

## Multi-Body Structure from Motion

Bc. Jana Krčka

Diplomová práce Jana Krčka se zabývá algoritmem pro řešení problému struktury z pohybu s více objekty, tzv. *Multi-Body Structure from Motion*. Cílem metody je na základě sekvence obrazů rozpoznat 3D polohy klíčových bodů detekovaných v obrazech a odhadnout vlastní pohyb kamery, přičemž se předpokládá, že scéna sestává z malého počtu pevných objektů, které se během pohybu kamery také pohybují. Jde o těžký problém z oboru počítačového vidění, kdy je třeba zároveň odhadovat parametry pohybu kamery, segmentovat scénu na objekty a odhadovat parametry jejich nezávislých pohybů. Na problému je nejtěžší určit počet nezávislých objektů ve scéně.

Předložená diplomová práce se nepřímo zabývá právě tímto problémem. Diplomant měl analyzovat existující řešení, označovaného v práci jako YASFM, navrhnout na základě této analýzy vylepšení a demonstrovat funkčnost zvoleného výsledného řešení. Z problémů, které YASFM měl, si diplomant vybral malou výtěžnost interpretace dat, jejímž projevem byla malá hustota výsledného 3D modelu. Diplomant si povšiml toho, že YASFM často považuje koherentní podmnožiny (sekvencí) 3D bodů za odlehle a vylučuje je z výsledku. Navrhl tedy metodu, která takto odmítnutá data reviduje a přidává do výsledného modelu. Zvolený postup je možno označit za hladový heuristický algoritmus, který doplňuje YASFM o mechanismus identifikace koherentních struktur v bodech odmítnutých YASFM a jejich vložení do řešení, po odhadnutí příslušné geometrické transformace. Formálně bylo zadání splněno. Předložená práce je napsána srozumitelnou angličtinou. Má dostačující rozsah co do počtu stran. Je doplněna dostatečně obsáhlým seznamem bibliografických referencí. Má logickou strukturu. K práci mám však řadu závažných připomínek po obsahové a technické stránce a po stránce kvality zpracování výkladových a dokumentačních částí textu.

### Shrnující hodnocení

- Práce je psána velmi zkratkovitě. Nevyhovuje zcela standardům stylu technického textu.
- Zcela chybí formální, matematickým jazykem provedený popis úloh, které diplomant implementoval. Pseudokódy uvedené v práci naprosto nepostačují k porozumění tomu, jak algoritmy fungují.
- Jediná formálnější část, kap. 4.2, popisuje standardní algoritmus, ale zcela chybným způsobem, se závažnými chybami v matematickém popisu a notaci.
- Popis YASFM je velmi povrchní, bez hlubšího porozumění metodám v něm použitým.
- Experimentální část práce popisuje jen několik ad-hoc validačních experimentů, bez vyhodnocení výsledků a příslušné diskuse o příčinách pozorovaných selhání.

### Konkrétní výhrady

1. Není příliš jasné, co se míní pod pojmem *complementary model* a proč by měl fungovat mechanismus jeho postupné segmentace. Rozumím tomu tak, že YASFM odmítne interpretovat část dat a označí je za vychýlené z modelu (*outliers*). Navržená metoda postupuje tak, že reviduje toto rozhodnutí na základě stejného (?) statistického modelu.

Proč místo toho není vylepšen statistický model v YASFM? Jak je zaručeno, že výsledná procedura poskytne hustší 3D model stejné kvality, bez zhoršení výsledků YASFM?

2. Kap. 1.3 se sice snaží o vysvětlení základních pojmů, jako je *model*, *segment* a *group*, ale toto vysvětlení je nedostatečné, protože popis je jen neformální. Ale pro porozumění práce by byl potřeba přesný formální popis, jinak mnoho části textu nedává smysl.
3. Kap. 2 bloudí mezi různými komponentami standardního postupu SfM a jeho variant, bez celkového pohledu na řešenou úlohu, takže text působí značně bezcílne a chaoticky.
4. Kap. 2.2 měla nejprve vysvětlit problém segmentace objektů a teprve potom popisovat metody řešení tohoto problému.
5. Není vůbec jasné jak probíhá *merging*. Co je kritériem, podle kterého se rozhodne o sloučení? Jak se sloučení realizuje (vzorce)? Ani není jasné, že pro odhad transformace musí být k dispozici společné body (to je zmíněno až v závěru). Co když k dispozici nejsou?
6. Algoritmy by měly být popsány normálním textem, čtenář je odkázán na komentáře v pseudokódu, ty ale nejsou dostatečné k pochopení. Výsledkem je nesrozumitelný, nepřehledný a nedostatečný popis zatížený zbytečnými implementačními detaily. Popis algoritmu a popis jeho implementace mají být odděleny.
7. Co ilustruje obr. 3? Jak to je relevantní pro popisovanou úlohu?
8. Ke kap. 4.2 mám několik výhrad:
  - (a) Vzorec (1) nedává smysl,  $A$ ,  $B$  jsou množiny a  $R$ ,  $t$ ,  $s$  jsou parametry pohybu.
  - (b) Proč vzorec (2) nepoužívá standardní notaci pro vektory popisující polohu bodů v prostoru, například  $\mathbf{x}_A^i$ ?
  - (c) Proč se vyžaduje stejný počet  $n$  prvků v množinách  $A$  a  $B$ ? To v praxi není skoro nikdy splněno.
  - (d) Proč se uvažuje škálování  $s$ , když v problému jde o pohyb tuhého tělesa?
  - (e) Popis použití SVD rozkladu matice  $\mathbf{v}$  (4) na str. 17 je velmi nestandardní směs počítačového kódu a matematického zápisu. Postup by mohl být popsán takto:  
Let  $\mathbf{UDV}^\top$  be the SVD decomposition of  $3 \times 3$  matrix  $\mathbf{K}$ . Let  $s_1 = \det \mathbf{U}$ ,  $s_2 = \det \mathbf{V}$ , and  $\mathbf{S} = \text{diag}(1, 1, s_1 s_2)$ . Then the solution for the rotation matrix  $\mathbf{R}$  is given by  $\mathbf{R} = \mathbf{USV}^\top$ .
  - (f) Termín *direction correlation matrix* není pro kovarianční matici standardní. Takové použití vyžaduje bibliografickou referenci.
  - (g) Co popisuje vzorec (5)? Co je argumentem normy? Jaká norma je použita?
9. Kap. 5.1, která má popisovat implementaci, má formu, jakou mohou mít komentáře v gitu, ale ne formu technického textu. Čtenář například očekává popis toho, co bylo netriviálního na implementaci popsaných doplňků.
10. Ke kap. 6 mám tyto výhrady
  - (a) Zcela chybí diskuse k experimentálním výsledkům.
  - (b) Chybí dokumentace selhání modifikované metody.
  - (c) V kap. 6.2 chybí definice prahů.

- (d) Chybí komentář k výsledku experimentu 1. Proč YASFM selhal? Uspěl by, kdyby pohyby kamer byly menší? Větší? Kdyby ve scéně nebylo silně texturované pozadí? Kdyby se jinak nastavily prahy?
  - (e) Co by se stalo v exp. 2, kdyby se použila modifikace popsaná v diplomové práci? Co se myslí pod pojmem *higher-spec camera*?
  - (f) Co způsobilo, že exp. 3 byl úspěšný, přičemž exp. 1 byl neúspěšný? Jaké výsledky dá modifikovaná metoda?
  - (g) Výsledky exp. 4 se zdají demonstrovat selhání algoritmu faktorizace na strukturu a pohyb. Jak v takovém případě pomůže navrhovaná modifikace? Jaký je výsledek modifikovaného algoritmu?
  - (h) Co se myslí pod pojmem *more three-dimensional object*? Máme snad více a méně třídídimenzionální objekty?
  - (i) Podle čeho vidíme, že modely 3 a 4 v tabulce na str. 37 jsou komplementární? Snad ne podle počtu bodů. . .
11. Proč v seznamu literatury na konci práce chybí reference [8] a [11] ze seznamu doporučené literatury v zadání?
  12. Není zvykem kombinovat číselné bibliografické reference se jmény autorů. Pokud je nutné uvádět jména autorů v textu, lze použít jiný citační styl.
  13. Není nutné opakovat reference, jako se děje například u YASFM [1], to stačí jen jednou.
  14. Reference nemohou následovat po tečce na konci věty (str. 4).
  15. Číslované obrázky, tabulky, apod. musí mít v textu odkaz (a komentář).
  16. Reference jako Figure 1, Section 3, Table 8 se v angličtině píše s velkým písmenem (a bez členu), tedy nikoliv figure 1, apod.
  17. Věta nesmí začínat matematickým symbolem, jako např. na str. 17.
  18. Chybná angličtina: str. 8: vertice  $\rightarrow$  vertex, několikrát amount  $\rightarrow$  number.

## Otázky do diskuse

1. Jaké může mít metoda praktické využití? Zkoušel jste například metodu na prakticky významné instanci problému *Multi-Body SfM*, například segmentace a odhad polohy účastníků silničního provozu?
2. Když jste zjistil, že největším problémem YASFM je shlukovací algoritmus, proč jste nenavrhl a nevyzkoušel jeho modifikaci, místo toho, abyste se zabýval méně významným problémem?
3. Kdo implementoval algoritmus registrace “pomocí SVD”, popsany v kap. 4.2? Byl to vedoucí práce, jak je naznačeno na str. 19?
4. Jaké výsledky dá modifikovaný YASFM, když scénu nasnímám z velkého množství blízko položených kamer?

Kvalitu diplomové práce posuzuje oponent zejména podle kvality předloženého spisu. Ten je velmi neuspokojivý. Rozsah a úroveň textu práce odpovídá diplomové práci jen stěží. Vzhledem k tomu, že zadání bylo formálně splněno, doporučuji práci k prezentaci a navrhuji hodnocení stupněm **E – dostatečně**.

Doc. Dr. Techn. Ing. Radim Šára  
katedra kybernetiky FEL ČVUT  
1. 6. 2017