

# **Képek illesztése: kamera regisztráció és képi lényegkiemelés CNN architektúrával**

Ph. D. értekezés tézisei

Szlávik Zoltán

Tudományos vezető:

Dr. Roska Tamás

Témavezető:

Dr. Szirányi Tamás

Analogikai és Neurális Számítások Laboratórium  
Számítástechnikai és Automatizálási Kutatóintézet  
Magyar Tudományos Akadémia

Budapest, 2006



# Tartalom

<i>Tartalom</i> .....	3
<i>Bevezetés</i> .....	4
<i>Alkalmazott vizsgálati módszerek</i> .....	7
<i>A disszertáció tézisei</i> .....	8
<i>Lehetséges alkalmazási területek</i> .....	14
<i>A szerző publikációi</i> .....	15
<i>Hivatkozások</i> .....	17

# Bevezetés

Napjainkban az élet legkülönbözőbb területein merül fel az igény a számítógépes rendszerek használata iránt az emberi vagy egyéb tevékenységek megfigyelésére, lehetőleg automatizált és előzetes információt nem igénylő eljárásokkal. Például szolgálhat hasonló rendszerekre a közlekedés megfigyelése és irányítása, az utazók megfigyelése a pályaudvarokon vagy autósok megfigyelése a parkolóknál. Egyetlen kamera használata korlátozza a megfigyelő rendszer alkalmazhatóságát, ugyanis az egykamerás megfigyelés leggyakoribb problémái a kitakarások és a kamera korlátozott látómezeje. Ezen okoknál fogva képtelenség a teljes megfigyelendő területet „szemmel” tartani. Még a legegyszerűbb feladatok esetén is elengedhetetlen a többkamerás rendszerek használata. Tipikus helyszínei a sokkamerás megfigyelésnek a repterek, a pályaudvarok, a bankok. Vagyis az olyan helyszínek, ahol a környezet nagysága vagy bonyolultsága miatt egyetlen kamerával a megfigyelés nehéz feladatnak bizonyul. Természetes igény az ilyen rendszerekkel/algorithmusokkal szemben a gyors és biztonságos működés. A kép- és videófeldolgozás területén a mai napig alapproblémának számít a gyors és/vagy robusztus algoritmusok implementálása, kidolgozása. Egy tipikus sokkamerás rendszer által végrehajtandó feladatok a következők:

1. változásdetektáció;
2. objektum detektálás;
3. objektum meghatározása – helyzet, osztály, jellemzők;
4. objektum követése;
5. eseménydetektáció.

A munkám során a fenti problémák egy részére kerestem választ. Egyfelől gyors algoritmusokat dolgoztam ki arcképek analízisére, másfelől robusztus eljárásokat terveztem kamerák képeinek illesztéséhez.

Az objektumok detektálása és követése egy jól definiált feladat, ha egy képről vagy egy képsorozatról van szó. Ebben az esetben a követés a detektált objektum megtalálását jelenti a későbbi képkockákon. Sokkamerás esetben a feladat hasonlóképpen értelmezhető, a detektált objektumot meg kell találnunk a többi

kamera képén. Egy lehetséges megoldás erre, ha a kamerák képeit illeszteni tudjuk. Az illesztés megoldásával az objektumok követése is robusztusabbá tehető a különböző nézetekből kapott információk összevetésével.

Az elmúlt évek eredményei sok olyan példát szolgáltatnak, amelyek esetében a kamerák képeinek illesztése sikeresen elvégezhető. Az ismert eljárások két nagy csoportba oszthatók: állókép- és mozgás-alapú eljárásokra. Ezen megoldások közös jellemzője, hogy valamilyen előfeltételezéssel élnek a látvány vagy objektum struktúráját vagy dinamikáját/mozgását illetően. Az állókép-alapú eljárások ([1][2][3][4]) esetében az olyan képi jellemzőket próbálják összeilleszteni, mint: élek, sarkok, kontúrok, színek, formák stb. Ezek az eljárások csak kis bázistávolságú képpárok esetében használhatók, amikor a különböző nézetekből adódó különbségek jelentéktelenek. A mozgás-alapú eljárások a kamerák által megfigyelt helyszín dinamikáját analizálják, amely még nagyon különböző nézetek esetén sem változik. A [5][6] munkákban a mozgó objektumokat próbálják detektálni és követni, valamint a nyomvonalak illesztésével regisztrálják a kamerák képeit. Ezen mozgásalapú eljárások gyakorlati alkalmazása sokkamerás rendszerekben csak korlátozottan lehetséges.

Mit csinájunk akkor, ha semmilyen *a priori* információval nem rendelkezünk a látványról vagy az általunk megfigyelt objektumokról? Továbbá, nem tudjuk őket sem detektálni, sem követni? Előzetes feltételezések esetén ilyenkor könnyen kaphatunk pontatlan, rossz megoldást. A munkám során olyan eljárást dolgoztam ki, mely mindenféle *a priori* információ nélkül képes összeegyeztetni a kamerák képeit.

Mikor több algoritmus – mozgásdetekció, követés, objektumanalízis stb. – integráltan van egy rendszerben, akkor elengedhetetlen a rendkívül gyors algoritmusok tervezése. Emberek azonosításakor videofelvételek és képek alapján az a legegyszerűbb és legtermészetesebb, ha az emberi arc analízise alapján döntünk. Mivel az emberi arc un. „nem-intruzív” biometrikus jellemző, vagyis az analízise elvégezhető anélkül, hogy a megfigyelni kívánt személyt külön eljárásnak vetnénk alá.

Számos, az emberi arc analízisével foglalkozó tudományos munka látott napvilágot az elmúlt években. Az arcdetekciós és arcfelismerő eljárások [9][10] a következő lépéseket valósítják meg: i) arc-gyanús képrészletek (továbbá arcjelöltek) detektálása (pl.: színszűrés segítségével); ii) az arcjelöltek megerősítése; iii) azonosítás. A megerősítés során az arc-jelöltekről kiderül, hogy valóban arcok-e vagy sem. Egy lehetséges módszer az arc-jelöltek verifikálására az arcjelölt és egy átlagos arcot vagy egy arcmodellt tartalmazó kép korrelációjának a kiszámolása [8]. Egy másik lehetséges megoldás, amikor az arc lényegi részei (szemek, száj és az orr) közötti távolságok egymáshoz való arányainak kiszámolásával kapnak olyan mérőszámokat, amikből megállapítható, hogy az arc-jelölt valóban arc-e vagy sem [7]. Az ilyen típusú megoldásnak alapfeltétele az arc lényegi részeinek a észlelése. Munkám során valós idejű eljárások kidolgozását tűztem ki célul az emberi arc jellemzőinek detektálására, ami alapján elvégezhetővé válik a későbbi, magasabb rendű döntés.

# Alkalmazott vizsgálati módszerek

A kutatások és kísérletek során felhasználtam az érintett szakterületek – képfeldolgozás és CNN elmélet – eddig elért eredményeit, amik számos konferencián és rangos folyóiratokban jelentek meg (konferenciák: pl. ICIP, ICPR, CNNA, ECCTD; folyóiratok: pl. PAMI, TCAS I.-II.).

A képi lényegkiemelő analogikai algoritmusok tervezése során felhasználtam az elődeim által kidolgozott analogikai algoritmusokkal kapcsolatos tapasztalatokat és eredményeket. Ezekre támaszkodva a feladat részfeladatainak megoldásával dolgoztam ki az eljárást. Az eljárás paramétereinek beállításához felhasználtam a laboratóriumunkban kidolgozott Aladdin számítógépes szimulációs rendszert valamint az ACE 16K celluláris mikroprocesszort a hozzátartozó szoftverrel együtt. A képi lényegkiemelő algoritmusok teszteléséhez egy nemzetközileg elfogadott képi adatbázist használtam [11].

Az automatikus képillesztő algoritmus kidolgozásához felhasználtam a laboratóriumunkban fejlesztett MDICam sokkamerás keretrendszert. Az algoritmusokat különböző kül- és beltéri videókon teszteltem. A kísérletek elvégzéséhez szimulátorokat fejlesztettem.

# A disszertáció tézisei

## I. Tézis

### Kamerák regisztrációja akkumulált mozgásstatisztikák segítségével

- a. **Az együttmozgási statisztikákra alapozva új eljárást adtam kamerák regisztrációjára.**

Együttmozgási statisztikákat definiáltam a különböző kameraképeken megfigyelt mozgások akkumulációjaként. Kísérletileg igazoltam, hogy az együttmozgási statisztikák hatékonyan használhatók kamerák regisztrációjára emberi beavatkozás nélkül, mindenféle *a priori* környezeti információt figyelmen kívül hagyva. Az eljárás pontossága a *subpixel*-es tartományon belül van.

*Kapcsolódó publikációk:* [2][6][16].

- b. **Megmutattam, hogy az együttmozgási statisztikákon alapuló kameraregisztáció esetén a videók időben egymáshoz illeszthetők a regisztráció hibáját minimalizálva.**

Eljárást adtam a videók időbeni illesztésére szinkronizálatlan kamerák esetére. A kidolgozott eljárás a kamerák regisztrációjakor keletkező hibát méri, és ezt minimalizálva határozza meg az időbeni eltérést a videók között.

*Kapcsolódó publikációk:* [2][6]

## II. Tézis

### Automatikus Bayes-i eljárás kamerák regisztrációjára tetszőleges struktúrájú objektumok, látványok esetére.

- a. **Új Bayes-i eljárást dolgoztam kamerák közös területeinek automatikus meghatározására.**

Sokkamerás rendszerek által szolgáltatott képek illesztésének előfeltétele a kameraképek közös területeinek ismerete/észlelése.



Bayes-i modellt dolgoztam ki az együtváltozó pontok vizsgálatára. Megmutattam, hogy a modell megoldása ekvivalens egy periodikus Markov lánc határvalószínűségeinek meghatározásával. A kidolgozott eljárás alkalmas a kameraképek közös részeinek automatikus detekciójára mellőzve mindenféle emberi beavatkozást vagy *a priori* információt a megfigyelt objektumokról.



1. ábra: egyazon helyszínről készített képek egymásra illesztése.

A szakirodalomból ismert képillesztő eljárások azt feltételezik, hogy a megfigyelt objektumok elég messze vannak a kamerától ahhoz, hogy síkon mozgó pontoknak tekinthessük őket. Ezáltal a feladat két sík összeillesztésére egyszerűsödik. Ez a feltételezés viszont csak kültéri rendszerek esetében érvényes. Beltéren a mozgó objektumok viszonylag közel vannak a kamerákhoz és a detektált mozgások a 3 dimenziós térben figyelhetők meg, azaz nem egy síkon. Viszont, ha olyan pontokat észlelünk melyek nagy valószínűséggel egy síkon (pl.: a padló síkja) mozognak, akkor továbbra is élhetünk a korábbi feltételezésünkkel. A következő altézis a fenti probléma megoldásával foglalkozik.

**b. Megmutattam, hogy a mozgástörténetek entrópiájának vizsgálatával hatékony eljárás készíthető a jellemző mozgásokhoz tartozó képpontok kinyerésére.**

A mozgástörténetek entrópiáját vizsgáló eljárást dolgoztam ki a megfigyelt mozgásokhoz tartozó képpontok klasszifikációjára: (i) jelentős, nem folyamatos mozgások; (ii) folyamatos mozgások vagy kamerazaj.

**c. Kísérletileg megmutattam, hogy egy árnyék-detekciós algoritmust az együttmozgások detekciójával kombinálva hatékony eljárás készíthető beltéri kamerák regisztrációjára.**

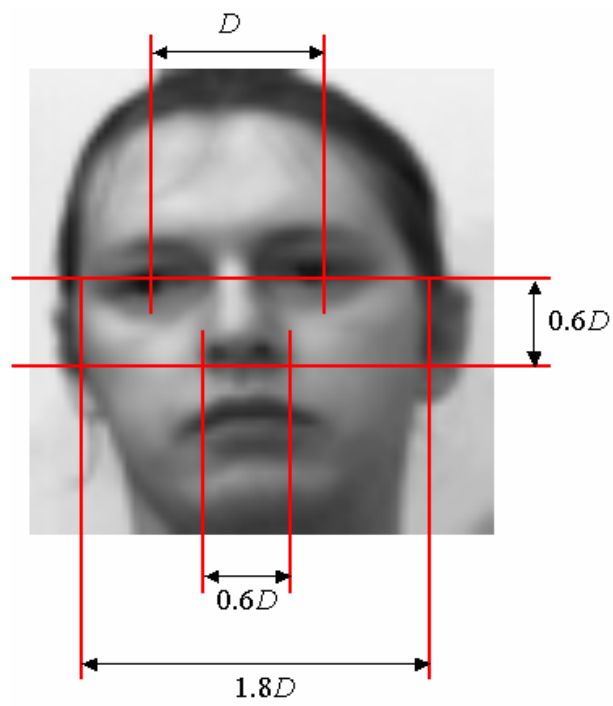
Az árnyékok ideális jellemzők beltéri kameraképek illesztéséhez; a legtöbb esetben a földön vagy padlón vannak, vagyis egy jól behatárolható síkon. Megmutattam, hogy a mozgó objektumok árnyékának detekciójával hatékony eljárás készíthető beltéri kamerák regisztrációjára.

### **III. Tézis**

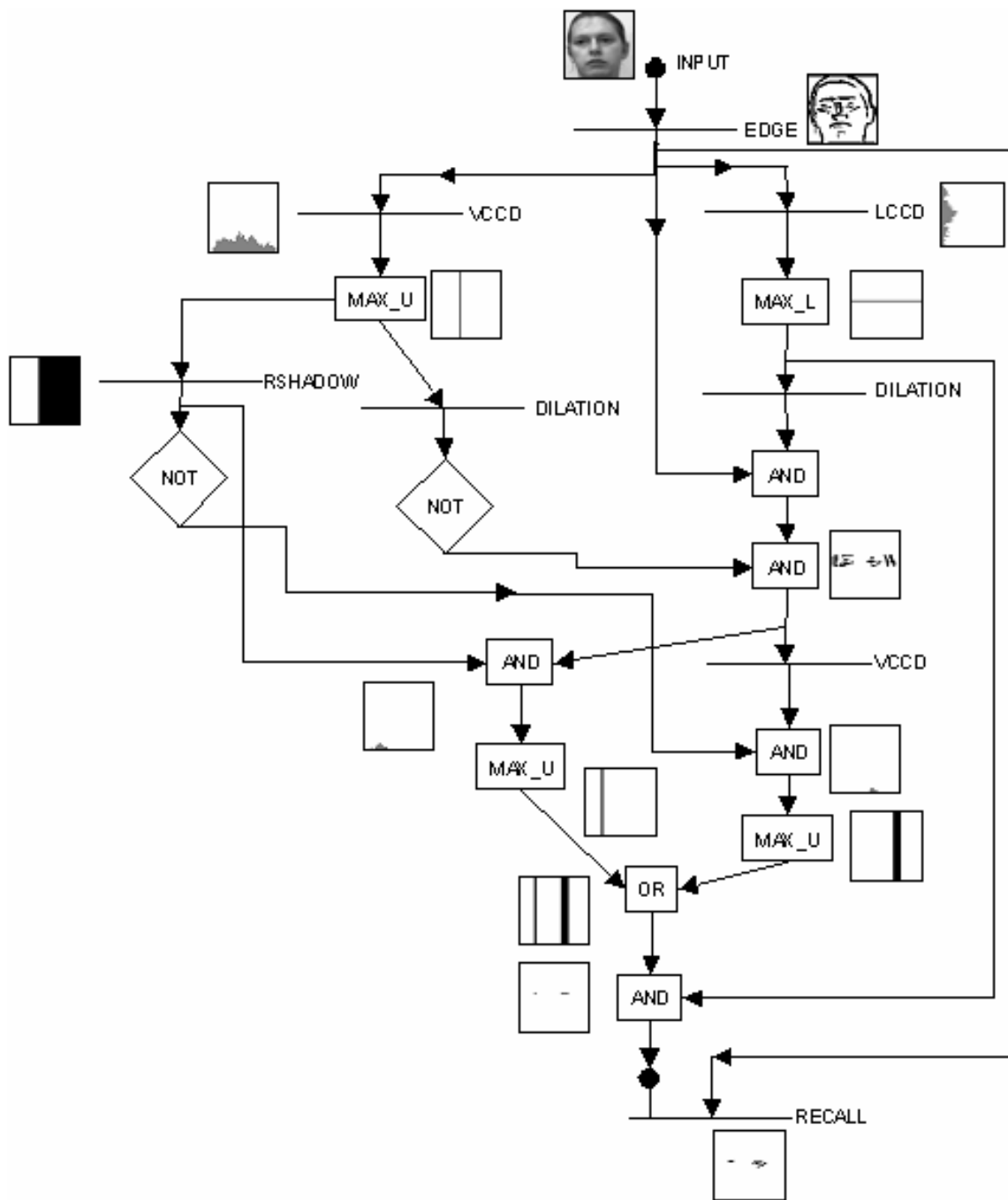
**Robusztus, tér-időbeli dinamika használatán alapuló képi lényegkiemelő analogikai algoritmusok, [3] [5]**

**a. Morfológiai műveletek tér-időbeli megvalósításán alapuló modellbázisú képi lényegkiemelő algoritmusokat dolgoztam ki az arc lényeges részeinek detektálására.**

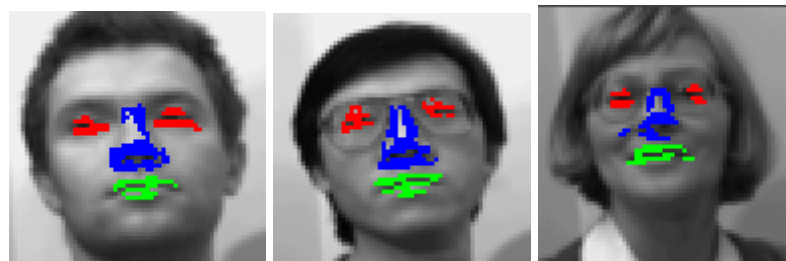
Új analogikai algoritmusokat dolgoztam ki az emberi arc lényeges részeinek – szemek, orr, száj – kiemelésére. Az algoritmusok azon a felismerésen alapulnak, hogy bizonyos geometriai összefüggések vannak az arc lényeges részei között [7]. Ezek az egyszerű összefüggések valamint a párhuzamos CNN architektúrában alkalmazott morfológiai műveletek biztosítják az algoritmus robusztus működését és valós idejű (50 képkocka/másodperc) futását a ACE4K vizuális mikroprocesszoron.



2. ábra: Az emberi arc alkotóelemei közötti geometriai összefüggés.



3. ábra: Szemeket detektáló algoritmus UMF-diagramja



4. ábra: Az emberi arc különböző alkotóelemeinek detektálása

**b. Frontális arcképek pozícióinvariáns azonosítása CNN architektúrán.**

Megmutattam, hogy a 2/a pontban vázolt algoritmus eredményeit felhasználva pozícióinvariáns arcfelismerő algoritmus készíthető. Az algoritmus lényege, hogy az arc lényegi részeinek megkeresésével és a közöttük lévő geometriai összefüggések felhasználásával megbecsülhető a fej pozíciója. Ismerve a fej pozícióját az arc visszatranszformálható frontális nézetbe és korrelációs módszerrel összevetve az adatbázisban lévő arcokkal elvégezhető az azonosítás.



<b>P4</b>	89%	80%	70%
92%	<b>P3</b>	83%	67%
84%	86%	<b>P1</b>	88%
72%	81%	90%	<b>P2</b>

5. ábra Arcfelismerés CNN architektúrával. Az arcadatbázis képei előzetesen négy csoportba lettek osztva a képeken látható arc pozíciója szerint. A fenti sorban az egyes osztályok mintapéldányai láthatók. A táblázat első sorában a P4 osztály képei lettek összehasonlítva többi osztály képeivel. Minden egyes kép P4-ből hozzá lett transzformálva a többi osztály képéhez és az algoritmus megpróbálta megkeresni a párját. Az azonosítás eredményei vannak feltüntetve a második, harmadik és negyedik oszlopban. A táblázat többi sorának értelmezése hasonló az elsőhöz.

# Lehetséges alkalmazási területek

A dolgozatban bemutatott algoritmusok konkrét alkalmazásokhoz kötődnek, melyek különböző projektekben merültek fel.

Az 1. tézisben megfogalmazott automatikus képillesztő eljárások általános megoldást biztosítanak szélesbázistávolságú sztereó-képek illesztésére tetszőleges körülmények esetén. Az algoritmusok a TeleSense, a MUSCLE Network of Excellence valamint „Video események szemantikus értelmezése, kategorizálása és indexálása” GVOP projektek keretében voltak megvalósítva és valós körülmények között tesztelve.

A 2. tézisben kidolgozott lényegkiemelő algoritmus elsősorban olyan rendszerek részeként alkalmazható, amelyekben emberi arcok analízisére van szükség egy konkrét probléma megoldásához. Az algoritmusnak CNN-UM chipes megvalósítása biztosítja a valós idejű működést (50 képkocka/másodperc) elősegítve ezzel a bonyolult rendszerekbe való hatékony integrálását. Az algoritmusok a TeleSense projekt keretében valósítottam meg.

# A szerző publikációi

## **Folyóirat publikációk:**

- [1] **Zoltán Szlávik**, Tamás Szirányi, “Stochastic view registration of overlapping cameras based on arbitrary motion”, *IEEE Transactions on Image Processing*, **accepted**, 2006
- [2] **Zoltán Szlávik**, Tamás Szirányi, László Havasi, “Video camera registration using accumulated co-motion maps”, *ISPRS Journal of Photogrammetry and Remote Sensing*, **accepted**, 2006
- [3] László Havasi, **Zoltán Szlávik**, Tamás Szirányi, “Detection of gait characteristics for scene registration in video surveillance system”, *IEEE Transactions on Image Processing*, **accepted**, 2006
- [4] László Havasi, **Zoltán Szlávik**, Tamás Szirányi, “Higher order symmetry for non-linear classification of human walk detection”, *Pattern Recognition Letters*, vol. 27(7), pp. 822-829, 2006

## **Nemzetközi konferencia publikációk:**

- [5] László Havasi, **Zoltán Szlávik**, Tamás Szirányi, “Use of human motion biometrics for multiple-view registration”, *Advanced Concepts for Intelligent Vision Systems, ACIVS 2005, LNCS*, vol. 3708, pp. 35-44, 2005
- [6] **Zoltán Szlávik**, László Havasi, Tamás Szirányi, “Estimation of common groundplane based on co-motion statistics”, *International Conference on Image Analysis and Recognition, ICIAR’04, LNCS*, vol. 3211, pp. 347-353, 2004
- [7] **Zoltán Szlávik**, Tamás Szirányi, László Havasi, “Analysis of dynamic scenes by using co-motion statistics”, *IEEE International Workshop on Visual Surveillance*, 2006
- [8] **Zoltán Szlávik**, Tamás Szirányi, “Bayesian estimation of common areas in multi-camera systems”, *IEEE International Conference Image Processing, ICIP, Atlanta*, **accepted**, 2006
- [9] Cs. Benedek, L. Havasi, T. Szirányi, **Z. Szlávik**, “Motion-based Flexible Camera Registration”, in *Proc. of IEEE Advanced Video and Signal-Based Surveillance, AVSS’05*, pp. 439-444, 2005
- [10] **Zoltán Szlávik**, László Havasi, Tamás Szirányi, Csaba Benedek, “Random motion for camera calibration”, *European Signal Processing Conference, EUSIPCO, Antalya*, 2005
- [11] **Zoltán Szlávik**, Tamás Szirányi, László Havasi, Csaba Benedek, “Optimizing of searching co-motion point-pairs for statistical camera calibration”, *IEEE International Conference on Image Processing, ICIP*, pp. 1178-1181, Genova, 2005

- [12] László Havasi, **Zoltán Szlávik**, Tamás Szirányi, “Eigenwalks: walk detection and biometrics from symmetry patterns”, IEEE International Conference on Image Processing, ICIP, pp. 289-292, Genova, 2005
- [13] László Havasi, **Zoltán Szlávik**, Csaba Benedek, Tamás Szirányi, “Learning human motion patterns from symmetries”, International Conference on Machine Learning Workshop on Machine Learning for Multimedia, Bonn, 2005, (on CD-ROM)
- [14] **Zoltán Szlávik**, Tamás Szirányi, “Face analysis using CNN-UM”, Proc. of Cellular Neural Networks and Applications, CNNA’04, pp. 190-196, Budapest, 2004
- [15] László Havasi, **Zoltán Szlávik**, “Symmetry feature extraction and understanding”, Proc. of Cellular Neural Networks and Applications, CNNA’04, pp. 255-261, Budapest, 2004
- [16] **Zoltán Szlávik**, László Havasi, Tamás Szirányi, “Image matching based on co-motion statistics”, 2<sup>nd</sup> IEEE Int. Symposium on 3D Data Processing, Visualization, and Transmission, 3DPVT, 2004, (on CD-ROM)
- [17] László Havasi, Csaba Benedek, **Zoltán Szlávik**, Tamás Szirányi, “Extracting structural fragments of overlapping pedestrians”, Proc. Of the 4<sup>th</sup> IASTED Int. Conference VIIP’04, pp. 943-948, Marbella, 2004
- [18] László Havasi, **Zoltán Szlávik**, Tamás Szirányi, “Pedestrian Detection using derived Third-Order Symmetry of Legs”, ICCVG’04
- [19] **Z. Szlávik**, T. Szirányi, “Face identification using CNN-UM”, Proceedings of European Conference on Circuit Theory and Design, ECCTD’03, vol. 2, pp. 81-85, Krakow, 2003
- [20] **Z. Szlávik**, D. Bálya, T. Roska, “Properties of the adaptive integration formula to compute the CNN dynamic equation”, Proceedings of European Conference on Circuit Theory and Design, ECCTD’03, vol. 1, pp. 105-109, Krakow, 2003



# Hivatkozások

- [21] S. T. Barnard, W. B. Thompson, "Disparity analysis of images," IEEE Trans. PAMI, vol. 2, pp. 333-340, 1980.
- [22] J. K. Cheng, T. S. Huang, "Image registration by matching relational structures," Pattern Recog., vol. 17, pp. 149-159, 1984.
- [23] J. Weng, N. Ahuja, T. S. Huang, "Matching two perspective views," IEEE Trans. PAMI, vol. 14, pp. 806-825, 1992.
- [24] Z. Zhang, R. Deriche, O. Faugeras, Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," Artificial Intelligence Journal, vol. 78, pp. 87-119, 1995.
- [25] L. Lee, R. Romano, G. Stein, "Monitoring activities from multiple video streams: establishing a common coordinate frame," IEEE Trans. PAMI, vol. 22, 2000.
- [26] Y. Caspi, D. Simakov, and M. Irani, "Feature-based sequence-to-sequence matching," in Proc. VAMODS (Vision and Modelling of Dynamic Scenes) workshop, with ECCV'02, Copenhagen, 2002.
- [27] Brunelli R, Poggio T. Face recognition through geometrical features. Proceedings of ECCV, S. Margherita Ligure, 1992; 792-800.
- [28] Brunelli R, Poggio T. Face recognition: features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 1993;15(10):1042-1052.
- [29] Chellappa R, Wilson C L, Sirohey S. Human and Machine Recognition of Faces: A Survey. Proceedings of IEEE 1995; 83:705-740.
- [30] Hjelmas E, Low B K. Face detection: A survey. Computer Vision and Image Understanding 2001;83(3):236-274.
- [31] Graham D B, Allinson N M. Characterizing Virtual Eigensignatures for General Purpose Face Recognition. Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences 1998;163:446-456.