

Szegedi Tudományegyetem
Informatika Doktori Iskola

Bináris képek geometriai és vizuális helyreállítása

PhD értekezés tézisei

Németh József

Témavezetők:

Dr. Csirik János

Dr. Kató Zoltán

Szeged

2015

Bevezetés

A bináris képek, melyeken csak két különböző intenzitásérték szerepel, fontos szerepet játszanak a képfeldolgozás területén, mivel

1. sok objektum képe alapján véve bináris, pl. dokumentum képek, kresz táblák, csontok vagy implantátumok a röntgen-felvételeken,
2. a legtöbb képfeldolgozási eljárás során keletkeznek bináris képek, szegmentálás útján.

Sok esetben az az információ, hogy a kép amellyel dolgozunk bináris segíthet a pontosabb eredmény elérésében, hisz leszűkíti a lehetséges megoldások halmazát. Például, az úgynevezett lineáris inverz problémák, mint amilyen a tomográfia vagy a dekonvolúció, a legtöbb esetben aluldefiniáltak, így több lehetséges megoldásuk van. Ha tudjuk, hogy a keresett megoldás bináris, leszűkíthetjük a keresési teret, ezért például tomográfia esetén a bináris képek általában kevesebb vetületből is helyreállíthatóak.

Sok esetben azonban az textúrázottság hiánya nehezíti is teheti a bináris képekkel való munkát. Az általános regisztrációs módszerek például gyakran pont-megfeleltetésekből dolgoznak, és ezeket a pontpárokat általában a pontok körül jellemző intenzitásérték mintázatok alapján nyerik ki. Bináris képek esetén ezek a módszerek általában nem tudnak megfelelő számú pont-megfeleltetést előállítani. Másrésről azonban bináris képek esetén a képek közötti intenzitásérték változás nem lép fel zavaró tényezőként. Így több módszert is kidolgoztak bináris képek regisztrálására, melyek csak az alakzatok pont-koordinátáit, és azokból számolt statisztikákat használják a képek illesztéséhez.

Ez a tézis összefoglalja a szerző bináris képek geometriai és vizuális helyreállításával kapcsolatos eredményeit. A 1. táblázat mutatja az egyes tézispontok és a publikációk kapcsolatát. A III. tézispont síkhomográfiára vonatkozó eredményei a [13] munkában is belettek mutatva.

	[9]	[10]	[12]	[11]	[4]	[8]
I.	•					
II.		•				
III.			•	•	•	
IV.						•

1. táblázat: A tézispontok és a publikációk kapcsolata.

I. Diszkrét tomográfia ismeretlen intenzitás értékekkel

Bemutatunk egy bináris tomográfiai módszert, amely magasabb fokú statisztikán alapuló diszkretizációs tagot használ. A $h \times w$ méretű digitális képet az $\mathbf{x} \in \{c_1, c_2\}^n$ oszlopvek-

torral reprezentáljuk, ahol $n = hw$ a pixelek száma, c_1 és c_2 pedig az előtérhez (objektumhoz) és a háttérhez tartozó intenzitásértékek. Tegyük fel, hogy a vetületeket k különböző szögből képezzük, és legyen l_i a mért vetületértékek száma az i . irányban. Az összes vetületi értéket a $\mathbf{p} \in \mathbb{R}^m$ oszlopvektor reprezentálja, ahol $m = \sum_{i=1}^k l_i$ vetületi értékek száma. A vetületképzési eljárás leírható a következő lineáris egyenletrendszerrel:

$$\mathbf{W}\mathbf{x} = \mathbf{p}, \quad (1)$$

ahol a $\mathbf{W} \in \mathbb{R}^{m \times n}$ mátrix írja le a vetületi geometriát. A leggyakrabban használt geometria típusok a tomográfiában és a bináris tomográfiában a párhuzamos-nyaláb, a legyező-nyaláb és a kúp-nyaláb.

A probléma folytonos változatát tekintjük. A rekonstrukciót az alábbi célfüggvény minimalizálásával végezzük:

$$E(\mathbf{x}, \alpha, \mu) = F(\mathbf{x}) + \lambda S(\mathbf{x}) + \mu D(\mathbf{x}, \alpha), \quad (2)$$

amely három elvárás fejze ki a megoldással kapcsolatban. Az első, úgynevezett *adat-megfelelési* tag azt fejze ki, hogy a megoldásnak ki kell elégítenie a vetületeket:

$$F(\mathbf{x}) = \frac{1}{2} \|\mathbf{W}\mathbf{x} - \mathbf{p}\|_2^2. \quad (3)$$

Ezt a tagot gyakran használják a diszkrét tomográfiában, hogy kifejezzék, hogy a megoldás vetületeinek nem szabad nagyon eltérniük a bemeneti $\mathbf{p}$ vetületektől. Mint általában a lineáris inverz problémák esetén, a (3) függvény minimalizálása elfajult megoldásokhoz vezetne a zajos vetületek miatt. Ezért regularizációra van szükség. Ezen célból a célfüggvény tartalmaz egy *simasági* tagot:

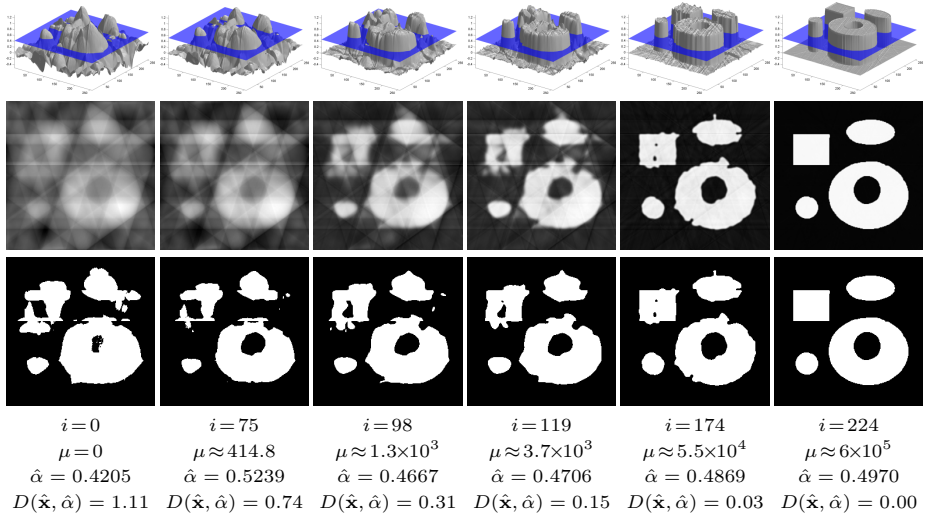
$$S(\mathbf{x}) = \frac{1}{2} \|L\mathbf{x}\|_2^2, \quad (4)$$

ahol L a Laplace regularizációs mátrix, azaz $L\mathbf{x}$ ugyanazt az eredményt adja, mint a 2-dimenziós konvolúció a Laplace-szűrővel. Ez a tag bünteti a túl nagy második deriváltakkal rendelkező megoldásokat, de megengedi az élek kialakulását. Ennek λ paraméterét nagyobbra állítva kikényszeríthetjük, hogy a megoldáson összefüggő területek alakuljanak ki, akkor is, ha a bemeneti vetületek zajosak. Csak az *adat-megfelelési* és a *simasági* tagokat használva a célfüggvény $F(\mathbf{x}) + \lambda S(\mathbf{x})$ konvex, és megoldása egy folytonos rekonstrukciót nyújt.

A bináris megoldásokat a *diszkrétizációs tag* írja elő:

$$D(\mathbf{x}, \alpha) = n \frac{\|\mathbf{x} - \alpha \mathbf{1}_n\|_4^4}{(\|\mathbf{x} - \alpha \mathbf{1}_n\|_2^2)^2} - 1, \quad (5)$$

ahol α a *diszkrétizációs paraméter*, $\mathbf{1}_n$ jelöli az n -hosszú oszlopvektort melynek minden eleme 1, és $\|\mathbf{x}\|_p = (\sum_{i=1}^n x_i^p)^{1/p}$ jelöli az általános vektor normát. Ezt a függvényt olyan bináris képek minimalizálják, melyeken a két intenzitásérték átlaga α , azaz létezik d , amelyre $|x_i - \alpha| = d$, minden $i = 1, \dots, n$ -re. Ilyen esetekben az értéke 0, míg a maximumát, melynek



1. ábra: A javasolt módszer illusztrálása. Az ábra oszlopai különböző iterációs szám utáni megoldásokat mutatnak, 5 vetületet használva. Az első sorban a megoldás 3-dimenziós ábrázolása látható, amelyen vízszintes sík jelzi a becsült $\hat{\alpha}$ középszintet. A második és harmadik sorok a megoldásokat és azok $\hat{\alpha}$ -küszöbölt változatait mutatják. A képek alatt a megfelelő μ diszkretizációs súlyok $\hat{\alpha}$ középszintek és $D(\hat{\mathbf{x}}, \hat{\alpha})$ diszkretizációs értékek találhatók.

értéke $n - 1$, akkor veszi fel, ha a képen egy kivétellel minden pixel értéke α . Így ez egy korlátos függvény, és kétszer folytonosan differenciálható, kivéve $\mathbf{x} = \alpha \mathbf{1}_n$ -ben, így standard gradiens alapú optimalizációs módszerek közvetlenül alkalmazhatók a minimalizálására. $D(\mathbf{x}, \alpha)$ lényegében egy standardizált momentum (kurtózis), azzal a különbséggel, hogy a másod-, és negyedfokú tagokat α -val normalizáljuk, az elemek átlaga helyett.

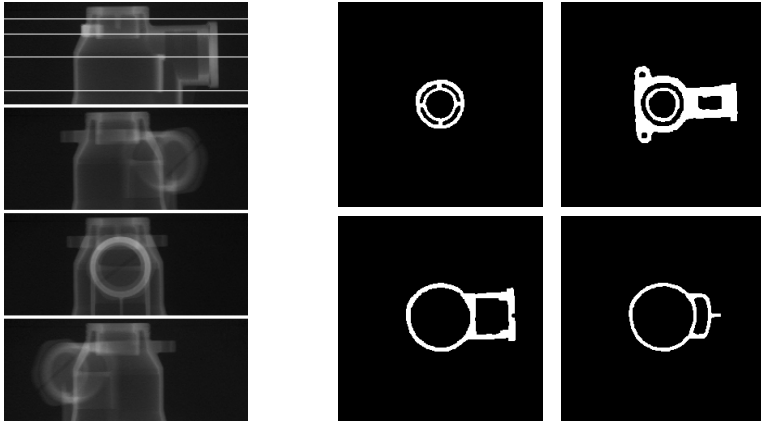
A diszkretizációs tag súlyának elég nagy értéket választva lényegében csak bináris megoldásokat engedélyezünk. A következő tétel kimondja, hogy tetszőleges $\xi > 0$ -hoz létezik egy megfelelően nagy μ úgy, hogy a diszkretizációs szint (azaz a diszkretizációs tag értéke) maximum ξ a célfüggvény minimumhelyén.

1. Tétel. *Bármely $\xi > 0$ esetén létezik $\mu(\xi) \in \mathbb{R}$ úgy, hogy ha $\mu \geq \mu(\xi)$ és $(\hat{\mathbf{x}}, \hat{\alpha}) = \arg \min_{\mathbf{x}, \alpha} E(\mathbf{x}, \alpha, \mu)$, akkor $D(\hat{\mathbf{x}}, \hat{\alpha}) \leq \xi$.*

Így a rekonstrukciós probléma adott $\xi > 0$ és $\hat{\mu} \geq \mu(\xi)$ esetén a következőképpen definiálható:

$$(\hat{\mathbf{x}}, \hat{\alpha}) = \arg \min_{\mathbf{x}, \alpha} E(\mathbf{x}, \alpha, \hat{\mu}). \quad (6)$$

Bináris dekonvolúció esetére korábban javasolták [6], hogy a középértéket és a megoldást egyszerre, ugyanabban a gradiens alapú optimalizáló eljárásban határozzák meg.



2. ábra: **Első oszlop:** A gáznyomás szabályozó homogén részének négy vetülete a 18-ból. **Többi oszlop:** A bemutatott módszer rekonstrukciós eredményei. A bemutatott eredményeken szereplő szeletek elhelyezkedését az első vetületi képen jelöltük.

Mi úgy találtuk, hogy a javasolt módszer esetén hatékonyabb, ha a középtérteket minden iterációs lépésben közvetlenül határozzuk meg az aktuális megoldáshoz, a diszkretizációs függvény minimumhelyének meghatározásával. Ezért a következő fokozatos optimalizálási módszert javasoljuk. Kiindulva egy kezdeti folytonos rekonstrukcióból, melyet a célfüggvény diszkretizációs tag nélküli minimalizálásával kapunk, minden iterációs lépésben

1. Meghatározzuk az $\hat{\alpha}$ középszintet az aktuális $\hat{x}$ megoldáshoz.
2. Növeljük a μ diszkretizációs súlyt.
3. Finomítjuk az $\hat{x}$ megoldást, azaz lokálisan minimalizáljuk a célfüggvényt az új $\hat{\alpha}$ és μ paraméterekkel.

A módszer fokozatosan kényszeríti ki a bináris megoldásokat, míg iteratíván közelíti a középszintet (lásd 1. ábra). Fontos megjegyezni, hogy a módszer nem igényli a háttér és az objektum intenzitás értékeinek becslését, sem a közbülső megoldások küszöbölését. Megmutattuk, hogy a célfüggvény Lipschitz-folytonos, ezért hatékonyan minimalizálható gradiens alapú optimalizáló módszerekkel. Bemutattuk, hogy a módszert lehet úgy implementálni, hogy a rekonstrukciós eredmény független legyen a képen lévő intenzitásértékektől.

A bemutatott módszer tesztelve lett egy szintetikus adathalmazon. A vizsgálatok megmutatták, hogy az algoritmus robusztus a vetületi zaj erősségével szemben. Összehasonlítva a PDM-DART módszerrel [15], az algoritmusunk versenyképesnek bizonyult, míg egy másik, konvex-programozási módszert egyértelműen felülmúlt.

A módszert sikeresen alkalmaztuk valós adatokon, egy gáznyomás-szabályozó homogén részének helyreállítására 18 vetületből (lásd 2. ábra).

A szerző hozzájárulása az eredményekhez

A szerző kidolgozott egy új, magasabb fokú statisztikát használó bináris tomográfiai technikát, amely azokban az esetekben is alkalmazható, amikor a képek intenzitásértékei ismeretlenek. Egy olyan célfüggvényt javasol, amelyben egy diszkretizációs tag írja elő a bináris megoldásokat. A célfüggvény minimalizálásához egy fokozatos optimalizálási stratégiát mutat be, amelyben a diszkretizációs tag súlya növekszik, így fokozatosan kényszeríti ki a bináris megoldásokat. Az intenzitásértékek középszintjének becslése a diszkretizációs tag minimalizálásával történik. A szerző megvizsgálta a módszer konvergencia-tulajdonságait és megmutatja hogy az algoritmus viselkedése független az intenzitásértékektől. Bemutatja a módszer robusztusságát a különböző mértékű vetületi zajokkal szemben. A módszert összehasonlította más korszerű módszerekkel és megmutatta, hogy az algoritmus jó alternatívát jelent. Sikeresen használta a módszert valós adatokon.

II. Bináris alakzatok dekonvolúciója diszkrét tomográfia segítségével

Célunk helyreállítani egy bináris $f \in \mathbb{R}^{u \times v}$ képet, az elmosódott és zajos $g \in \mathbb{R}^{u \times v}$ bináris képből. Az elmosódást a kép és az elmosódást leíró függvény (point spread function - PSF) konvolúciója adja, tehát:

$$g = h * f + n, \quad (7)$$

ahol $h \in \mathbb{R}^{p \times q}$ a PSF, $*$ jelöli a 2-dimenziós konvolúciót és $n \in \mathbb{R}^{u \times v}$ a zaj. Feltételezzük, hogy az elmosódást leíró h függvény ismert. Célunk az eredeti bináris kép meghatározása $b \in \{0, 1\}^{u \times v}$ alakban. Mivel azonban az intenzitásértékek nagyságrendje ismeretlen ezért bevezetünk egy ismeretlen s skálázási együtthatót a modellbe, tehát $f = sb$.

Legyen $\tilde{f}^\theta$ az f kép θ irányból képzett vetülete. Kihasználjuk, hogy ha $g = h * f$ akkor $\tilde{g} = \tilde{h} * \tilde{f}$, ahol $*$ jelöli az 1-dimenziós konvolúciót. Így az összefüggés az eredeti és a megfigyelt kép vetületei között a következőképpen fejezhető ki:

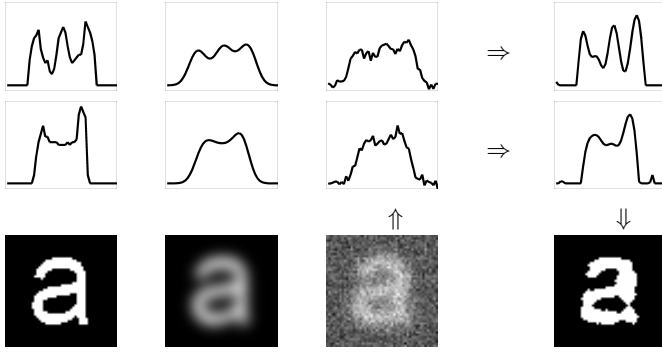
$$\tilde{g} = \tilde{h} * \tilde{f} + \tilde{n}. \quad (8)$$

Egy $\Omega = \{\theta_i | i = 1, \dots, k\}$ irányhalmaz esetén a javasolt módszer először helyreállítja a $\tilde{f}^{\theta_i}$ vetületeket a $\tilde{g}^{\theta_i}$ vetületekből, majd a második lépésben rekonstruálja az f képet a helyreállított vetületekből (lásd 3. ábra). A konvolúció egy lineáris művelet, tehát a (8) egyenlet a következőképpen írható:

$$\tilde{g} = H \tilde{f} + \tilde{n}, \quad (9)$$

ahol a H mátrix írja le a $\tilde{h}$ szűrővel végrehajtott konvolúciót. Mivel $\tilde{n}$ ismeretlen, az egyenletrendszer megoldása regularizációt igényel. A Tikhonov-regularizáció standard formája a következő:

$$\tilde{f}_\lambda = \arg \min_{\tilde{f}} \|H \tilde{f} - \tilde{g}\|_2^2 + \lambda^2 \|\tilde{f}\|_2^2, \quad (10)$$



3. ábra: A bemutatott módszer alapelemei. **Első három oszlop:** az eredeti kép, az elmosódott kép és az elmosott-zajos kép és ezek vetületei. **Negyedik oszlop:** a módszer helyreállítja a vetületeket és rekonstruálja az alakzatot.

ahol λ egy pozitív konstans a regularizációs paraméter amely a megoldás simaságát szabályozza. Adott λ esetén a közvetlen megoldást a

$$\check{f}_\lambda = (H^T H + \lambda^2 I)^{-1} H^T \check{g} \quad (11)$$

képlet adja, ahol I jelöli a megfelelő méretű egységmátrixot. A regularizációs paraméter megfelelő értékének meghatározásához az L-curve módszert használtuk, amely a maradvány-normájának és a megoldás-normájának log-log függvényét tekinti különböző regularizációs paraméterekre:

$$\mathcal{L} = \{(\log_2 \|H\check{f}_\lambda - \check{g}\|_2^2, \log_2 \|\check{f}_\lambda\|_2^2), \lambda \geq 0\}. \quad (12)$$

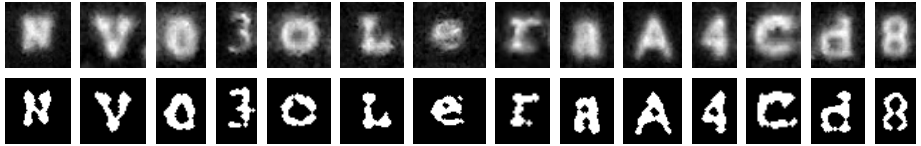
Az optimális regularizációs paramétert az a λ^* adja, melyre az $\mathcal{L}$ görbe görbülete a maximális.

A megoldásban jelentkező esetleges negatív elemek kiküszöbölésére egy további, általános Tikhonov regularizáción alapuló iteratív eljárást használtunk, mely fokozatosan eliminálja a negatív értékeket.

A diszkrét tomográfiai rekonstrukcióhoz a $\check{b}^{\theta_i} = \check{f}^{\theta_i}/s$ bináris vetületekre van szükség. Az s skálafaktort nem lehet közvetlenül kiszámítani, azonban ennek alsó és felső korlátja könnyen meghatározható, így definiálhatjuk a lehetséges értékeinek egy S halmazát. Minden $s \in S$ skálafaktorhoz vesszük a $\check{b}^{\theta_i} = [\check{f}^{\theta_i}/s]$ vektorokat, mint az ismeretlen b bináris kép vetületeinek becslését.

A vízszintes és függőleges $\check{b}^0$ és $\check{b}^{\pi/2}$ vetületekből való rekonstrukcióval foglalkoztunk. Közismert, hogy két vetület legtöbbször nem elegendő a bináris képek helyreállításához, azaz, adott vetület-párnak általában több kép is megfelelhet, míg bizonyos esetekben egy kép sem elégíti ki a vetületeket. Definíáljuk a tomográfiai-ekvivalencia halmazt:

$$U = U(\check{b}^0, \check{b}^{\pi/2}) = \{z \in \{0, 1\}^{u \times v} : z^0 = \check{b}^0, z^{\pi/2} = \check{b}^{\pi/2}\}. \quad (13)$$



4. ábra: Helyreállítási eredmények elmosódott dokumentum képekből kinyert karaktereken. Az eredeti képek az első, míg a helyreállított képek a második sorban láthatóak.

Habár általános esetben $|U| > 1$, minket az a megoldás érdekel, mely a legjobban hasonlít az eredeti bemeneti képhez, mint modellhez. Egy adott modell képnek leginkább megfelelő megoldás már egyértelműen meghatározott. Az m modell képet a bemeneti g képből hozzuk létre úgy, hogy a zaj szórásának (amely a vetületek helyreállítása után meghatározható) megfelelő mennyiségű zajt távolítunk el. Az az elvárás, hogy a z_s megoldásnak hasonlítania kell a modell képhez az jelenti, hogy azokon az (x, y) pozíciókon, ahol $z(x, y) = 1$ a modellképen magas intenzitásérték helyezkedik el. Ez a következő minimalizálási problémaként írható fel:

$$z_s = \arg \min_{z \in U} \left(- \sum_{x,y} z(x, y) m(x, y) \right). \quad (14)$$

Így a rekonstrukciós probléma visszavezethető a minimális-költségű maximális-folyam problémára [1]. Ebben a hálózatban a *ellátó* (*supply*) és az *igénylő* (*demand*) pontok a vetületek értékeit reprezentálják, az élek pedig a pixeleket. Minden egyes (x, y) pixelhez tartozó $S_x \rightarrow T_y$ él kapacitása 1, míg költsége $-m(x, y)$. A minimális-költségű maximális-folyam polinomiális időben meghatározható [14] és megadja a z_s megoldását a rekonstrukciós problémának. A $z_s(x, y)$ pixel akkor, és csakis akkor kap 1-es értéket, ha a folyam áthalad az $S_x \rightarrow T_y$ élen.

A különböző $s \in S$ skálafaktorokhoz a módszer meghatározza a különböző z_s rekonstrukciókat. Az optimális z megoldás meghatározásához összehasonlítunk minden egyes z_s megoldást a g bemeneti képpel és kiválasztjuk a legkisebb-négyzetes értelemben vett legközelebbit.

A módszer vizsgálatához egy szintetikus képi adatbázist hoztunk létre, amely 62 alfanumerikus karakterből (59×59 pixel méretben) és azok 1550 eltorzított változatából állt. Minden képet különböző erősségű Gauss-szűrővel simítottunk és különböző erősségű fehér-zajt adtunk hozzájuk. Összehasonlító teszt mutatja, hogy a módszer sok esetben jobb eredményeket ér el mint a széles körben használt Lucy-Richardson eljárás. A módszert teszteltük valós kamerával készített, életlen karakterképeken, néhány eredmény megtalálható a 4. ábrán.

A szerző hozzájárulása az eredményekhez

A szerző egy bináris tomográfia alapú bináris képhelyreállítási módszert javasol. Az elmosódott képek vetületeit Tikhonov regularizáció segítségével állítja helyre. A megfelelő regularizációs paraméter kiválasztásához az L-curve módszert javasolja. A helyreállított vetületekből egy maximum-folyam-alapú bináris tomográfiai módszerrel rekonstruálja az alakzatot. Összehasonlító teszt mutatja, hogy a módszer sok esetben jobb eredményeket ér el mint egy másik, széleskörben használt módszer. A szerző bemutat valós kamerával készített, elmosódott képeken elért eredményeket is.

III. Bináris alakzatok nemlineáris regisztrációja

Bemutatunk egy általános bináris regisztrációs módszert. Legyen $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ egy tetszőleges diffeomorfizmus, azaz egy differenciálható és invertálható transzformáció, melynek az inverze is differenciálható. Célunk meghatározni egy ilyen transzformáció paramétereit, mely fedésbe hozza a bemeneti bináris alakzatokat. Legyenek $\mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^2$ és $\mathbf{y} = [y_1, y_2]^T \in \mathbb{R}^2$ egymásnak megfelelő pontpárok a *sablon* és a *megfigyelés* képeken. Ekkor

$$\mathbf{y} = \varphi(\mathbf{x}) \Leftrightarrow \mathbf{x} = \varphi^{-1}(\mathbf{y}), \quad (15)$$

Ez az összefüggés érvényben marad, ha mindkét oldalon hat egy $\omega : \mathbb{R}^2 \rightarrow \mathbb{R}^n$ függvény [3, 12, 11]:

$$\omega(\mathbf{y}) = \omega(\varphi(\mathbf{x})) \Leftrightarrow \omega(\mathbf{x}) = \omega(\varphi^{-1}(\mathbf{y})). \quad (16)$$

Mindkét oldalon integrálva a következő egyenlethez jutunk:

$$\int_{\mathcal{F}_o} \omega(\mathbf{y}) d\mathbf{y} = \int_{\mathcal{F}_t} \omega(\varphi(\mathbf{x})) |J_\varphi(\mathbf{x})| d\mathbf{x}. \quad (17)$$

ahol $\mathcal{F}_t$ és $\mathcal{F}_o$ az objektum régiók a *sablon* és a *megfigyelt* alakzatokon, míg $|J_\varphi| : \mathbb{R}^2 \rightarrow \mathbb{R}$ jelöli a transzformáció Jacobi-determinánsát:

$$|J_\varphi(\mathbf{x})| = \begin{vmatrix} \frac{\partial \varphi_1}{\partial x_1} & \frac{\partial \varphi_1}{\partial x_2} \\ \frac{\partial \varphi_2}{\partial x_1} & \frac{\partial \varphi_2}{\partial x_2} \end{vmatrix}. \quad (18)$$

A javasolt módszer alapötlete a regisztrációs probléma visszavezetése megfelelő számú független ω függvény segítségével megkonstruált egyenletrendszerre. Mivel egy $\omega : \mathbb{R}^2 \rightarrow \mathbb{R}^n$ ($n > 1$) függvény segítségével létrehozott egyenlet felbontható n különböző egyenletre, minden megszorítás nélkül feltehetjük, hogy $n = 1$.

Tegyük fel, hogy a φ transzformációnak k paramétere van és legyen $\omega_i : \mathbb{R}^2 \rightarrow \mathbb{R}$, ($i = 1, \dots, \ell$) a felhasznált függvények halmaza. Az összes ismeretlen meghatározásához legalább k egyenletre van szükségünk, tehát $\ell \geq k$. Így a következő egyenletrendszerhez jutunk:

$$\int_{\mathcal{F}_o} \omega_i(\mathbf{y}) d\mathbf{y} = \int_{\mathcal{F}_t} \omega_i(\varphi(\mathbf{x})) |J_\varphi(\mathbf{x})| d\mathbf{x}, \quad i = 1, \dots, \ell, \quad (19)$$

amelyhez minden egyes ω_i függvény egy egyenlettel járul hozzá. Az ω_i függvények az alakzatok egy-egy színezésének tekinthetők, az integrálok pedig az ω_i függvények alakzatok fölötti térfogatát adják. Az egyenletek tehát ezen térfogatokat hasonlítják össze. Minden egyenlet újabb megszorítást ad, és az egyenletrendszer megoldása szolgáltatja a transzformáció paramétereinek becslését.

A következő transzformációs osztályokat vizsgáltuk:

1. A **síkhomográfia** egy projektív transzformáció ugyanannak a síkbeli objektumnak a képei között. Fontos szerepet játszik a számítógépes látás területén.
2. A **polinom transzformációkat** gyakran használják más transzformációs modellek közelítésére.
3. A **thin plate spline** (TPS) modellt gyakran alkalmazzák általános deformációk közelítésére.

A kiugró numerikus értékek elkerülése érdekében normalizációt hajtunk végre mind a pixel koordinátákon, mind a ω_i függvényeken. Ezen célból egyrészt mindkét alakzat koordinátáit normalizáljuk a $[-0.5, 0.5] \times [-0.5, 0.5]$ egységnyizetbe, másrészt olyan ω_i függvényeket alkalmazunk, melyek értékészlete korlátos (például $[-1, 1]$).

A normalizálások ellenére az ω_i függvények eltérő karakterisztikája miatt a különböző egyenletek különböző mértékben járulhatnak hozzá a célfüggvényhez az egyenletrendszer megoldásakor. Ezért további normalizációként minden egyenletet leosztunk egy megfelelő konstanssal, amit az adott egyenletnél alkalmazott ω_i függvény $\frac{\sqrt{2}}{2}$ sugarú kör fölötti integrálja ad:

$$N_i = \int_{\|\mathbf{x}\| \leq \frac{\sqrt{2}}{2}} |\omega_i(\mathbf{x})| d\mathbf{x}, \quad (20)$$

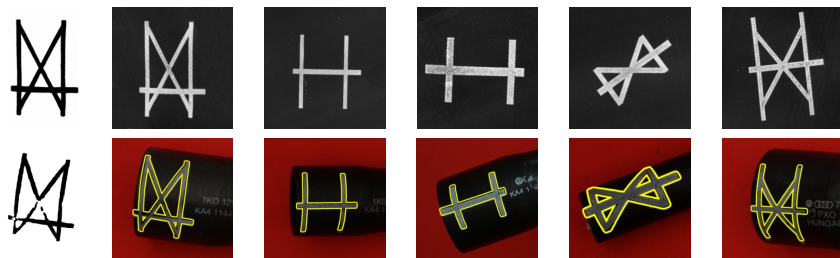
és így a (19) egyenletrendszer normalizált változata:

$$\frac{\int_{\mathcal{F}_o} \omega_i(\mathbf{y}) d\mathbf{y}}{N_i} = \frac{\int_{\mathcal{F}_t} \omega_i(\varphi(\mathbf{x})) |J_\varphi(\mathbf{x})| d\mathbf{x}}{N_i}, \quad i = 1, \dots, \ell. \quad (21)$$

A (21) egyenletrendszert folytonosan vezettük le, a gyakorlatban azonban véges felbontású digitális képekkel dolgozunk. Ezért az integrálokat közelítenünk kell véges összegekkel az előtérpíxelek felett. Jelölje az $\mathcal{F}_t$ and $\mathcal{F}_o$ folytonos halmazokhoz tartozó véges pixel halmazokat F_t és F_o . Ekkor a (21) közelíthető a következő egyenletrendszerrel:

$$\frac{1}{N_i} \sum_{\mathbf{y} \in F_o} \omega_i(\mathbf{y}) = \frac{1}{N_i} \sum_{\mathbf{x} \in F_t} \omega_i(\varphi(\mathbf{x})) |J_\varphi(\mathbf{x})|, \quad i = 1, \dots, \ell. \quad (22)$$

A keresett transzformációs paramétereket az egyenletrendszer megoldása adja. Bár az egyenletek nemlineárisak, ahogy az eredmények mutatják, hogy az egyenletrendszer hatékonyan megoldható a *Levenberg-Marquardt algoritmussal* [7]. Síkhomográfia esetében az egyenletek további három formában is felírhatóak az inverz transzformáció segítségével.



5. ábra: Regisztrációs eredmények beépítési jeleken. A felső sorban a *sablon* képként, az alsó sorban a *megfigyelés* képként használt felvételek láthatóak. A bal oldali képpár a második oszlopban szereplő képek szegmentált változatai. A *regisztrált* képek kontúrjait rávetítettem a *megfigyelés* képekre. (A képeket a ContiTech Fluid Automotive Hungária Kft. bocsájtotta rendelkezésemre.)

Ezek az egyenletek további megszorításokat jelentenek, így segítve az optimum megtalálását.

A módszer hatékonyságát megvizsgáltuk különböző $\{\omega_i\}$ függvény halmazok alkalmazása esetén. Teszteltünk például hatvány, polinom és trigonometrikus függvény halmazokat. Az eredmények alapján alacsony fokú polinom függvényeket javasunk, melyek számítási hatékonyság szempontjából is előnyösek. A síkhomográfia esetében megmutattuk, hogy az algoritmus hatékonysága javítható, ha a transzformáció Taylor sorba fejtetten változatán alkalmazzuk a keretrendszert.

A javasolt módszer hatékonyságát nagy, szintetikus képi adatbázison teszteltük mind-egyik transzformációs modell esetén. Az eredményeket összehasonlítottuk a Shape Context [2] módszer által elért eredményekkel, módszerünkkel pontosabb eredményeket értünk el. Vizsgáltuk a módszer robusztusságát 4 különböző típusú szegmentálási hibával szemben a síkhomográfia transzformáció esetén. Az eredmények alapján a javasolt módszer robusztus azokban az esetekben, amikor a szegmentálási hibák egyenletesen oszlanak el az alakzatokon. Azokban az esetekben amikor a szegmentálási hiba nagy, összefüggő régióból állt, a módszer kevésbé pontos eredményeket ad a Shape Context eredményeihez képest.

A bemutatott módszert alkalmaztuk különböző valós alkalmazási területeken, mint például kézzel írott karakterek (TPS modell) vagy KRESZ táblák regisztrációjára (síkhomográfia), és egy ipari minőség ellenőrzési problémára (probléma specifikus transzformációs modell, lásd 5. ábra).

A szerző hozzájárulása az eredményekhez

A szerző bináris alakzatok nemlineáris regisztrációjával foglalkozik. Egy általános regisztrációs keretrendszert használ, amely a regisztrációs problémát egy egyenletrend-

szer megoldására vezeti vissza. A keretrendszert különböző nemlineáris transzformációs osztályok esetében is alkalmazza, mint például síkhomográfia, polinomiális transzformációk és thin-plate-spline transzformáció. Bemutatja az implementációs részleteket és azt, hogy az egyenletek oda-vissza történő felírásával hogyan lehet javítani a módszer eredményein. A síkhomográfia esetében megmutatja, hogy az algoritmus hatékonysága javítható, ha a transzformáció Taylor sorba fejtett változatán alkalmazzuk a keretrendszert. Különböző ω függvény halmazokat vizsgál az egyenletrendszer felírásához, és összehasonlítja őket. A szerző az egyenletek normalizálásával éri el, hogy azok azonos módon járuljanak hozzá a célfüggvényhez. A szerző összehasonlítja a módszert más korszerű technikákkal szintetikus adatokon és megvizsgálja a javasolt algoritmus robusztusságát a szegmentálási hibákkal szemben. A szerző bemutatja a módszer eredményeit több valós probléma esetén, mint például röntgenképek vagy KRESZ táblák illesztése.

IV. Bináris alakzatok projektív regisztrációja affin invariánsok segítségével

Bemutatunk egy két lépéses módszert bináris képek közötti síkhomográfia transzformáció meghatározására. Az ilyen transzformáció paraméterei a 3×3 -as $\mathbf{H}$ mátrix H_{ij} elemei ($H_{33} = 1$ rögzített). A H_{31} és H_{32} paraméterek a *perspektív* torzulást írják le, míg a többiek az affin transzformáció paraméterei. A transzformáció a következőképpen bontható fel:

$$h = h^a \circ h^p \quad (23)$$

ahol $h^p : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $h^p(\mathbf{x}) = [h_1^p(\mathbf{x}), h_2^p(\mathbf{x})]^T$ egy nemlineáris transzformáció:

$$\begin{aligned} h_1^p(\mathbf{x}) &= \frac{x_1}{p_1 x_1 + p_2 x_2 + 1} \\ h_2^p(\mathbf{x}) &= \frac{x_2}{p_1 x_1 + p_2 x_2 + 1}, \end{aligned} \quad (24)$$

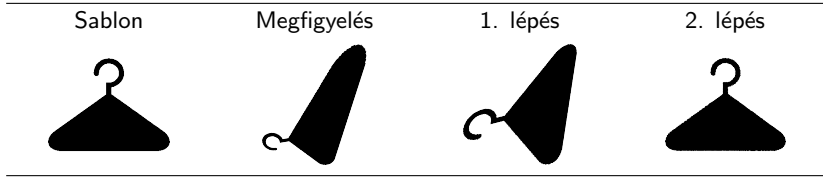
mely perspektív torzulást eredményez, míg $h^a : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $h^a(\mathbf{x}) = [h_1^a(\mathbf{x}), h_2^a(\mathbf{x})]^T$ egy affin transzformáció:

$$\begin{aligned} h_1^a(\mathbf{x}) &= a_{11}x_1 + a_{12}x_2 + a_{13} \\ h_2^a(\mathbf{x}) &= a_{21}x_1 + a_{22}x_2 + a_{23}. \end{aligned} \quad (25)$$

Így a sablon (template) és a megfigyelés (observation) alakzatok közötti összefüggés a következőképpen írható fel:

$$\mathcal{F}_t = (h^a \circ h^p)(\mathcal{F}_o) = h^a(h^p(\mathcal{F}_o)) \quad (26)$$

A javasolt módszer a perspektív h^p komponens p_i paramétereit, és az affin h^a komponens a_i paramétereit két külön lépésben határozza meg (lásd 6. ábra), végül a (23) egyenlet alapján megkapjuk a h transzformáció H_{ij} paramétereit.



6. ábra: A regisztrációs folyamat: Az első lépés csak a perspektív torzulást állítja helyre a megfigyelt képen, míg a második lépés meghatározza az affin transzformációt így megkapjuk az eredeti sablon képet.

Ha a (26) egyenlet teljesül, akkor bármely affin invariáns $I : \mathbb{R}^2 \rightarrow \mathbb{R}$ függvényre:

$$I(\mathcal{F}_t) = I(h^p(\mathcal{F}_o)). \quad (27)$$

Egy független affin invariáns $I_i : \mathbb{R}^2 \rightarrow \mathbb{R}$, $i = 1 \dots n$ függvényhalmazt véve a következő egyenletrendszert kapjuk:

$$I_i(\mathcal{F}_t) = I_i(h^p(\mathcal{F}_o)). \quad (28)$$

A perspektív torzulás h^p paramétereit ezen egyenletrendszer megoldásaként kapjuk. Bár könnyen látható, hogy ennek egyenletrendszernek nincs közvetlen meghatározható megoldása, az eredmények mutatják, hogy az általunk használt általános nemlineáris optimalizációs módszerrel hatékonyan megoldható.

A javasolt módszert úgynevezett Affin Momentum Invariánsok (Affine Moment Invariants, AMIs [5]) esetére dolgoztuk ki, mivel használatukkal hatékonyan számítható egyenletekhez jutunk. A (28) egyenletrendszer bal oldalai nem függenek a h^p paramétereiktől, tehát közvetlenül számíthatóak a sablon alakzat pontkoordinátái segítségével. Az $\mathcal{F}$ alakzat $(r + s)$ rendű m_{rs} geometriai momentuma:

$$m_{rs}(\mathcal{F}) = \int_{\mathcal{F}} x_1^r x_2^s dx. \quad (29)$$

Az affine momentum invariánsok a centrális momentumokon alapszanak:

$$\mu_{rs}(\mathcal{F}) = \int_{\mathcal{F}} (x_1 - c_1)^r (x_2 - c_2)^s dx \quad (30)$$

ahol az alakzat súlypontjait a következőképpen számoljuk:

$$c_1 = m_{10}(\mathcal{F})/m_{00}(\mathcal{F}) \quad \text{and} \quad c_2 = m_{01}(\mathcal{F})/m_{00}(\mathcal{F}). \quad (31)$$

Például, az első kettő affin momentum invariáns:

$$\begin{aligned} I_1 &= (\mu_{20}\mu_{02} - \mu_{11}^2)/\mu_{00}^4 \\ I_2 &= (-\mu_{30}^2\mu_{03}^2 + 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} - 4\mu_{30}\mu_{12}^3 - 4\mu_{21}^3\mu_{03} + 3\mu_{21}^2\mu_{12}^2)/\mu_{00}^{10}. \end{aligned} \quad (32)$$

Most megmutatjuk, hogy adott h^p paraméterek esetén a (28) egyenletrendszer jobb oldalai hogyan számíthatóak anélkül, hogy ténylegesen előállítanánk a $h^p(\mathcal{F}_o)$ alakzatot. Ehhez a h^p transzformáció $|J_{h^p}|$ Jacobi determinánsát használjuk. Egy h^p transzformációval eltranszformált $\mathcal{F}$ alakzat geometriai momentuma a következőképpen számítható:

$$m_{rs}(h^p(\mathcal{F})) = \int_{\mathcal{F}} [h_1^p(\mathbf{x})]^r [h_2^p(\mathbf{x})]^s |J_{h^p}(\mathbf{x})| d\mathbf{x} \quad (33)$$

ahol a perspektív transzformáció Jacobi determinánsa:

$$|J_{h^p}(\mathbf{x})| = \frac{1}{(p_1 x_1 + p_2 x_2 + 1)^3},$$

Egy perspektíven torzított $h^p(\mathcal{F})$ alakzat centrális momentumait a következőképpen számíthatjuk:

$$\mu_{rs}(h^p(\mathcal{F})) = \int_{\mathcal{F}} [h_1^p(\mathbf{x}) - c_1]^r [h_2^p(\mathbf{x}) - c_2]^s |J_{h^p}(\mathbf{x})| d\mathbf{x}, \quad (34)$$

ahol

$$c_1 = m_{10}(h^p(\mathcal{F}))/m_{00}(h^p(\mathcal{F})) \quad \text{and} \quad c_2 = m_{01}(h^p(\mathcal{F}))/m_{00}(h^p(\mathcal{F})). \quad (35)$$

Adott p_1 és p_2 paraméterek esetén a (28) egyenletrendszer jobboldalain lévő $I(h^p(\mathcal{F}))$ affin momentum invariánsok a (34) egyenletben leírt centrális momentumok segítségével számíthatók. Így a momentumok számításához elkerülhető az eltranszformált $h^p(\mathcal{F})$ alakzat tényleges előállítása, ami nagyon időigényes feladat lenne.

Miután a perspektív paramétereket meghatároztuk, a következő lépésben az $\mathcal{F}_t$ és $h^p(\mathcal{F}_o)$ alakzatok közötti h^a transzformációt kell megbecsülnünk. Ezen célból a [3] módszert használjuk, melyet úgy módosítottunk, hogy a $|J_{h^p}|$ determináns használatával elkerüljük a $h^p(\mathcal{F}_o)$ alakzat előállítását. Az a_{ij} affin paraméterekre a következő egyenletrendszer írható fel:

$$\int_{\mathcal{F}_t} y_k^n d\mathbf{y} = |J_{h^a}| \sum_{i=1}^n \binom{n}{i} \sum_{j=0}^i \binom{i}{j} a_{k1}^{n-i} a_{k2}^{i-j} a_{k3}^j \int_{\mathcal{F}_o} h_1^p(\mathbf{x})^{n-i} h_2^p(\mathbf{x})^{i-j} |J_{h^p}(\mathbf{x})| d\mathbf{x}, \quad (36)$$

ahol $n = 1, 2, 3$ és $k = 1, 2$. Ez az egyenletrendszer 6 darab 3-ad fokú polinom egyenletből áll amely így megoldható a 6 ismeretlenre. Az affin transzformáció Jacobi determinánsa állandó az egész síkon, és könnyen számítható az alakzatok területeinek arányaként:

$$|J_{h^a}| = \frac{\int_{\mathcal{F}_t} dy}{\int_{\mathcal{F}_o} |J_{h^p}(\mathbf{x})| d\mathbf{x}} \quad (37)$$

Bár a (36) egyenletrendszernek több megoldása lehet, közülük könnyen kiválaszthatjuk azt a valós gyököt, amellyel a (37) egyenletben meghatározott determinánshoz tartozik. Megjegyezzük, hogy a megoldás nem egyértelmű, ha az alakzat affin-szimmetrikus.

Összetéve a h^p perspektív és a h^a affin transzformációkat megkapjuk a keresett h síkhomográfiát. Vizsgálataink során a I_3, I_4, I_5, I_6 invariánsokat használtuk és a (28)



7. ábra: Regisztrációs eredmények KRESZ táblákon. A megfigyelés képek az első sorban, alattuk pedig a megfelelő *sablon* képek találhatók. A regisztrációs eredmények kontúrjait rávetítettük a *sablon* képekre.

egyenletrendszer megoldásához az úgynevezett *differenciál evolúció* módszert használtuk. Szintetikus adatbázison futtatott tesztek megmutatták, hogy a javasolt módszer jobb eredményeket ad az előző pontban bemutatott algoritmushoz képest, amely nem képes kezelni azokat a helyzeteket, amikor az alakzatok között 90 foknál nagyobb mértékű elforgatás van. Mivel az affin momentum invariánsok érzékenyek a szegmentálási hibára, a módszer robusztusságát ezen hibákkal szemben tesztelni kell a későbbiekben. Mindazonáltal sikerült jó eredményeket elérnünk több valós KRESZ táblákat ábrázoló képpáron is (lásd 7. ábra).

A szerző hozzájárulása az eredményekhez

A szerző egy affin momentum invariánsokon alapuló módszert dolgozott ki bináris alakzatok közötti síkhomográfia transzformáció paramétereinek becslésére. Megmutatja, hogyan lehet felbontani a transzformációt egy perspektív és egy affin összetevőre. A paraméterek meghatározása két, egymást követő lépésben történik. Az első lépés csak a perspektív torzulást állítja helyre a megfigyelt képen, míg a második lépés meghatározza az affin transzformációt így megkapjuk az eredeti sablon képet. A szerző összehasonlítja módszert az előző pontban bemutatott technikával és bemutat valós képeken elért eredményeket is.

Hivatkozások

- [1] Kees Joost Batenburg. A network flow algorithm for reconstructing binary images from discrete x-rays. *Journal of Mathematical Imaging and Vision*, 27(2):175–191, 2007.
- [2] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recog-

nitition using shape context. *Transaction on Pattern Analysis and Machine Intelligence*, 24(4):509–522, April 2002.

- [3] Csaba Domokos and Zoltan Kato. Parametric estimation of affine deformations of planar shapes. *Pattern Recognition*, 43:569–578, March 2010.
- [4] Csaba Domokos, Jozsef Nemeth, and Zoltan Kato. Nonlinear shape registration without correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):943–958, May 2012.
- [5] Jan Flusser, Tomás Suk, and Barbara Zitová. *Moments and Moment Invariants in Pattern Recognition*. Wiley & Sons, October 2009.
- [6] Jeongtae Kim and Soohyun Jang. High order statistics based blind deconvolution of bi-level images with unknown intensity values. *Optics Express*, 18(12):12872–12889, June 2010.
- [7] Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2):431–441, 1963.
- [8] Jozsef Nemeth. Recovering projective transformations between binary shapes. In *Proceedings of Advanced Concepts for Intelligent Vision Systems*, volume 7517, pages 374–383, Brno, Czech Republic, 2012.
- [9] Jozsef Nemeth. Discrete tomography with unknown intensity levels using higher-order statistics. *Journal of Mathematical Imaging and Vision*, 2015. Online first, DOI: 10.1007/s10851-015-0581-0.
- [10] Jozsef Nemeth and Peter Balazs. Restoration of blurred binary images using discrete tomography. In *Proceedings of Advanced Concepts for Intelligent Vision Systems*, volume 8192, pages 80–90, Poznan, Poland, 2013.
- [11] Jozsef Nemeth, Csaba Domokos, and Zoltan Kato. Nonlinear registration of binary shapes. In *Proceedings of International Conference on Image Processing*, pages 1101–1104, Cairo, Egypt, November 2009. IEEE.
- [12] Jozsef Nemeth, Csaba Domokos, and Zoltan Kato. Recovering planar homographies between 2D shapes. In *Proceedings of International Conference on Computer Vision*, pages 2170–2176, Kyoto, Japan, September 2009. IEEE.
- [13] József Németh. Síkhomográfia paramétereinek becslése bináris képeken. In *Proceedings of National Scientific Students' Associations Conference*, April 2009. Supervisors: Zoltán Kató and Csaba Domokos. Note: in Hungarian.

- [14] James B. Orlin. A polynomial time primal network simplex algorithm for minimum cost flows. In *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '96, pages 474–481, Philadelphia, PA, USA, 1996. SIAM.
- [15] Wim van Aarle, Kees Joost Batenburg, and Jan Sijbers. Automatic parameter estimation for the discrete algebraic reconstruction technique (DART). *IEEE Transactions on Image Processing*, 21(11):4608–4621, 2012.