

**Mozgásalapú háromdimenziós  
objektum-rekonstrukció  
gyenge perspektíva esetén**

**Hajder Levente**

PhD értekezés tézisei

Témavezetők:

dr. Csetverikov Dmitrij (MTA SZTAKI)

dr. Vajk István (BME AAIT)

Számítástechnikai és Automatizálási Kutató Intézet

Magyar Tudományos Akadémia

Automatizálási és Alkalmazott Informatikai Tanszék

Budapesti Műszaki és Gazdaságtudományi Egyetem

Budapest 2007.

## 1. Előzmények és célkitűzések

A mozgásalapú objektum-rekonstrukció a háromdimenziós számítógépes látás egyik érdekes, kihívásokkal teli kutatási területe. A tudományág az elmúlt húsz évben nagy fejlődésen ment keresztül, ám ez nem azt eredményezte, hogy sikerült minden feladatra megoldást találni, hanem éppen az ellenkezőjét: a publikált algoritmusok alkalmazásával újabb és újabb megoldandó problémák bukkantak elő.

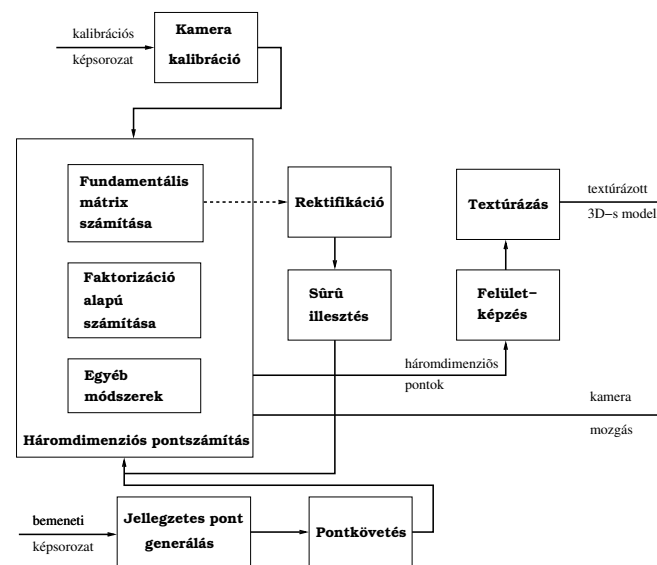
Az objektum-rekonstrukció eredményeit már most is számtalan helyen lehet alkalmazni, pedig a kidolgozott módszerek sebessége, pontossága és automatizáltsága sok esetben még nem teszi lehetővé az elméletileg jól működő eljárások felhasználását a gyakorlatban.

A lehetséges alkalmazási területek közül kiemelhetjük a robotlátást, hiszen a rekonstrukció során a harmadik dimenzió megjelenítése a felismerésekhez fontos többletinformációt juttat a kétdimenziós látórendszerekhez képest. Másik fontos terület a szórakoztatóipar, elsősorban a filmtrükkök világa: az ezredforduló óta számtalan világsiker befutott film alkalmazza a tudományterület eredményeit. Ezen a két területen kívül is sikerrel alkalmazzák a tudományos eredményeket, a lopásvédelemtől a közlekedésrendészeti feladatokig lehetne sorolni a felhasználási lehetőségeket.

A rekonstrukciós feladat megértéséhez, és a javasolt megoldások alkalmazásához több területet is ismernie kell a kutatóknak: a matematikai ismereteken túl a videófelvételekből származó képek és a 3D→2D leképzés geometriájának megértéséhez optikai tudással kell rendelkezni, a képek elemzéséhez a digitális képfeldolgozás módszereinek

magas szintű ismerete szükséges, a valóság-hű rekonstrukcióhoz és a filmtrükkök kidolgozásához elengedhetetlen a számítógépes grafika alapjainak elsajátítása.

Kutatási munkám során a fő célkitűzésem az volt, hogy egy moduláris felépítésű rekonstrukciós rendszert dolgozzak ki kollégák és egyetemi hallgatók segítségével. Ezt a célt sikerült elérni. A rendszerünk működését az alábbi ábrával szemléltethetjük:



A modularitásnak köszönhetően a rendszer elemeit folyamatosan lehet fejleszteni, a komponenseket jobbra cserélni. Tudományos eredményeim is (az utolsó tézist leszámítva) ebbe a rendszerbe beépíthető új módszereket jelentenek.

## 2. Módszertani összefoglalás

### A faktorizációs feladat

A faktorizáció feladata a kétdimenzióban követett pontok alapján meghatározni az objektum háromdimenziós pontjainak a koordinátáit, továbbá a kamera mozgását az objektumhoz képest. Gyenge perspektív vetítést használva ez a következőképpen írható le:

Ha adott valamely merev mozgó objektum  $P$  darab (jellegzetes) pontja, amelyeket  $F$  számú képkockán keresztül követtünk, és a  $p$ -edik követett pont koordinátáit az  $f$ -edik képen  $x_{fp} = (u_{fp}, v_{fp})^T$  jelöli, akkor merőleges vetítés esetén a pont helye a következő módon számítható:

$$x_{fp} = R_f s_p + t_f, \quad (1)$$

ahol  $R_f = [r_{f1}, r_{f2}]^T$  ortonormált mátrix első két sora,  $s_p$  a jellegzetes pont háromdimenziós koordinátája,  $t_f$  pedig az objektumkoordinátarendszerének vetületének helye az origóhoz képest. Gyenge perspektíva esetén az összefüggés kiegészül a gyenge perspektíva  $q_f$  (nem nulla) skalárszorójával:

$$x_{fp} = q_f R_f s_p + t_f, \quad (2)$$

A  $t_f$  eltolás kiküszöbölhető, ha az objektum saját koordinátarendszerében a középpontját úgy választjuk meg, hogy ezt az origó vetületét minden képen ismerjük. Praktikus okokból a pontok átlagát, azaz a súlypontot szokták origónak választani, hiszen merőleges vetítés esetén kétdimenzióban és háromdimenzióban is a koordináták egyszerű átlagolásával meghatározható a súlypont.

Ebben az esetben felírhatjuk, hogy

$$x_{fp} = q_f R_f s_p. \quad (3)$$

Ha az összes pontot figyelembe vesszük az adott képen, a fenti egyenletből mátrixegyenletet készíthetünk:

$$\underbrace{W_f}_{2 \times P} = (x_{f1} \dots x_{fP}) = \underbrace{M_f}_{2 \times 3} \cdot \underbrace{S}_{3 \times P} \quad (4)$$

ahol  $M_f$  neve *mozgásmátrix*,  $S = (s_1, \dots, s_P)$  pedig a *struktúramátrix*. Merőleges vetítés esetén  $M_f = R_f$ , míg gyenge perspektíva alatt  $M_f = q_f R_f$ .

Ha vesszük az összes képkockát, a 4. összefüggés alapján a mérési mátrixot is kifejezhetjük:

$$\underbrace{W}_{2F \times P} = \underbrace{M}_{2F \times 3} \cdot \underbrace{S}_{3 \times P}, \quad (5)$$

ahol  $W^T = [W_1^T, W_2^T, \dots, W_F^T]$  és  $M^T = [M_1^T, M_2^T, \dots, M_F^T]$ .

A tényleges feladat a  $W$  mérési mátrix szorzattá bontása (faktorizálása), és az  $M$  mozgási,  $S$  struktúramátrix becslése.

Ha figyelembe vesszük a súlypontok helyét is, akkor a faktorizációt a következő alakba írhatjuk:

$$W = \left[ M \mid t \right] \begin{bmatrix} S \\ \mathbf{1}^T \end{bmatrix}. \quad (6)$$

ahol  $t = [t_1^T t_2^T \dots t_F^T]^T$  vektor tartalmazza a súlypontok helyeit a képeken.

## A Tomasi-Kanade faktorizáció

Tomasi-Kanade faktorizációs módszer két lépésben végzi el a faktorizációs feladatot: először egy rangcsökkentéssel zajszűrést hajt végre, így affin mozgás- és struktúramátrixokkal faktorizálhatunk:  $W = M_{aff}S_{aff}$ . Második lépésben a kapott altérben elvégzünk egy  $Q$  lineáris transzformációt, amely végén megkapjuk a kívánt (valós) mátrixokkal végzett faktorizációt:

$$W = MS \quad (7)$$

ahol  $M = M_{aff}Q$  és  $S = Q^{-1}S_{aff}$ .

Megmutatható, hogy sok pont esetén legalább három képkocka, sok képkocka esetén legalább négy pont szükséges ahhoz, hogy a gyenge perspektív faktorizációt el lehessen végezni.

### Tesztelési környezet szintetikus adatokon

A téziseimet egységes tesztelési környezetben vizsgáltam meg. A tesztadatok generálását az alábbi algoritmus segítségével végeztem el:

1. Mesterséges háromdimenziós objektumokat készítettem Gauss véletlenszám-generátor segítségével. A várható érték nulla volt, a szórás egy paraméter segítségével állítható be.
2. Megadott mozgás segítségével különböző szögekkel elforgattam az objektumot.
3. Gyenge perspektíva segítségével levetítettem a képsíkra.
4. Kétdimenziós normális zajt (várható értéke nulla, a szórás

beállítható paraméter) adtam a vetített koordinátákhoz.

Az így kapott koordinátákra futtattam le a tézisekben ismertetett algoritmusokat. Az eredményeket könnyen ki lehet értékelni, hiszen mind a kameramozgás, mind az objektum eredeti háromdimenziós koordinátái ismertek.

### Tesztelési környezet valós adatokon

Valós képsorozatokon KLT sarokdetektáló segítségével jellegzetes pontokat határoztam meg, ezeket a pontokat korreláció alapú mintaillesztés segítségével követtem a képeken keresztül. Tipikusan 5...10 képkockán át van remény arra, hogy a pontok nagy részét követni tudjuk.

Az így kapott, követett kétdimenziós pontokat adtam az algoritmusok bemeneteire. Először mindig a második tézisben ismertetett szegmentáló eljárások kapták meg a ponthalmazt, és azok választották ki a mozgó objektumhoz illeszkedő kétdimenziós pontokat.

A szegmentált pontok szolgáltatották a nem robusztus algoritmusok bemenetét.

### Az értekezés felépítése

Az értekezésem alapvetően két részre bontható: az első részben a rekonstrukció problematikájával és a vetítéseket magyarázó geometriával foglalkozom, a második részben pedig bemutatom az új tudományos eredményeket.

A 2.fejezet (közvetlenül a bevezető után) a rekonstrukciós rendszer

felépítésével és az egyes részproblémák megoldási lehetőségeivel foglalkozik, bemutatva az egyes feladatokhoz rendelt szakirodalmat. A vetítési modellekkel foglalkozó 3. fejezet ismerteti a három leggyakrabban alkalmazott vetítési módszert: a merőleges, a gyengén perspektív és a perspektív vetítést.

A második, új tudományos eredményekkel foglalkozó rész a három tézisnek megfelelően három fejezetet takar. Az első tézisben (4. fejezet) a kamera mozgásának és a háromdimenziós objektum térbeli pontjainak a rekonstrukciójával foglalkozom, a második tézisben (5. fejezet) a mozgó objektumok szegmentálásával, harmadik tézisemben (6. fejezet) pedig tagolt objektumok mozgásalapú rekonstrukciójával foglalkozom.

### 3. Új tudományos eredmények

A doktori értekezés új tudományos eredményeit a következő három tézis foglalja össze.

#### 1. tézis A Tomasi-Kanade faktorizáció iteratív alapú javítása [2,3,8,16,17]

Kidolgoztam egy új módszert, amelyik a Tomasi-Kanade faktorizáció javítása.

Az új eljárás iteratív, mely a megadott feltétel esetén a futását véges időben befejezi. Az iteráció minden egyes ciklusa három lépést tartalmaz:

**S-lépés** Az S-lépés a háromdimenziós struktúrát javítja. Mivel a struktúramátrix értéke tetszőleges lehet, a struktúrafinomítás legkisebb négyzetes értelemben optimálisan számítható:

$$S^{(k)} = M^{(k-1)\dagger} W \quad (8)$$

**M-lépés** Az M-lépés a kameraparamétereket (mozgásmátrixot) finomítja. Gyenge perspektíva esetén képkockánként 4 paraméterrel írható le a kameramozgás, amelyet a kiegészítő lépésnek köszönhetően zárt alakban megoldottam háromdimenziós pontfelhő-regisztrációja segítségével. (A megoldást levezetéssel együtt a függelékben ismertetem.)

**Kiegészítő lépés** A kiegészítés célja az M-lépés során alkalmazott zárt alakú megoldás segítése.

Szintetikus adatokon mutattam meg, hogy a javított módszer

jobb eredményt ad az eredeti faktorizációnál. Azt is megmutattam, hogy a javítás elsősorban alacsony pontszám esetén jelentős. A javított módszert valódi képsorozatokon is sikerrel alkalmaztam.

## **2. tézis Mozgó objektumok pontjainak szegmentálása gyenge perspektíva esetén [2,4,6,9,10]**

Megoldásokat kínáltam mozgó objektumok pontjainak szegmentálására, azaz olyan módszereket dolgoztam ki, fejlesztettem tovább, amelyek képesek mozgó objektumok kétdimenzióban követett pontjait különválasztani, a mozgó objektumokhoz nem tartozó (például rosszul követett) pontokat pedig elkülöníteni.

### **2.a. Hibás pontok kiszűrése Monte-Carlo módszerrel**

Továbbfejlesztettem a Trajković-Kurata módszert, amely a követett pontokból elkülöníti azokat a pontokat, amelyek nem tartoznak a domináns mozgó objektumhoz.

A továbbfejlesztés két területen történt: (1) az LMedS robusztus statisztikát LTS-re cseréltem, (2) az affin mozgásmátrix helyett a valós mozgás felhasználását javasoltam. Az új módszerek pontosabb eredményt adnak, és LTS esetén magasabb hibáspont-arányt engednek meg, mint az eredeti algoritmus.

Kidolgoztam ezentúl egy RANSAC alapú eljárást is, amely a szakmában elterjedt RANdom SAMpling Consensus robusztus algoritmus alkalmazása a konkrét problémára. Affin és valós mátrixokra egyaránt kiterjesztettem a módszert.

A domináns mozgás meghatározásához egy mértéket vezettem be, amelyről bebizonyítottam, hogy korlátos, és ennek következtében küszöböléssel el lehet dönteni, hogy valamely domináns mozgás korrekt-e.

Ezek az algoritmusokat szintetikus teszteken összehasonlítottam, és az eredményekkel igazoltam, hogy a medián alapú módszerek esetén a hibás pontok számának aránya maximum 50% lehet, LTS esetében ezt meghaladhatja. Azt is igazoltam, hogy a valós mátrixokat használó módszerek pontosabb eredményeket szolgáltatnak az affin mozgást használóknál.

Az algoritmusokat valós felvételekből származó tesztadatokon is sikerrel futtattam.

### **2.b. Régió alapú szegmentálás**

Kidolgoztam a 2.a. altézisben megfogalmazott módszerre alapozva egy új, régió alapú szegmentáló módszert, amely a pontok szegmentálását tovább javítja a kompakt és merev mozgó objektumok sajátosságainak figyelembevételével. A régió alapú megközelítés tudomásom szerint újítás a mozgásalapú, háromdimenziós objektum-szegmentálás területén.

Azt is bebizonyítottam, hogy a szegmentálás alapját képező ún. visszavetítési hiba a koordináta-transzformáció többértelműségére érzéketlen.

Az algoritmus valós képekből származó tesztadatokon vizsgáltam meg, és meggyőző eredményeken kaptam: a szegmentálás még akkor is sikerült, ha a mozgó objektum lassú szögsebességgel forgott.

### 3. tézis Tagolt objektumok rekonstrukciója gyenge perspektíva alatt [11,12,13,14,18]

Kidolgoztam két új módszert, melynek segítségével el lehet dönteni két merev, mozgó objektumról, hogy tagolt objektumot alkot-e. Az első módszer pontszerűen kapcsolt, a második tengelyesen rögzített tagolt objektumok kapcsolódásának eldöntésével foglalkozik. Az első módszer esetén a kapcsolódás helyének becslését is megoldottam.

**3.a. Kapcsolódási pont becslése.** Feltételezem, hogy a merev mozgó objektumokat korábban szegmentáltuk és a mozgás- és struktúrainformációkat az alábbiak szerint kiszámítottuk:

$$W_1 = \left[ M_1 \mid t_1 \right] \begin{bmatrix} S_1 \\ 1_1^T \end{bmatrix}, \quad (9)$$

$$W_2 = \left[ M_2 \mid t_2 \right] \begin{bmatrix} S_2 \\ 1_2^T \end{bmatrix}. \quad (10)$$

A kapcsolódási pontra az alábbi költségfüggvény esetében legkisebb négyzetes értelemben optimális megoldás adtam:

$$J = \left\| W_1 - \left[ M_1 \mid t \right] \begin{bmatrix} S'_1 \\ 1_1^T \end{bmatrix} \right\|_F^2 + \left\| W_2 - \left[ M_2 \mid t \right] \begin{bmatrix} S'_2 \\ 1_2^T \end{bmatrix} \right\|_F^2. \quad (11)$$

A megoldás zárt alakban fejeztem ki:

$$\begin{bmatrix} \hat{o}_1 \\ \hat{o}_2 \end{bmatrix} = [M_1 \mid -M_2]^\dagger [t_2 - t_1], \quad (12)$$

ahol  $o_1$  és  $o_2$  jelöli a közös pontot 3D-ben,  $t_1$  és  $t_2$  pedig 2D-ben.

### 3.b. Pontszerűen kapcsolt tagolt objektumok csoportosítása.

A csoportosítási feladat megoldására a 2.a. altézis során definiált becslés hibáját alkalmaztam:

$$\frac{1}{F} \|M_1 \hat{d}_1 + t_1 - M_2 \hat{d}_2 - t_2\|_2. \quad (13)$$

A csoportosítási feladat eldöntésére egy küszöböt kell a hibára meghatározni. A hiba mértékegysége pixel, ami a szükséges küszöb meghatározását megkönnyíti.

### 3.c. Tengelyesen kapcsolt tagolt objektumok csoportosítása.

Megmutattam, hogy a közös tengelyű tagolt objektumok esetén a tengellyel összekötött két objektum relatív mozgását leíró mátrixok sorvektorai közös tengelyű köríveket írnak le. A közös tengelyű körívek meghatározására két részből álló algoritmust készítettem:

1. Párhuzamos síkok illesztése az adathalmazokra.
2. Közös tengelyű körök meghatározása a síkokon.

A döntést itt is küszöbölés segítségével végzem el, a szükséges

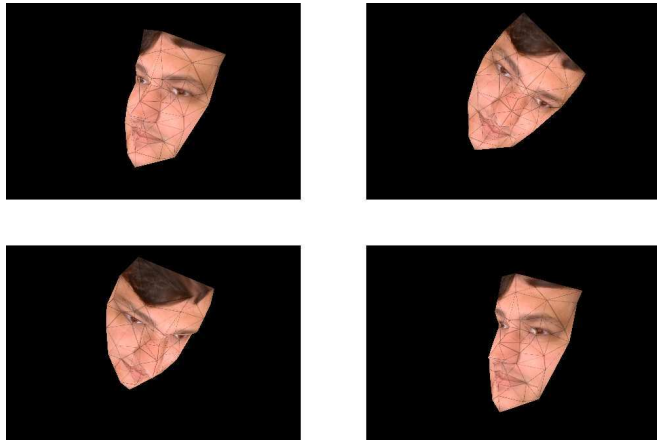
hibát a körillesztés pontatlansága adja.

Valamennyi altézist szintetikus és valós adatokon is megvizsgáltam, és a működést demonstráltam.

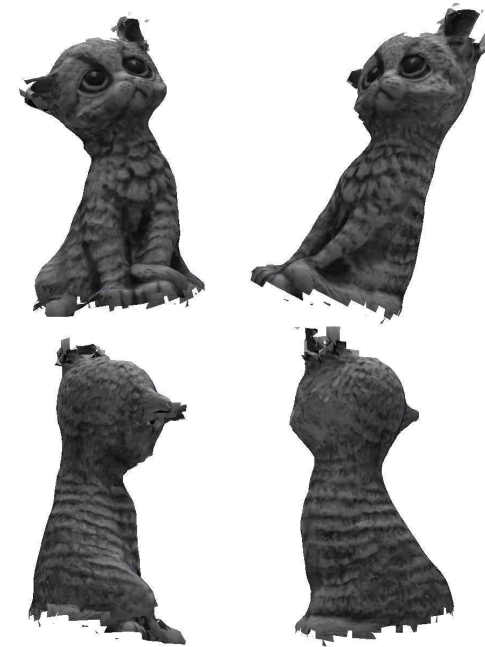
#### 4. A tudományos eredmények alkalmazása

Ahogy az már az 1. fejezetben említettem, a kutatómunka célja egy teljes rekonstrukciós rendszer kiépítése volt. Az értekezésben az egyes tézisek működését példákon mutatjuk be, itt, a tézisfüzetben a teljes rendszer működésére mutatok be két példát, amelyet a dolgozat szintén tartalmaz.

Az első példa során emberi arcról készített három darab képről készítettünk gyenge perspektív módszerek segítségével háromdimenziós rekonstrukciót. Az eredményt az alábbi képek szemléltetik:



Másik példánk során teljes körbefordulás során 80 képből készítettünk egy macskát ábrázoló szoborról textúrázott rekonstrukciót valódi perspektívát alkalmazó módszerek segítségével:



A rendszer, bár kész van, korántsem tökéletes. A kutatómunka tovább folytatódik: az egyes részfeladatokra jobb megoldásokat lehet készíteni, akár más tudományos eredményeinek a megvalósításával, akár új módszerek kifejlesztésével.

Céлом, hogy a megismert módszereket a műegyetemi oktatásba is áttemeljem, kiegészítve a már meglévő, számítógépes látással foglal-



kozó kurzusokat. A fejlesztési feladatba már eddig is sok hallgatót sikerült bevonni, reményeim szerint a jövőben nemcsak diplomamunkák és TDK dolgozatok, hanem akár további értekezések is születnek ezen a rendkívül érdekes, jövőbemutató tudományterületen.

## 5. A szerző publikációi a dolgozat témakörében

- [1] **Hajder Levente**: Mozgásérzékelés webkamerával. In *Elektrotechnika*, 98:203-205, 2005.
- [2] **Levente Hajder**, D. Chetverikov: Weak-perspective Structure from Motion for Stongly Contaminated Data. In *Pattern Recognition Letters*, 27:1581-1589, 2006.
- [3] **Levente Hajder**: Shape and Motion from Video. In *Proc. II. Magyar Számítógépes Grafika és Geometria Konferencia*, pp. 98–102, 2003.
- [4] **Levente Hajder**: Robust Structure from Motion by Outlier Rejection. In *Proc. Képfeldolgozók és Alakfelismerők Konferenciája*, pp. 30–36, 2004.
- [5] **Hajder Levente**, Chetverikov D., Kardos I., Renner G.: Aktív kontúrok és Fast Marching eljárás alkalmazása az orvosi képfeldolgozásban. In *Proc. Képfeldolgozók és Alakfelismerők Konferenciája*, pp. 90–96, 2004.
- [6] **Levente Hajder**, Dmitry Chetverikov: Robust Structure from Motion under Weak Perspective. In *Proc. 2<sup>nd</sup> International Symposium on 3D Data Processing, Visualization & Transmission*, pp. 828–835, 2004.
- [7] I. Kardos, **Levente Hajder**, D. Chetverikov D, G. Renner: Bone Surface Reconstruction from MR/CT Images using Fast Marching and Level Set Methods. In *Proc. Joint Hungarian–Austrian Conference on Image Processing and Pattern Recognition*, pp. 41–48, 2005.

- [8] **Levente Hajder**: An Iterative Improvement of the Tomasi-Kanade Factorization. In *Proc. III. Magyar Számítógépes Grafika és Geometria Konferencia*, pp. 30–36, 2005.
- [9] **Levente Hajder**, D. Chetverikov: Robust 3D Segmentation of Multiple Moving Objects under Weak Perspective. In *ICCV Workshop on Dynamical Vision*, CD-ROM, 2005.
- [10] **Levente Hajder**, D. Chetverikov: Robust 3D Segmentation of Multiple Moving Objects under Weak Perspective. In *Lecture Notes on Computer Science*, 4358, pp. 48–59, 2007.
- [11] **Levente Hajder**: 3D Motion Grouping of Articulated Objects. In *Computer Vision Winter Workshop (oral presentation after double blinded review)* pp. 125–130 ,2006.
- [12] **Levente Hajder**: 3D Motion Grouping of Articulated Objects. In *Automation and Applied Computer Science Workshop* pp. 131–142, 2006.
- [13] **Hajder Levente**: Tagolt objektumok rekonstrukciója videó alapján *Intelligens Rendszerek Fiatal Kutatók Szimpóziuma* (poszter) 2006.
- [14] **Levente Hajder**: Optimal Joint Estimation of Articulated Objects under Weak Perspective *Proc. Képfeldolgozók és Alakfelismerők Konferenciája*, pp. 71–78, 2007.
- [15] Csaba Kazó, Ákos Pernek, **Levente Hajder**: Texturing 3D Models from Images *Proc. Képfeldolgozók és Alakfelismerők Konferenciája*, pp. 148–156, 2007.

- [16] **Levente Hajder**, Ákos Pernek, Csaba Kazó : Structure from Motion by Fast Alternation. *Proc. Workshop of the Austrian Association for Pattern Recognition (OAGM/AAPR)*, Accepted for publication, 2007.
- [17] **Levente Hajder**, Ákos Pernek, Csaba Kazó : Fast and Precise Weak-Perspective Factorization. Submitted to *The 12th International Conference on Computer Analysis of Images and Patterns*, Vienna, Austria, 2007.
- [18] **Levente Hajder** : Grouping of Articulated Objects with Common Axis. Submitted to *The 12th International Conference on Computer Analysis of Images and Patterns*, Vienna, Austria, 2007.