

Személyek lokalizálása és magasságuk becslése többszörös születés és halál dinamikával többkamerás környezetben*

Utasi Ákos, Benedek Csaba

Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutatóintézet
Elosztott Események Elemzése Kutatócsoport
1111 Budapest, Kende u. 13-17.

Kivonat A jelen cikk egy új eszközt mutat be személyek lokalizálásához többkamerás környezetben, kalibrált kamerákat használva. Ezenfelül becslést adunk az egyes személyek magasságára is. A bemutatott módszer bemenete jelenleg a testek előtér sziluettjei, de ez később könnyedén kicserélhető más, széles körben alkalmazott, objektum (pl. fej, láb, test) detektáló eljárások kimeneteire. A módszer első lépésben az összes előtér képpontot a föld síkjára, valamint ezzel párhuzamos, különböző magasságú síkokra vetíti. Ezután jellemzőket nyerünk ki, melyek a kétdimenziós képlétrehozás fizikai tulajdonságain alapulnak. A végső konfigurációt (pozíciók és magasságok) egy iteratív sztochasztikus optimalizáló eljárással, a többszörös születés és halál dinamikával, kapjuk meg.

1. Bevezetés

Személyek detektálása, valamint lokalizálása a legtöbb megfigyelő alkalmazásban kulcsprobléma, és a más személyek, valamint különböző statikus objektumok által okozott takarások magas aránya miatt még mindig nagy kihívást jelent zsúfolt, városi helyszíneken. Ennek következményeként egy objektum sziluett maszk esetén nem feltételezhetjük, hogy az csak egyetlen személyhez tartozik, valamint az egyes maszkok akár darabokra is széteshetnek. Ilyen körülmények között az egykamerás lokalizálás vagy követés gyakran megvalósíthatatlan. A javasolt módszer képes több kamera használatával pontosan lokalizálni a személyek pozícióját a talaj síkján. Ennélfogva sok különböző, magas szintű gépi látás feladatban alkalmazható, mint például helyszín értelmezés, több objektum követése, vagy csoportok, illetve tömeg viselkedésének elemzése. Ezenkívül a módszerünk még az egyes személyek magasságát is képes megbecsülni. A javasolt módszer feltételezi, hogy a helyszínt több, kalibrált kamera figyeli meg, illetve a kinyert előtér sziluettek rendelkezésre állnak. Ezeket a sziluetteket a talaj síkjára, valamint ezzel párhuzamos síkokra vetítjük. A bemutatott módszer semmilyen színi, vagy alak modellt nem használ a helyszínen látható személyek megkülönböztetésére. Ehelyett kihasználjuk a több kamera által nyújtott előnyöket, és a több síkra történő

* A cikk eredményei az alábbi publikációban jelent meg: Á. Utasi, Cs. Benedek: Multi-camera people localization and height estimation using multiple birth-and-death dynamics. In Proceedings of The 10th International Workshop on Visual Surveillance (2010)

vetítés eredményeit felhasználva, két hasonló geometriai jellemzőt nyerünk ki minden egyes 2D pozícióban: egyiket a talaj síkján, a másikat pedig a többi párhuzamos síkon. Végül a személyek optimális konfigurációjának megtalálásához a kinyert jellemzőket egy sztochasztikus optimalizáló eljárásban használjuk egy geometriai interakciós taggal.

A cikk többi része a következőképpen épül fel. A 2 fejezetben röviden bemutatjuk a többkamerás személy lokalizáláshoz kapcsolódó munkákat. A javasolt módszert a 3 fejezet tárgyalja. Módszerünket egy publikus adatbázison értékeltük ki, melynek eredményeit a 4 fejezet tartalmazza. A cikket végül egy összefoglaló fejezet zárja.

2. Kapcsolódó munkák

Az elmúlt évtizedekben az egykamerás személy detektálás és követés jelentős fejlődésen ment keresztül. A [1] cikk átfogó áttekintést ad a korszerű módszerekről. Ezek a módszerek azonban csak korlátozottan képesek a zsúfolt helyszíneket kezelni, ahol a takarások aránya magas. Ilyen körülmények között a többkamerás módszerek jobb megoldást biztosítanak, sokkal pontosabban becslik meg az egyes személyek pozícióit. Mikic és mtsai [2] egy előtér folt alapú módszert javasoltak, ahol egy objektumot minden egyes nézeten egy folt reprezentál. Az objektum 3D súlypontját becslik meg, amit egy túlhatározott lineáris rendszer legkisebb négyzetek megoldásával származtatnak, a mérések pedig az egyes nézeteken található koordináták. A [3] cikk szerzői az egyes nézeteken történő szegmentáláshoz a személyeket pozíciójukkal és színi megjelenésükkel modellezik. Ez egyben elősegíti az egyes objektumokhoz tartozó előtér régiók elkülönítését is. A [4] cikkben a kinyert előtér foltok legalsó képpontjainak súlypontját vetítik a talaj síkjára. Ezt az információt a 2D befoglaló téglalap sarkaival kiegészítve hoznak létre egy mozgás modellt. A [5] cikkben bemutatott módszer azt feltételei, hogy az egyes objektumokat több, a fej magasságába helyezett, kamera figyeli meg. A talaj síkja egy diszkrét ráccsal van felosztva, majd minden rács pozícióban egy átlagos gyalogos magasságával megegyező méretű téglalapot definiálnak. Ezeket a téglalapokat vetítik az egyes nézetekre, ezzel modellezve, hogy az egyes személyek milyen pozíciót foglalnak el a rácson. [6] módszere több kamera nézet bizonyítékait egyesíti, hogy megtalálja a helyszín azon pozícióit, ahol személyek találhatóak. Ehhez a homográfia foglaltsági korlátozást javasolják, ami az egyes nézetek előtér valószínűségi információit egyesíti a személyek lokalizálásához. Ehhez egy referencia nézetet választanak ki, melyre a többi nézet valószínűség értékeit vetítik rá. A több síkra vetítést az olyan speciális esetek kezeléséhez használják, amikor a foglaltság időszakos, pl. amikor egy személy ugrik, vagy fut. A mi módszerünk szintén több síkra vetítést használ, azonban teljesen más céllal. Az egyes kamera nézetek előtér maszkjait vetítjük a talaj síkjára, valamint ezzel párhuzamos síkokra, amit ezután jellemző kinyeréshez használunk. Esetünkben az egyes személyek hipotetikus pozíciója és magassága mindig két különböző síkon számított bizonyíték kombinációja: egyrészt a talaj síkján, másrészt a fej hipotetikus síkján. Ebből egy olyan diszkriminatív jellemzőt kapunk, amihez a 3D objektum 2D képleképzésének tulajdonságait használjuk fel. A [7] cikkben bemutatott módszer hosszútávú statisztika tanulást alkalmaz a térbeli magasság eloszlás meghatározásához, amit a mozgó személy magasságának becsléséhez használnak fel. A mi módszerünk-

ben nincs szükség hosszú távú tanulásra, hiszen minden egyes személy magasságát az optimalizálási folyamat során becsüljük meg a pozíciójukkal együtt.

A másik fontos kérdés az objektum modellezéshez kapcsolódik. A *direkt* módszerek az objektumokat primitívekből építik fel, ilyenek lehetnek pl. sziluett foltok [8], vagy szegmentált objektum részek. Habár ezek a módszerek nagyon gyorsak, nem működnek, amennyiben a primitíveket nem lehet megbízhatóan detektálni. Ettől eltérően az *inverz* módszerek [9] illeszkedés értéket rendelnek hozzá minden lehetséges objektum konfigurációhoz, majd egy optimalizálási folyamat keresi meg a legnagyobb konfidenciával rendelkező konfigurációt. Így rugalmas megjelenési modellek alkalmazására nyílik lehetőség, és egyszerűen lehet akár alakzati priort vagy objektum kölcsönhatási tényezőzt a modellbe illeszteni. Azonban a populáció magas dimenziójú terében a keresés számítási költsége nagyon nagy, továbbá az illeszkedési függvény lokális maximuma félrevezetheti az optimalizálást.

A javasolt modellben egyesítjük az alacsony szintű és az objektum szintű megközelítések előnyeit. Az alkalmazott Többszörös Születés és Halál (Multiple Birth and Death) technika az objektum populációt két lépés váltakozásával alakítja ki: objektum javaslat (születés) és eltávolítás (halál). A két lépés szimulált lehűtés keretrendszerben fut, valamint az objektumok ellenőrzése robusztus *inverz* megközelítést követ.

3. Javasolt módszer

A javasolt módszer bemenete az azonos helyszínt megfigyelő, kalibrált kamrák képeiből kinyert előtér sziluettek. A kalibráláshoz Tsai módszerét [10] és szoftverét használjuk. A jelenlegi implementációban az előtér maszkok kinyeréséhez először Kevert Gauss Modellt (KGM) illesztünk [11] minden egyes pontban a képek színi koordinátáira, majd az így kapott modelleket egy paraméter frissítés nélküli előtér-háttér elkülönítő módszerben [12] használjuk. Módszerünk alapötlete, hogy a kinyert sziluetteket egyrészt a talaj síkjára, másrészt ezzel párhuzamos, a személyekével megegyező magasságú síkra vetítjük (1. ábra). A vetítés által egy olyan jól kivehető vizuális jellemzőt kapunk, ami egy virtuális madártávlati nézőpontból a talaj síkján figyelhető meg. Azonban az egyes személyek magasságairól nincs előzetes információnk, valamint a különböző személyek magassága szintén különbözhet. Ezért vetítjük a sziluett maszkokat több párhuzamos síkra, melyek magassága a tipikus emberi magasság tartományába esik. A zsúfolt helyszíneken az átfedések aránya nagyon magas lehet, emiatt sérülhet a hipotézisünk. Ezt úgy oldjuk meg, hogy az összes nézet vetítéseinek eredményeit egyesítjük egy közös síkra. A javasolt módszer az alábbi fő lépésekből épül fel, melyeket a későbbi fejezetek tárgyalnak részletesen:

1. *Vetítés több síkra:* A sziluetteket a talaj, és ezzel párhuzamos, különböző magasságú síkokra vetítjük.
2. *Jellemző kinyerés:* Minden sík minden pozíciójában olyan jellemzőt nyerünk ki, amely pozitív választ ad valós magasság és valós pozíció esetén. A jellemző alapja a 2D képleképzés fizikai tulajdonságai és a több síkra vetítés.
3. *Sztokasztikus optimalizálás:* Az optimális konfigurációt egy iteratív folyamatban keressük, amihez kinyert jellemzőket és geometriai interakciós tagokat használunk.

3.1. Vetítés több síkra

Jelölje P_0 a talaj síkját, valamint P_z azt az ezzel párhuzamos síkot, amely P_0 -tól z távolságra helyezkedik el. A javasolt módszer az első lépésben a detektált sziluetteket a kamerák kalibrációs adatait felhasználva a P_0 síkra és különböző P_z ($z > 0$ magasságba tolt) síkokra vetíti. Az 1. ábra szemlélteti ennek hatékony számítását, azaz csak a P_0 síkra vetítünk, majd a következő összefüggést használjuk. Legyen (x_c, y_c) egy kamera pozíciója a talaj síkján, magassága pedig h_c . Jelölje továbbá (x_0, y_0) egy sziluett adott pontjának a talaj síkjára vetített pozícióját, azaz $h_0 = 0$. Ekkor ugyanannak a pontnak egy másik párhuzamos z magasságú síkra vetített (x_z, y_z) pozícióját a következőképpen fejezhetjük ki:

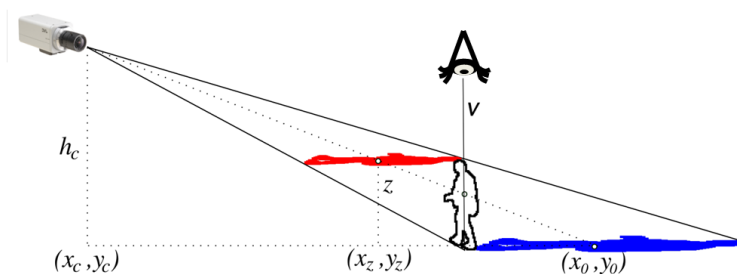
$$x_z = x_0 - (x_0 - x_c) z/h_c \quad (1)$$

$$y_z = y_0 - (y_0 - y_c) z/h_c \quad (2)$$

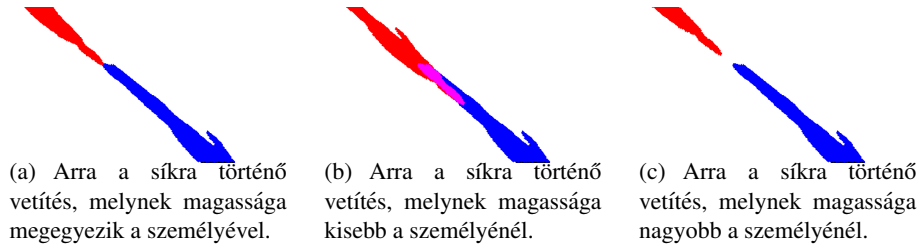
Az 1. ábrán, valamint a cikk további részeiben a sziluettek P_0 síkra történő vetületét kézzel, és egy P_z síkra történőt pirossal jelöljük.

3.2. Jellemző kinyerés

Az egyes személyek hipotetikus pozíciója és magassága a 3D objektum 2D vetületének fizikai tulajdonságain alapszik. Tekintsük az 1. ábrán látható h magasságú személyt, akinek a sziluettjét a P_0 talaj síkra (kézzel jelölve), valamint a személlyel megegyező magasságú P_z síkra (azaz $z = h$, pirossal jelölve) vetítettük. Legyen továbbá v a személy függőleges tengelye, amely merőleges a P_0 síkra. Ekkor azt figyelhetjük meg, hogy a sziluett pontok, amelyeket a $P_z|_{z=h}$ síkra vetítettünk a kamera irányában, míg a P_0 síkon levő sziluett vetületek éppen v másik oldalán helyezkednek el. A 2. ábrán tovább vizsgáljuk ezt a jellemzőt: kiválasztunk egy P_z fölötti nézőpontot, majd ebből merőlegesen a talaj irányába nézve jelenítjük meg a helyszínt. Itt a P_z és P_0 síkokra vetített sziluett pontokat egy közös $x - y$ síkon egyszerre ábrázoltuk pirossal valamint kézzel, az átfedésekhez pedig lila színt használtunk. A 2(a) ábrán jól látható, hogy amennyiben a magasság becslésünk helyes (azaz $z = h$), akkor a két terület épphogy érinti egymást a $p = (x, y)$ pontban, ami megfelel a személy talaj síkján található



1. ábra. A sziluetteket a talaj síkjára (kék), és ezzel párhuzamos síkokra (piros) vetítjük.



2. ábra. Jellemzőink alapja a 2D képleképzés tulajdonságai és a több síkra történő vetítés. Egy sziluett talaj síkra történő vetületét kézzel, a három különböző magasságú P_z síkra történő vetületét pirossal jelöltük.

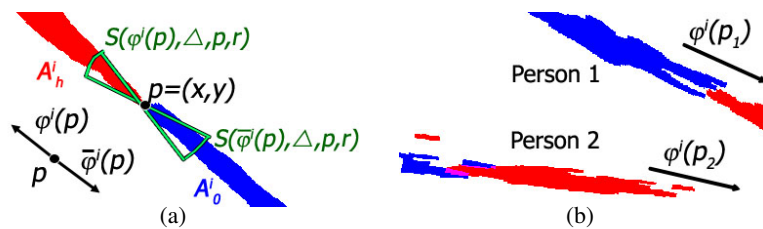
pozíciójának. Azonban amikor a z értékét alulbecsüljük (azaz $z < h$), akkor a két sziluett átfedi egymást, ahogy a 2(b) ábra mutatja. Végül amikor a z értékét túlbecsüljük (azaz $z > h$), a sziluettek eltávolodnak egymástól, ezt szemlélteti a 2(c) ábra.

Ezután létrehozunk egy illeszkedési függvényt, mely a helyszínen található $p = (x, y)$ talaj pozíciója és h magasságú objektum feltevésünket kiértékeli. Ehhez több kamerából származó információt használunk. Legyen (x_c^i, y_c^i) az i -edik kamera pozíciója a talaj síkján. Az i -edik kamera p ponttól mért vízszintes irányát a $\varphi^i(p)$ szöggel írjuk le:

$$\varphi^i(p) = \arctan\left(\frac{y - y_c^i}{x - x_c^i}\right). \quad (3)$$

Használni fogjuk továbbá az ellentétes irány definícióját is: $\bar{\varphi}^i(p) = \varphi^i(p) + \pi$. A két irányt a 3(a) ábra illusztrálja.

A fenti megfigyelések alapján egy (x, y, h) objektum hipotézis releváns az i -edik kamera adatai alapján, amennyiben a következő két feltétel teljesül. Elsőként, sziluett pontokat kell találnunk a P_0 síkon (azaz a kézzel jelölt pontok) $p = (x, y)$ pont környezetében a $\bar{\varphi}^i(p)$ irányban, másrésztől büntetjük azokat a pontokat, amelyek ugyanazon környezet $\varphi^i(p)$ irányába esnek. Ezeket a feltételeket figyelembe véve definiáljuk az



3. ábra. (a) Jelölések és területek melyeket az $f_0^i(p)$ és $f_z^i(p)$ jellemzők számításához használunk. (b) Sziluett vetületek a P_0 és egy adott z távolságú P_z síkon egy két személyt tartalmazó helyszínen. Az 1. személy magasságát pontosan megtaláltuk (azaz $h_1 = z$), azonban a 2. személy magasságát alulbecsültük. (azaz $z < h_2$).

$f_0^i(p)$ jellemzőt a talaj síkján:

$$f_0^i(p) = \frac{\mathbf{Ter}(A_0^i \cap S(\bar{\varphi}^i(p), \Delta, p, r)) - \alpha \cdot \mathbf{Ter}(A_0^i \cap S(\varphi^i(p), \Delta, p, r))}{\mathbf{Ter}(S(\bar{\varphi}^i(p), \Delta, p, r))}, \quad (4)$$

ahol A_0^i az i -edik kamera modellel a P_0 síkra vetített sziluett pontok halmaza; $\mathbf{Ter}(\cdot)$ a terület, $S(\bar{\varphi}, \Delta, p, r)$ és $S(\varphi, \Delta, p, r)$ jelöli a p középpontú körcikkeket a $[\bar{\varphi} - \Delta; \bar{\varphi} + \Delta]$ illetve $[\varphi - \Delta; \varphi + \Delta]$ szög tartományban (zöld színnel jelölve a 3(a) ábrán), továbbá r egy előzetesen beállított konstans sugár paraméter.

A fentiekhez hasonlóan bevezetjük az $f_z^i(p)$ jellemzőt a P_z síkon, a $p = (x, y)$ pont körül a $\varphi^i(p)$ irányban:

$$f_z^i(p) = \frac{\mathbf{Ter}(A_z^i \cap S(\varphi^i(p), \Delta, p, r)) - \alpha \cdot \mathbf{Ter}(A_z^i \cap S(\bar{\varphi}^i(p), \Delta, p, r))}{\mathbf{Ter}(S(\varphi^i(p), \Delta, p, r))}. \quad (5)$$

Ezután mind az $f_0^i(p)$, mind az $f_z^i(p)$ jellemzőt levágjuk, hogy a $[0, \bar{f}]$ tartományon vegyenek fel értékeket, végül normalizáljuk \bar{f} -el. Itt \bar{f} azt vezérli, hogy mekkora terület arány kell a maximális kimenet eléréséhez.

Amennyiben az (x, y, h) paraméterekkel definiált objektum teljesen látható az i -edik kamerából, akkor az $f_0^i(p)$ és $f_z^i(p)$ jellemzők nagy értéket vesznek fel a $p = (x, y)$ pontban és $z = h$ magasságban. Azonban a rendelkezésre álló nézeteken más személyek vagy statikus objektumok teljesen vagy részlegesen kitakarhatják egy személy lábát vagy fejét, ami erősen rontja a jellemzők értékét. Habár az $f_0^i(p)$ és $f_z^i(p)$ jellemzők az egyes nézeteket tekintve gyengék, létrehozhatunk egy erős osztályozót amennyiben az összes kamera adatát együttesen tekintjük és kiszámítjuk az egyes nézeteken számított jellemzők átlagainak szorzatát:

$$f(p, z) = \sqrt{\frac{1}{N} \sum_{i=1}^N f_0^i(p) \times \frac{1}{N} \sum_{i=1}^N f_z^i(p)}. \quad (6)$$

Két kamera nézetet használva a 4. ábra szemlélteti a 6. egyenlet kimenetét.

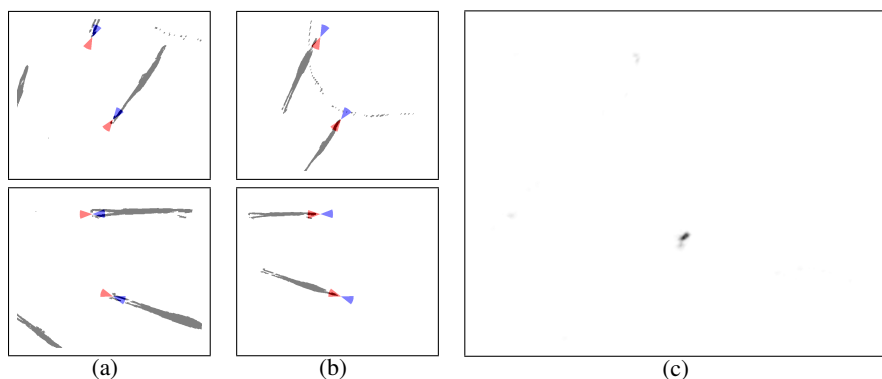
A fenti jellemző definíciók után a helyszínen található személyek megtalálását egy globális optimalizálási folyamat végzi. Mivel a személyek száma szintén ismeretlen, és egy személyt az x , y és h paraméterek jellemeznék, a konfigurációs tér dimenziója nagy. Emiatt hatékony optimalizálási eljárás szükséges.

3.3. Jelölt pontfolyamat modell

Célunk, hogy detektáljuk és elkülönítsük a helyszínen lévő személyeket és meghatározzuk a helyzetüket és magasságukat. Ezért egy egyszerűsített modellt használunk: az embereket a befoglaló hengerekkel írjuk le a 3D térben. Tegyük fel, hogy a földfelszín sík és az emberek a talajon állnak. Egy adott u objektum hengert ekkor az $x(u)$, $y(u)$ talajkoordinátái, valamint a $h(u)$ magasság határozzák meg, ahogy az 5(a) ábra mutatja.

Legyen \mathcal{H} az u objektumok tere. Az Ω konfigurációs teret az alábbi módon definiáljuk [9]:

$$\Omega = \bigcup_{n=0}^{\infty} \Omega_n, \quad \Omega_n = \{\{u_1, \dots, u_n\} \in \mathcal{H}^n\}. \quad (7)$$

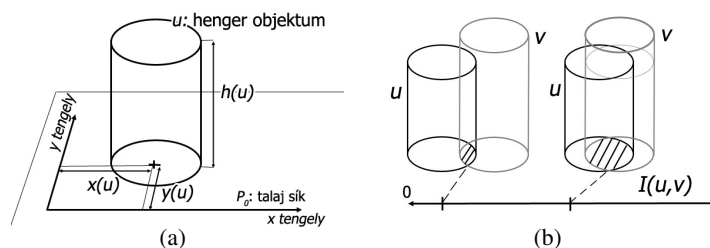


4. ábra. (a) Két kamera nézetéről kinyert jellemzők a talaj síkján a 4. egyenletet használva. (b) Kinyert jellemzők a $P_{168\text{cm}}$ síkon az 5. egyenletet használva. (c) A 6. egyenlettel az (a) és (b) értékekből számított $f(\cdot, 168\text{cm})$ egyesített jellemzők. A kamera irányát piros, az ellentétes irányt kék szín jelöli.

Legyen ω egy tetszőleges $\{u_1, \dots, u_n\} \in \Omega$ objektum konfiguráció. Definiálunk egy \sim szomszédossági relációt \mathcal{H} -ban: $u \sim v$, ha a hengerek metszik egymást. A globális képi bemenetre \mathcal{D} -vel hivatkozunk, ami az előtérmaszkokat tartalmazza minden kameranézetben és a kalibrációs mátrixokat.

Bevezetünk egy inhomogén bemenetfüggő energiafüggvényt a konfigurációs téren: $\Phi_{\mathcal{D}}(\omega)$, ami *negatív valószínűségi* értéket rendel valamennyi lehetséges objektum populációhoz. Az energiát felosztjuk bemenetfüggő ($J_{\mathcal{D}}$) és prior (I) részekre:

$$\Phi_{\mathcal{D}}(\omega) = \sum_{u \in \omega} J_{\mathcal{D}}(u) + \gamma \cdot \sum_{\substack{u, v \in \omega \\ u \sim v}} I(u, v), \quad (8)$$



5. ábra. (a) Henger objektumok modellezik a személyeket a 3D térben. A pozíciót és a magasságot becsüljük. (b) Hengerek metszete a 3D térben, mint geometriai interakciós tag.

ahol $J_{\mathcal{D}}(u) \in [-1, 1]$, $I(u, v) \in [0, 1]$ és γ súlyozó tényező a két energiateg között. Az optimális konfigurációt Maximum Likelihood (ML) becsléssel kapjuk meg:

$$\omega_{\text{ML}} = \arg \min_{\omega \in \Omega} [\Phi_{\mathcal{D}}(\omega)] .$$

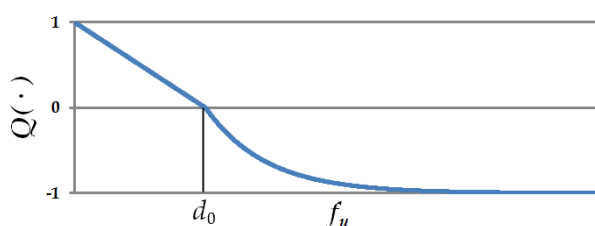
A következő kulcsfeladat az I prior és a $J_{\mathcal{D}}$ bemenetfüggő potenciálfüggvények megfelelő definiálása úgy, hogy az ω_{ML} konfiguráció hatékonyan becslje az emberek valódi csoportját a helyszínen. Mindenekelőtt el kell kerülnünk azon konfigurációkat, ahol sok objektum azonos, vagy nagy mértékben átlapolódó pozícióban található. Ezért az $I(u, v)$ interakciós potenciálokkal prior geometriai kényszereket valósítunk meg, azaz büntetjük a hengerek metszetét a 3D térben (5(b) ábra):

$$I(u, v) = \mathbf{Ter}(u \cap v) / \mathbf{Ter}(u \cup v) . \quad (9)$$

Másrésről, a $J_{\mathcal{D}}(u)$ unáris potenciálok egy adott $u = (x, y, h)$ objektum jelöltet jellemeznek a környezet képi adatai alapján, de függetlenül a populáció többi objektumától. A negatív potenciállal rendelkező hengereket *attraktív objektumoknak* hívjuk. Figyelembe véve a 8. egyenletet megfigyelhetjük, hogy az optimális populáció kizárólag attraktív objektumokból kell, hogy álljon: ha $J_{\mathcal{D}}(u) > 0$, akkor u -t törölve a konfigurációból kisebb $\Phi_{\mathcal{D}}(\omega)$ globális energiaértékhez jutunk. Ezen a ponton felhasználjuk modellünkben az $f_u = f(p(u), h(u))|_{p(u)=(x(u), y(u))}$ jellemzőt, melyet részletesen a 3.2 fejezetben tárgyaltunk. Emlékezzünk, hogy az f_u egyesített jellemző egy adott u személy-hipotézist értékel ki a kamerarendszer képei alapján úgy, hogy magas f_u értékek tartoznak az attraktív objektum jelöltekhez. Ezért a jellemzőteret egy monoton csökkenő függvénnyel a $[-1, 1]$ intervallumba vetítjük (6. ábra):

$$J_{\mathcal{D}}(u) = Q(f_u, d_0, D) = \begin{cases} \left(1 - \frac{f_u}{d_0}\right) & \text{ha } f_u < d_0 \\ \exp\left(-\frac{f_u - d_0}{D}\right) - 1 & \text{ha } f_u \geq d_0 \end{cases} \quad (10)$$

ahol d_0 és D paraméterek. Következésképpen, u objektum attraktív a $J_{\mathcal{D}}(u)$ tag alapján akkor és csak akkor, ha $f_u > d_0$, míg D adat normalizálást végez.



6. ábra. A $Q(f_u, d_0, D)$ függvény rajzolata.

3.4. Optimalizálás

Az optimális objektum konfigurációt a Többszörös Születés és Halál algoritmussal [9] becsüljük, mely a következő lépéseket tartalmazza:

Inicializáció: induljunk ki üres populációból, azaz $\omega = \emptyset$, és illesszünk egy 2D képpontrácsot a P_0 talaj síkra.

Fő program: állítsuk be a b_0 születési arányt, a kezdeti $\beta = \beta_0$ inverz hőmérséklet paramétert, és a $\delta = \delta_0$ diszkretizációs lépcsőt, ezután alternáljuk a születés és halál lépéseket.

1. *Születés:* sorban vizsgáljuk meg a talaj síkra illesztett rácson lévő s képpontokat, és ha nincs s középpontú objektum a jelenlegi ω konfigurációban, futtassuk a születés folyamatot s -ben δb_0 valószínűséggel.

Születés folyamat s -ben: generáljunk egy új u objektumot $[x(u), y(u)] := s$ középponttal, és állítsuk be a $h(u)$ magasság paramétert véletlenszerűen előre meghatározott maximális és minimális értékek között. Végül, adjuk u -t az aktuális ω konfigurációhoz.

2. *Halál:* tekintsük az aktuális $\omega = \{u_1, \dots, u_n\}$ objektum konfigurációt és rendezzük az objektumokat $J_{\mathcal{D}}(u)$ alapján, csökkenő irányban. Az objektumokat vizsgáljuk meg ebben a sorrendben, és minden egyes u -ra számítsuk ki az alábbi értéket $\Delta\Phi_{\omega}(u) = \Phi_{\mathcal{D}}(\omega/\{u\}) - \Phi_{\mathcal{D}}(\omega)$, amely az u törlésével létrejövő potenciális energiaváltozást határozza meg. A halálozási arány így:

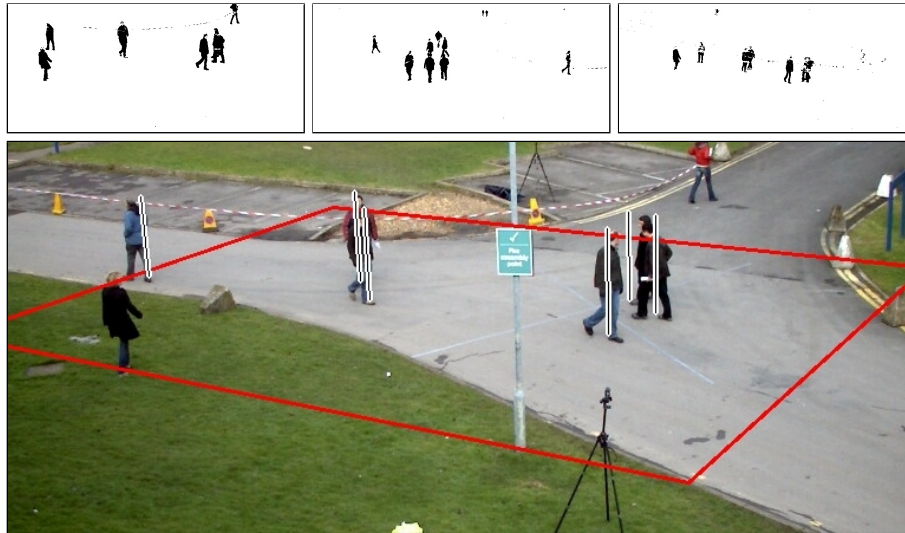
$$d_{\omega}(u) = \frac{\delta a_{\omega}(u)}{1 + \delta a_{\omega}(u)}, \quad \text{ahol } a_{\omega}(u) = e^{-\beta \cdot \Delta\Phi_{\omega}(u)}$$

Ezután $d_{\omega}(u)$ valószínűséggel töröljük u -t ω -ból.

Konvergencia teszt: amíg a folyamat nem konvergál, növeljük a β inverz hőmérsékletet, csökkentünk a diszkretizációs lépcsőt δ geometriai sémával, és lépünk vissza a születés lépésre. A konvergencia tényét akkor állapítjuk meg, ha az egymást követő ciklusok végén már nem változik a konfiguráció.

4. Kísérletek

A javasolt módszer teszteléséhez a PETS 2009 adatbázis [13] *City center* képeit használtuk, mely kameránként összesen 400 képet tartalmazott. Továbbá a rendelkezésre álló kamerák közül a nagy látószögűeket választottuk ki (View_001, View_002, valamint View_003). Kísérleteinkben a vetítéseket a talaj síkján egy kézzel kijelölt téglalap alakú területére korlátoztuk, amely mindegyik nézetből jól látható volt. A KGM háttérmodellt a CIE $L^*U^*V^*$ színtérben definiáltuk, majd a paraméterbecslő folyamat után kézzel megnöveltük a csatornák kovarianciáját, hogy minimum értékük 25.0 (chroma csatornák), illetve 49.0 (luma csatorna) legyen, ezzel csökkentve az árnyékok hatását. Végül az előtér-háttér elkülönítéshez a széles körben használt [12] módszert használtuk az alábbi paraméter beállításokkal: $T = 0.6$ modalitás paraméter, $I = 3.0$ illeszkedési feltétel.



7. ábra. Felül: az előtér-háttér elkülönítés eredménye. Alul: a személyek becsült talaj pozícióit és magasságait egy vonal ábrázolja. A megfigyelt területet piros téglalap jelöli.

A jellemző kinyerés során (3.2 fejezet) $r = 25\text{cm}$ sugarat feltételeztünk, Δ értékét konstans 30° -ra állítottuk, a büntető paraméter $\alpha = 1.0$ volt, a terület arány küszöb pedig $\tilde{f} = 0.75$. Az optimalizálási folyamat paramétereinek beállításához azt feltételeztük, hogy legalább egy nézet megfelelően tartalmazza egy személy lábát, egy másik pedig a fejét, amiből $d_0 = 1/3$ objektum jelölt elfogadási küszöb következik. Azonban az előtér maszkok zajossága miatt kísérleteinkben egy kevésbé korlátozó $d_0 = 0.28$ értéket használtunk. D értéke konstans 8, a személyek közötti minimális távolság feltétel (azaz az 5(a) ábrán látható henger sugara) pedig 50cm . Ami a többszörös szünetés és halál optimalizálási folyamat paramétereit illeti, követtük az eredeti cikk [9] útmutatásait, és $\delta_0 = 20000$, $\beta_0 = 50$ értékeket használtunk, a geometriai hűtési tényező pedig $1/0.96$ volt. Minden képkocka esetén az optimalizálási folyamatot 20 iterációra korlátoztuk, a kapott eredményeket nem használtuk fel a következő képkocka elemzése során. Az eredmények megjelenítéséhez visszavetítettük a becsült pozíciókat az első kamera neztre, majd egy vonalat húztunk a becsült talaj és fej pozíció között (7. alsó kép), a megfigyelt terület határát egy piros téglalap jelöli. A 7. felső képei az előtér-háttér elkülönítés eredményét mutatja. Végül vizuálisan értékeltük ki az eredmény pontatlansági arányát (azaz a helyes detektálás, de alul vagy túlbecsült magasság érték, ez 6.27% lett), valamint szintén kiszámítottuk a hamis negatív detektálás arányát is (1.75% értéket kaptunk). További kísérleti eredmények a <http://web.eee.sztaki.hu/~ucu/vs10-location-results.avi> címen érhetőek el.

5. Összefoglalás

A jelen cikkben egy új módszert mutattunk be a személyek lokalizálásához, több kalibrált kamera használatával. A feladathoz először a 2D képleképzés fizikai tulajdonságain alapuló jellemzőt nyertünk ki, ami magas választ (bizonyítékot) ad egy személy valós pozíciója és valós magassága esetén. Eszközünknek robusztusnak kell lennie zsúfolt helyszíneken, ahol magas a takarások aránya. Ezért a módszerünk mindegyik kamera, több síkra vetítés által kapott, bizonyítékát egyesíti. Végül a pozíciók és magasságok becslése egy optimalizálási folyamattal, a Többszörös Születés és Halál dinamikával, történik. A jelenlegi implementációban előtér-háttér elkülönítést [12] végzünk az előtér képpontok kinyeréséhez. A kiértékeléshez egy publikus kültéri adatbázis három kamera nézetének képeit használtuk. Vizsgálataink szerint a javasolt módszer zsúfolt környezetben is pontos becslést biztosít, ahol teljes vagy részleges takarások gyakran előfordulnak. A jövőben szeretnénk megvizsgálni a jellemző kinyerés és az optimalizálás különböző paramétereinek hatásait. Ezenkívül szeretnénk megvizsgálni az aktuális eredmények felhasználásának előnyeit a következő lépésben. Továbbá jelentős javulást hozhat a bizonyítékok létrehozásában a robusztus testrészt detektáló eljárások (pl. [14]) használata. Ez könnyedén, minimális változtatással beilleszthető a javasolt algoritmusba.

Köszönetnyilvánítás

Köszönet az EU THIS projekt által nyújtott támogatásért. A második szerző munkáját a Bolyai János Kutatási Ösztöndíj is támogatta.

Hivatkozások

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. *ACM Computing Surveys* **38** (2006) 13
2. Mikic, I., Santini, S., Jain, R.: Video processing and integration from multiple cameras. In: *Proc. of the Image Understanding Workshop*. (1998) 183–187
3. Mittal, A., Davis, L.S.: M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo. *Int. J. of Computer Vision* **51** (2002) 189–203
4. Kang, J., Cohen, I., Medioni, G.: Tracking people in crowded scenes across multiple cameras. In: *Proc. of the Asian Conf. on Computer Vision*. (2004)
5. Fleuret, F., Berclaz, J., Lengagne, R., Fua, P.: Multicamera people tracking with a probabilistic occupancy map. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **30** (2008) 267–282
6. Khan, S.M., Shah, M.: Tracking multiple occluding people by localizing on multiple scene planes. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **31** (2009) 505–519
7. Havasi, L., Szilávik, Z.: Using location and motion statistics for the localization of moving objects in multiple camera surveillance videos. In: *Proc. of the IEEE Int. Workshop on Visual Surveillance*. (2009)
8. Benedek, Cs., Szirányi, T.: Bayesian foreground and shadow detection in uncertain frame rate surveillance videos. *IEEE Trans. on Image Processing* **17** (2008) 608–621
9. Descombes, X., Minlos, R., Zhizhina, E.: Object extraction using a stochastic birth-and-death dynamics in continuum. *J. of Math. Imaging and Vision* **33** (2009) 347–359

10. Tsai, R.Y.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. of Robotics and Automation* **3** (1987) 323–344
11. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. of Royal Statistical Society, Series B* **39** (1977) 1–38
12. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **22** (2000) 747–757
13. PETS: Dataset - Performance Evaluation of Tracking and Surveillance (2009) <http://www.cvg.rdg.ac.uk/PETS2009/a.html>.
14. Wu, B., Nevatia, R.: Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses. *Int. J. of Computer Vision* **82** (2009) 185–204