagini, noi adaptăm parametrii rețelei astfel încât potrivirea să minimizeze eroarea, măsurată ca suma distanțelor dintre corespondențele prezise și cele de referință.

4.3 Abordare

Lucrarea [14] a introdus o factorizare a matricei $\mathbf{M}$ ce expune in mod explicit structura de graf a setului de puncte si scorurile dintre noduri si muchii

$$\mathbf{M} = [\text{vec}(\mathbf{M}_p)] + (\mathbf{G}_2 \otimes \mathbf{G}_1)[\text{vec}(\mathbf{M}_e)](\mathbf{H}_2 \otimes \mathbf{H}_1)^\top \tag{4.3}$$

Calcularea celui mai mare vector propriu $\mathbf{v}^*$ a matricei de afinitate $\mathbf{M}$ poate fi realizată prin metoda puterii directe

$$\mathbf{v}_{k+1} = \frac{\mathbf{M}\mathbf{v}_k}{\|\mathbf{M}\mathbf{v}_k\|} \tag{4.4}$$

Faza de gradienți. Pentru a calcula gradientii, exprimăm variația pierderii și identificăm derivatele parțiale necesare. Prin utilizarea tehnicilor de *backpropagation* matriceal, putem exploata factorizarea specială (4.3) a matricei $\mathbf{M}$, pentru a face operațiile eficiente atât în termeni de memorie cât și de timp.

Extractor de Trăsături Profunde

In: $\mathbf{I}_1, \mathbf{I}_2$
Out: matrice de trăsături $\mathbf{F}^1, \mathbf{F}^2$ și $\mathbf{U}^1, \mathbf{U}^2$ asa cum ar fi calculate de orice CNN e.g. *VGG-16* [10]

$\rightarrow$

Matricea de Afinitate

In: $\mathbf{F}^1, \mathbf{F}^2, \mathbf{U}^1, \mathbf{U}^2$
Structura de graf: $\mathbf{G}_1, \mathbf{G}_2, \mathbf{H}_1, \mathbf{H}_2$
Calcule: construiește $\mathbf{M}_e$ și $\mathbf{M}_p$
Out: $\mathbf{M}$ ca in eq. (4.3)

$\rightarrow$

Metoda Puterii Directe

In: $\mathbf{M}$
Calcule: $\mathbf{v}_0 \leftarrow \mathbf{1}$, $\mathbf{v}_{k+1} = \mathbf{M}\mathbf{v}_k / \, |\mathbf{M}\mathbf{v}_k|$
Out: $\mathbf{v}^*$

$\rightarrow$

$\rightarrow$

Bi-Stocastizare

In: $\mathbf{v}^*$
Calcule: transformă $\mathbf{v}^*$ într-o matrice și o bi-stocastizează
Out: matricea dublu-stocastică de încredere $\mathbf{S}$

$\rightarrow$

Votare

In: $\mathbf{S} \in \mathbb{R}^{n \times m}$
Calcule: softmax$(\alpha \mathbf{S})$
Parametri: scala α
Out: vectorul de deplasare $\mathbf{d}$

$\rightarrow$

Pierdere

In: $\mathbf{d}, \mathbf{d}^{gt}$
Out: $L(\mathbf{d}) = \sum_i \phi(\mathbf{d}_i - \mathbf{d}_i^{gt})$

Figura 4-1: Modelul nostru complet antrenabil de potrivire de grafuri. La antrenare, gradientii de la funcția de pierdere sunt transmiși printr-o ierarhie de extragere de trăsături profunde, prin factorizarea matricei de afinitate rezultate, prin soluția problemei de potrivire și prin stratul de asociere bazat pe vot.

$$\frac{\partial L}{\partial \mathbf{M}_e} = \sum_k \mathbf{G}_1^\top \left(\frac{(\mathbf{I} - \mathbf{v}_{k+1}\mathbf{v}_{k+1}^\top)}{\|\mathbf{M}\mathbf{v}_k\|} \frac{\partial L}{\partial \mathbf{v}_{k+1}} \right)_{n \times m} \mathbf{G}_2 \odot \odot \, \mathbf{H}_1^\top (\mathbf{v}_k)_{n \times m} \mathbf{H}_2 \tag{4.5}$$

$$\frac{\partial L}{\partial \mathbf{M}_p} = \sum_k \frac{(\mathbf{I} - \mathbf{v}_{k+1}\mathbf{v}_{k+1}^\top)}{\|\mathbf{M}\mathbf{v}_k\|} \frac{\partial L}{\partial \mathbf{v}_{k+1}} \odot \mathbf{v}_k \tag{4.6}$$

Cu o aranjare corectă a operațiilor, complexitățile devin $O(\max(m^2 q, n^2 p))$ și $\Theta(pq)$.

4.4 Rezultate experimentale

Prezentăm câteva exemple calitative de stabilire de corespondențe pentru două seturi de date *state-of-the-art*, în fig. 4-2 și în fig. 4-3.

Figura 4-2: Patru exemple calitative ale metodei noastre, peste setul de test CUB-200-2011. Imaginile cu un contur negru reprezintă sursa, pe când cele cu un contur roșu reprezintă țintele. Imaginile cu contur verde prezintă corespondențele de referință. Culorile cercurilor identifică în mod unic 15 locații semantice.

Figura 4-3: Douăsprezece exemple calitative ale metodei noastre peste setul de test PASCAL VOC. Pentru fiecare pereche de exemple, în stânga se află imaginea sursă iar în dreapta cea țintă. Culorile identifică corespondențele calculate dintre puncte.

Bibliografie

[1] A. Berg, T. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *Proceedings of Computer Vision and Pattern Recognition*, 2005.

[2] T. Brox, C. Bregler, and J. Malik. Large displacement optical flow. In *Proceedings of Computer Vision and Pattern Recognition*, 2009.

[3] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):500–513, 2011.

[4] O. Duchenne, F. Bach, I. Kweon, and J. Ponce. A tensor-based algorithm for high-order graph matching. In *Proceedings of Computer Vision and Pattern Recognition*, 2009.

[5] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3), 1981.

[6] M. Leordeanu, A. Zanfir, and C. Sminchisescu. Semi-supervised learning and optimization for hypergraph matching. In *ICCV*, 2011.

[7] C. Liu, J. Yuen, and A. Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):978–994, 2011.

[8] Julian Quiroga, Thomas Brox, Frédéric Devernay, and James Crowley. Dense semi-rigid scene flow estimation from rgbd images. In *Computer Vision–ECCV 2014*, pages 567–582. Springer, 2014.

[9] Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, and Cordelia Schmid. Epicflow: Edge-preserving interpolation of correspondences for optical flow. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[10] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[11] D. Sun, S. Roth, and M.J. Black. Secrets of optical flow estimation and their principles. In *Proceedings of Computer Vision and Pattern Recognition*, 2010.

[12] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic huber-l1 optical flow. In *British Machine Vision Conference*, 2009.

[13] L. Xu, Jiaya Jia, and Yasuyuki Matsushita. Motion detail preserving optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 34(9), 2012.

[14] Feng Zhou and Fernando De la Torre. Factorized graph matching. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 127–134. IEEE, 2012.