

# Környező járművek mozgásának meghatározása irány becslésre és kép-párok közötti epipoláris geometriára építve

Golarits Marcell,<sup>1,2</sup> Rózsa Zoltán<sup>1,2</sup> és Szirányi Tamás<sup>1,2</sup>

<sup>1</sup> Gépi Érzékelés Kutatólaboratórium, Számítástechnikai és Automatizálási Kutatóintézet (SZTAKI),  
Eötvös Loránd Kutatói Hálózat (ELKH), Budapest

<sup>2</sup> Anyagmozgatási és Logisztikai Rendszerek Tanszék, Közlekedésmérnöki és Járműmérnöki Kar,  
Budapesti Műszaki és Gazdaságtudományi Egyetem (BME), Budapest

---

## Abstract

*A változó környezetben történő tájékozódáshoz elengedhetetlen a környező tárgyak mozgásának ismerete. A cikkben javasolunk egy módszert, amely egy mozgó kamera képén megjelenő dinamikus objektumok mozgására ad becslést. A módszer lényege, hogy a lehető legkevesebb információból azaz két egymást követő képből végezzük el a becslést. Ennek első lépéseként a kamera mozgás becslését végezzük, majd a képeken szegmentált járművek 3D pontjai között keressük merev transzformációt. A becslés során feltételezzük a járművek hosszirányban történő elmozdulását. A kevesebb információs szükséglet mellett a bemutatott két képet felhasználó módszer legalább kétszeres sebességgel képes működni, mint a három képet használó módszerek. A javasolt eljárást több autonóm vezetési adatbázison különböző közlekedési helyzetekre is teszteltük.*

---

## 1. Bevezető

A relatív kameramozgás becslése a gépi látás egyik alapvető problémája és a mai napig jelentősen kutatott területe, rengetek alkalmazással. Drónok, mobil robotok, autonóm autók stb. navigációja esetén elengedhetetlen a jármű elmozdulásának ismerete. A képfeldolgozásban olyan jól ismert eljárások, mint a Simultaneous Localization and Mapping (SLAM) és Structure from Motion (SfM) kvázi statikus környezetben képesek számítani a kamera elmozdulásokat és a környezetet feltérképezni. A térképezést további egyszerű fedélzeti szenzorokkal kiegészítve (pl.: elfordulás érzékelő, IMU, GPS) közvetve vagy közvetlenül a valós skála is meghatározható. Földi járműre szerelt kamera esetén ismert magasságot feltételezve a rekonstrukción sík illesztést alkalmazva becsülhető is a valós skála. Az ilyen rekonstrukciókból azonban a mozgó tárgyak szűrésre kerülnek, mivel a rajtuk detektált pontok nem illeszkednek a statikus környezet által meghatározott kamera mozgáshoz tartozó vetítési egyenletekbe. Azonban közlekedés szempontjából a mozgó objektumok nagyobb veszélyt is jelenthetnek, mint a statikus akadályok. Elég csak a gyalogosokra vagy más járművekre gondolni. Ezeknek a mozgó objektumoknak az azonnali szűrése kritikus a biztonság szempontjából<sup>12</sup>. Amíg más megoldásoknak több képek

sorozata szükségeltetik a környező objektumok mozgásának meghatározásához, az általunk javasolt megoldásnak mindössze két egymást követő felvételre van szüksége. Ez a különbség meghatározó lehet a féktávolság tekintetében.

A képpont egyeztetéseket alapul véve a dinamikus objektumok is rekonstruálhatóak egy relatív skálán<sup>15</sup>. Azonban ez a skála nem fog szükségszerűen egyezni a statikus rekonstrukció skálájával, sőt általában jelentősen eltér tőle. Ahhoz, hogy az objektum távolságának és sebességének meghatározásához szükséges a két rekonstrukció közötti skála transzformáció. A cikkben ennek a skála transzformációnak a meghatározására javasolunk egy új fajta megközelítést. Ehhez az elmozdulás és esszenciális mátrixának becslésére, a mozgó objektum rekonstrukciójára és arra feltételezésre építünk, hogy az objektum a képpárok között hosszirányú (első főtengely menti) elmozdulása a meghatározó. A legtöbb esetben ez a feltevés helytálló a közlekedési eszközökre, még kanyarodás közben is. Ez a megkötés könnyen transzformálható más jellegű objektumokra például emberekre, ahol az elmozdulás általában a harmadik főtengely szerinti az elmozdulás.

A cikkben javasolt hozzájárulások:

- Új megoldás a kamera mozgástól eltérően mozgó objek-

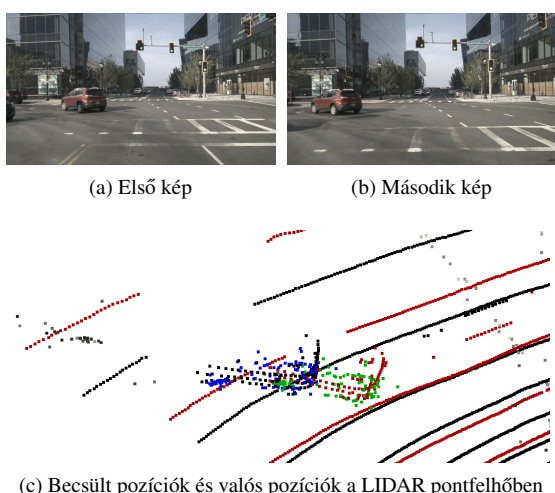


Figure 1: Példa a mozgó objektumok rekonstrukciójára egymást követő kamera képen. 1a és 1b a Nuscenes adat-szett<sup>2</sup> kép egymást követő képe. 1c ábrán pedig a a rekonstrukciók és a LIDAR pontfelhők látható közösen: vörös - első LIDAR mérés, fekete - második LIDAR mérés, zöld - első kép rekonstrukció, kék - második rekonstrukció. Az első LIDAR mérés és kamera rekonstrukció manuálisan lett illesztve az illusztráció érdekében.

tumok mozgás becslésére egy más követő mono kamera képek alapján.

- A mozgás predikció mindössze két egymást követő felvétel alapján, ami jelentősen csökkenti a becslés elvégzésének idejét.
- Túl mutat a jelenlegi state of the art technikákon, mivel általánosan alkalmazható 3D-ben történő becslésre nem úgy mint a korábbi megoldások (mint<sup>5, 8</sup>).

## 2. Kapcsolódó irodalmak

Az SfM vagy SLAM<sup>13</sup> algoritmusok számára kihívást jelentenek a dinamikusan változó helyszínek. A téma részletes megismerését a<sup>16</sup> összefoglaló cikk segíti. Az irodalomban már vannak olyan SfM amelyek képesek rekonstruálni a környezetben lévő merev vagy nem merev<sup>7</sup> testeket, azonban a gyakorlatban ez problémát okozhat az SfM köteget behangolás lépésénél<sup>10</sup>. A legtöbb esetben ezt a problémát faktorizációval oldják meg<sup>15</sup>.

A legfőbb probléma a testek mozgás rekonstrukciójával a relatív skála meghatározása a statikus rekonstrukcióhoz képest. A kulcspontokat és a vetítési egyenleteket felhasználva relatív skálán meghatározható az objektum trajektóriája, azonban a skála meghatározásához mindenképpen szükséges további megkötéseket bevezetni. Ezek a kényszerek származtathatóak különböző feltételezésekből, mint a<sup>4</sup> munkában, mély tanulási eljárással számított mélységkép-

ből vagy objektum alakmodelljének illesztéséből<sup>19</sup>. Mozgás modellekre építve<sup>5</sup>, mint állandó sebesség<sup>8</sup> is meghatározhatóak a megfelelő megkötések. A leggyakrabban használt feltételezés, hogy az objektumok mozgása merőleges a talaj normálisra<sup>3</sup> és a talajtól vett távolságuk állandó<sup>1</sup>. A fenti eljárások vagy kettőnél több kameraképre vagy speciális kamera és/vagy objektum mozgásra vagy a talaj sík irányának ismeretére építenek. Az úton valószínűsíthető gyér talajpont rekonstrukció (Figure 1c -en láthó szürke pontok) miatt valós körülmények között nem megbízhatóak ezek a megoldások. A javasolt eljárás előnyei ezzel szemben:

- Csak két kép szükséges a rekonstrukcióhoz.
- Nem feltételezzük meghatározott kamera mozgást és csak vizuális odometria szükséges (a statikus környezet rekonstrukciós nem szükséges);
- Nem feltételezzük az objektumok síkbeli mozgását (éppen ezért a talaj sík becslése sem szükséges); emellett az objektumok 3D-ben is mozoghatnak (pl. repülő);
- Nem szükséges mélytanulás vagy objektum modell.

## 3. A javasolt módszer

A következőkben részletesen bemutatjuk a módszert. A skála- és transzformáció becslésünket négy lépésre osztjuk.

### 3.1. Előfeldolgozás

Az algoritmusunk bemeneti adatainak generálásához három előfeldolgozási részfeladat végrehajtását javasoljuk. Először is meg kell határozni a kamerák belső paramétereit<sup>20</sup>. Ezután meg kell becsülni a relatív kamerapozíciókat. Kísérleteink során a COLMAP<sup>17</sup> segítségével robusztusan rekonstruáltuk a környezetet. Végül a mozgó objektumot detektálni kell (Yolo<sup>11</sup>), és egyeztetni az egymást követő képkockákon<sup>9</sup>. Az alfejzet módszerei nem képezik a javasolt módszer lényegi részét. Azonban a valós életben végzett kísérletek azonban azt mutatták, hogy a fenti módszerek jó teljesítményt nyújtanak a becslésünkhöz szükséges bemeneti adatok biztosításához.

### 3.2. Esszenciális mátrix becslés és rekonstrukció

Dinamikus objektumok esetében:

$$[p_{2m} \ 1]K_2^{-1}E_{mov}K_1^{-T}[p_{1m} \ 1]^T = 0 \quad (1)$$

ahol  $p_{2m}$  egy mozgó pont a második képen, és  $p_{1m}$  egy mozgó pont az első képen, a  $K_i$ ,  $i = 1, 2$  a kamera vetítési mátrixai (mono kamera esetünkben  $K_1 = K_2$ ). A 2 ábrán az általunk használt transzformációs mátrixokat és koordinátarendszert mutatjuk be.

$E_{mov}$  felbontható (a statikus háttér esszenciális mátrixához hasonlóan):

$$E_{mov} = S_{mov}R_2R_{mov,g}R_1^{-1} \quad (2)$$

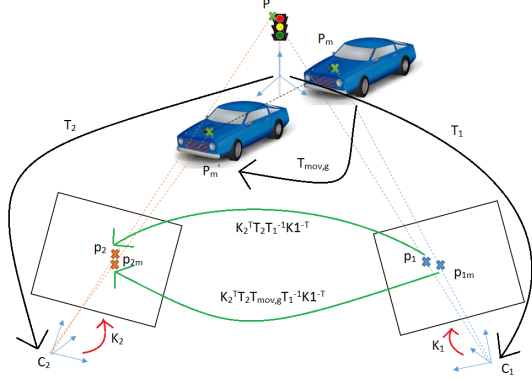


Figure 2: A transzformációk illusztrációja. A kék autó a mozgó objektumot ábrázolja (a forgalmi oszlop egy statikus objektum), két kameraállásból vizsgálva. A  $P$  egy statikus pont, a  $P_m$  és a  $P_m'$  a globális koordináta-rendszerben meghatározott mozgó 3D pontnak felel meg.  $T_{mov,g}$  írja le az autó (és a rajta lévő  $P_m$ ) mozgását.

ahol  $S_{mov}$  egy ferdén-szimmetrikus mátrix, a  $R_i$ ,  $i = 1, 2$  a globálisból a kamerakoordináta-rendszerbe leképező transzformációk forgatási mátrixai, hasonlóan, mint a a mozgó objektum mozgását leíró transzformációs mátrix:  $T_{mov,g} = [R_{mov,g} | t_{mov,g}]$

A felbontással definiálhatjuk a virtuális kamerapózok ( $v$  indexszel jelölve) vetítési mátrixait, pl. az első virtuális kamera koordináta-rendszerében:

$$S_{mov} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} = E_{mov} R_1 R_{mov,g}^{-1} R_2^{-1} \quad (3)$$

$$P_{v,1} = K_1^T [I | 0]^T = K_1^T T_{v,1} \quad (4)$$

$$P_{v,2} = K_2^T [R_v | \lambda_1 t]^T = K_2^T T_{v,2} \quad (5)$$

ahol  $R_v = R_1 R_{mov,g}^{-1} R_2^{-1}$ ,  $t = [t_x \ t_y \ t_z]^T$  és  $\lambda_1$  az ismeretlen skála.  $T_{v,1}$  és  $T_{v,2}$  transzformációs mátrixokat fogunk használni az általánosan használt specifikus leírás helyett (ahol a globális koordináta-rendszer megegyezik az első kamerapozíció koordináta-rendszerével). Teszteink során a COLMAP <sup>18</sup> segítségével rekonstruáltuk a dinamikus objektumokat egy virtuális kamerapár mozgásának becslésével.

### 3.3. Mozgási irány becslése

#### Koordináta transzformáció:

Az esszenciális mátrix határozza meg az objektum

mozgását egy adott léptékben (pl.  $\lambda_1 = 1$ ) és a 3D pontfelhő alakját.

A korábban kiháromszögelt pontfelhő átranzformálható a virtuális kamerák koordináta-rendszerébe:

$$[X_i \ Y_i \ Z_i \ 1]^T = T_{v,i}^{-1} [X_v \ Y_v \ Z_v \ 1]^T \quad (6)$$

ahol  $X_v$ ,  $Y_v$  és  $Z_v$  a háromszögelt 3D pontok pontkoordinátái, az  $i = 1 : 2$  pedig a virtuális kamerákat jelöli.

Megjegyzés: A virtuális kamerák és a statikus háttér rekonstrukciójának koordináta-rendszere megegyezik. És így, kiszámítjuk a dinamikus objektum koordinátáit a statikus rekonstrukció globális koordináta-rendszerében:

$$[P_m \ 1]^T = T_1 [\lambda_1 X_1 \ \lambda_1 Y_1 \ \lambda_1 Z_1 \ 1]^T \quad (7)$$

$$[P_m' \ 1]^T = T_2 [\lambda_1 X_2 \ \lambda_1 Y_2 \ \lambda_1 Z_2 \ 1]^T \quad (8)$$

Statistikai outlier-eltávolító algoritmust <sup>14</sup> alkalmazunk a háromszögelt objektumfelhő zajtalanítására. Végül a  $P_m$  és  $P_m'$  (a globális koordináta-rendszerünkben a mozgás előtti és utáni 3D-s mozgópontok) kiszámíthatók a kiválasztott skálán.

Főkomponens analízis (PCA): A háromszögelt és transzformált pontfelhőn a PCA <sup>6</sup> segítségével meghatározzuk a határoló téglatestet, és azt az irányt használjuk, amelyben a pontfelhő a legnagyobb szórással rendelkezik (hosszirány)  $\underline{N}$ , mint az objektum középponti mozgási irányára vonatkozó becslés. A 3 ábrán látható, hogy a mozgás iránya hogyan határozza meg a megfelelő skálát.

### 3.4. Relatív skála kiszámítása

Itt azt a feltételezést használtuk, hogy az objektum (vagy legalábbis súlypontja) olyan transzformáción megy keresztül, amely a hossz tengelyével párhuzamos translációként közelíthető:

$$(R_2^{-1} P_{c,2} - R_1^{-1} P_{c,1} \ \underline{N} \ C_2 - C_1) \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ 1 \end{pmatrix} = \underline{0} \quad (9)$$

ahol  $P_{c,1}$  és  $P_{c,2}$  a pontfelhő súlypontja az első és a második kamerakoordináta-rendszerben. A  $C_1$  és a  $C_2$  kameraközéppontok a globális statikus koordináta-rendszerben, a  $\lambda_2$  pedig a  $\underline{N}$  irányú elmozdulás mértéke.

Ez a három skaláregyenlet két ismeretlennel megoldható legkisebb négyzetes értelemben a  $\lambda_1$  esszenciális mátrix skálázására és a  $\underline{N}$  irányvektor  $\lambda_2$  skálázására.

A mozgó objektum transzformációs mátrixa:

Table 1: A javasolt módszer Vehicle Trajectory adatszetten lévő eredményeinek összehasonlítása a következővel: <sup>1</sup>. A <sup>1</sup> a teljes képsorozatot felhasználva kiválasztott feltételek nézetpárjait használja a méretarány becsléséhez. A becsléshez mi az egymást követő nézetpárok geometriai átlagát használtuk.

Hiba	Átlagos skálaarány eltérés					Átlagos trajektória hiba [m]				
	Lancer	Lincoln	Smart	Golf	Van	Lancer	Lincoln	Smart	Golf	Van
<sup>1</sup>	0.04	<b>0.04</b>	<b>0.04</b>	<b>0.06</b>	0.08	0.20	0.23	<b>0.33</b>	0.33	0.47
Javasolt	<b>0.03</b>	<b>0.04</b>	0.08	<b>0.06</b>	<b>0.05</b>	<b>0.15</b>	<b>0.22</b>	0.63	<b>0.23</b>	<b>0.35</b>

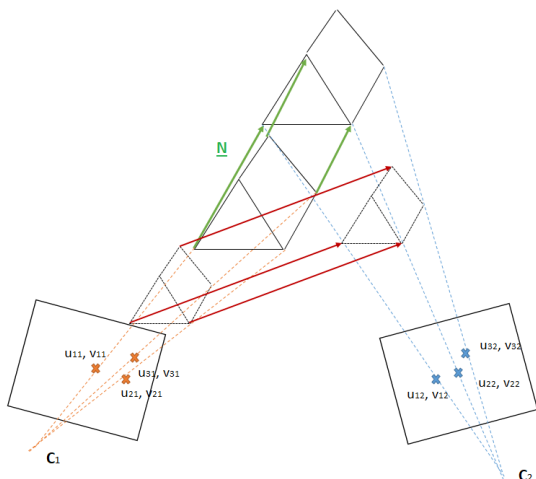


Figure 3: Az objektum mozgásbecslésének szemléltetése ismert alakokkal és relatív mozgásiránnyal. A mozgó tárgy egy háromszög alakú prizma. A mozgás iránya (itt a hossz tengelyével párhuzamos) határozza meg a megfelelő skálát. Csak a nagyobb háromszögprizmapár skálájának választása eredményezi a  $\underline{N}$  jellel jelölt mozgási irányt (zöld nyilak iránya).

$$T_{mov,g} = T_2^{-1} T_{v,2} T_{v,1}^{-1} T_1 \quad (10)$$

Végül megkaptuk az objektum mozgás transzformációs mátrixát.

#### 4. Teszt eredmények

A javasolt módszert a <sup>1</sup> Vehicle Trajectory adatszettjén értékeltük ki, amely legjobb tudomásunk szerint jelenleg az egyetlen megfelelő adatkészlet az eoru-motion becsléséhez mozgó mono kamerával. (Más adatkészletekből hiányoznak a pontos ground truth adatok a skálaarány becsléséhez).

A Vehicle Trajectory adatkészlet egy virtuális adatkészlet, amely hét szekvenciát (útvonalat), és mindegyik szekvenciához öt autót tartalmaz. A ground truth objektummaszkok

az adatkészlettel együtt kerültek közzétételre így ezek is felhasználásra kerültek annak érdekében, hogy az eredményeket az előfeldolgozás ne befolyásolja. Az adatkészletet mono és sztereó kamerák jármű útvonalának becslésére tervezték. Mivel mi mono kamerával dolgozunk, a bal oldali képek kerültek felhasználásra. Ez igaz a hetedik szekvencia (Bumpy Road) kivétel, mivel ott párhuzamos kamera és objektum mozgás van (ami az algoritmusunk számára degenerált eset). Ebben az esetben a bal és a jobb oldali kamera képeit felváltva (különböző időpillanatokban) használtuk a teljes adathalmaz kiértékeléséhez (megjegyzés: csak két egymás utáni képkocka, nem degenerált esettel elegendő lenne a relatív skála becsléséhez).

A skálabcslés két esetét különböztetjük meg. 0,75-nél nagyobb vagy egyenlő (megbízhatónak ítélt) degeneráltsági fokkal (az egységvektorok skaláris szorzatának abszolút értéke a becsült tárgy és a kamera mozgása irányában) rendelkező, vagy azalatti esetek. Csak az előbbieket kerültek felhasználásra a globális kiértékeléshez. Voltak azonban olyan pályák, ahol nem volt e degeneráltsági fok alatti kamerapozíció. Ebben az esetben a rendelkezésre álló becsléseket használtuk. A számított skálák geometriai átlagát (az egymást követő képkockákból) használtuk a pálya globális skálájának meghatározásához.

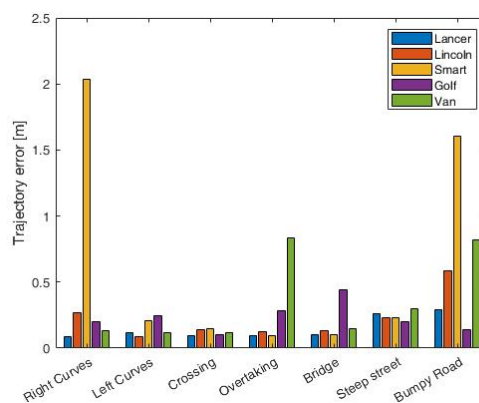


Figure 4: Különböző szekvenciák és járművek trajektória hibája a Vehicle Trajectory adatszetten <sup>1</sup> a javasolt módszerrel

Az átlagos skálát legalább olyan pontosan becsültük meg, mint a baseline az adatkészlet öt autójából négyben. A trajektória hibák mind a 35 különböző esetben 4 ábrán láthatók. Ha összehasonlítjuk a 4 ábrát a <sup>1</sup>-ben közzétett hasonló ábrával, az látható, hogy az esetek többségében felülmúltuk a baseline-t. A 1 táblázatban a teljesítménymérőszámokat összehasonlítjuk a <sup>1</sup> értékeivel. Egy kivétellel minden esetben pontosabb volt a trajektória becslésünk.

## 5. Konklúzió

Ebben a cikkben bemutattunk egy megközelítést két képkocka között a környező járművek mozgásának rekonstrukciójára. Ehhez megbecsüljük az objektumhoz és a háttérrekonstrukcióhoz kapcsolódó relatív skálát. A módszer nem igényli a talaj előzetes becslését, a kamera mozgásának korlátozását, az objektum alakmodelljét, vagy bármilyen más feltételezést vagy tanulást. A skála propagálása lehetséges, amint azt az értékelés is mutatja. Mivel csak két képkockát használ a módszer, kétszer gyorsabban képes megjósolni a mozgó tárgy pozícióját és sebességét, mint egy három képkockát használó módszer. Előnyös az autonóm vezetési alkalmazásokban, ahol az előrejelzés sebessége (különösen egy veszélyes mozgó objektum esetében) elengedhetetlen. A módszer kvantitatív eredményei ezt igazolják. A legtöbb esetben felülmúltuk a state-of-the art módszert teljesítményét. A jövőben más objektumok mozgásbecslését tervezzük.

## 6. Köszönetnyilvánítás

A publikációban szereplő kutatást, amelyet a SZTAKI valósított meg, az Innovációs és Technológiai Minisztérium és a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal támogatta az Autonóm Rendszerek Nemzeti Laboratórium keretében. Továbbá a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal (NKFIH) OTKA (K 139485) pályázata támogatta.

## References

1. Sebastian Bullinger, Christoph Bodensteiner, Michael Arens, and Rainer Stiefelhagen. 3D vehicle trajectory reconstruction in monocular video data using environment structure constraints. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.
2. Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
3. Chang Yuan and G. Medioni. 3D reconstruction of background and objects moving on ground plane viewed from a moving camera. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2261–2268, 2006.
4. Kemal Egemen Ozden, Kurt Cornelis, Luc Van Eycken, and Luc Van Gool. Reconstructing 3D trajectories of independently moving objects using generic constraints. *Computer Vision and Image Understanding*, 96(3):453 – 471, 2004. Special issue on model-based and image-based 3D scene representation for interactive visualization.
5. Jiahui Huang, Sheng Yang, Tai-Jiang Mu, and Shi-Min Hu. Clustervo: Clustering moving instances and estimating visual odometry for self and surroundings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
6. Alexander Ilin and Tapani Raiko. Practical approaches to principal component analysis in the presence of missing values. *Journal of Machine Learning Research*, 11(66):1957–2000, 2010.
7. S. Kumar, Y. Dai, and H. Li. Multi-body non-rigid structure-from-motion. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 148–156, 2016.
8. A. Kundu, K. M. Krishna, and C. V. Jawahar. Real-time multibody visual SLAM with a smoothly moving monocular camera. In *2011 International Conference on Computer Vision*, pages 2080–2087, 2011.
9. Matt L. Miller, Harold S. Stone, , Ingemar J. Cox, and Ingemar J. Cox. Optimizing murty's ranked assignment method. *IEEE Transactions on Aerospace and Electronic Systems*, 33:851–862, 1997.
10. K. E. Ozden, K. Schindler, and L. Van Gool. Multi-body structure-from-motion in practice. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1134–1141, 2010.
11. J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, 2017.
12. Z. Rozsa and T. Sziranyi. Object detection from a few lidar scanning planes. *IEEE Transactions on Intelligent Vehicles*, 4(4):548–560, 2019.
13. Zoltan Rozsa, Marcell Golarits, and Tamas Sziranyi. Localization of map changes by exploiting SLAM residuals. In Jacques Blanc-Talon, Patrice Delmas, Wilfried Philips, Dan Popescu, and Paul Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems*, pages 312–324, Cham, 2020. Springer International Publishing.
14. Radu Bogdan Rusu, Zoltan Csaba Marton, Nico

- Blodow, Mihai Dolha, and Michael Beetz. Towards 3D point cloud based object maps for household environments. *Robot. Auton. Syst.*, 56(11):927–941, November 2008.
15. R. Sabzevari and D. Scaramuzza. Multi-body motion estimation from monocular vehicle-mounted cameras. *IEEE Transactions on Robotics*, 32(3):638–651, 2016.
  16. Muhamad Risqi U. Saputra, Andrew Markham, and Niki Trigoni. Visual SLAM and structure from motion in dynamic environments: A survey. *ACM Comput. Surv.*, 51(2), February 2018.
  17. Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
  18. Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.
  19. S. Yang and S. Scherer. Cubeslam: Monocular 3-D object SLAM. *IEEE Transactions on Robotics*, 35(4):925–938, 2019.
  20. Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.