



PROJEKT Č. VI20172020068

NÁSTROJE A METODY ZPRACOVÁNÍ VIDEO A OBRAZU
PRO ZVÝŠENÍ EFEKTIVITY OPERACÍ BEZPEČNOSTNÍCH A
ZÁCHRANNÝCH SLOŽEK (VRASSEO)

TECHNICKÁ ZPRÁVA 2018

**AUTOMATICKÁ KALIBRACE
SLEDOVÁNÍM DOPRAVY**

Vojtěch Bartl, Adam Herout

Fakulta informačních technologií
Vysokého učení technického v Brně
Božetěchova 1/2
612 66 Brno, Česko

Prosinec 2018

Obsah

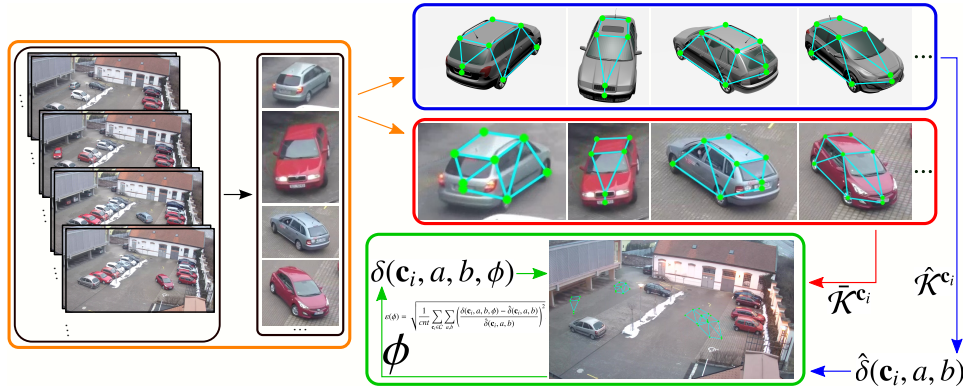
1	Úvod	1
2	Metoda automatické kalibrace z pozorování dopravy	3
3	Experimentální vyhodnocení	5

Abstrakt

Technická zpráva popisuje výzkum a vývoj algoritmu pro plně automatickou kalibraci dohledové kamery z průjezdů vozidel (osobních automobilů). Vstupem algoritmu je obraz z libovolné stacionární dohledové kamery zabírající alespoň z části vozovku s automobily (nejen přímou silnici, ale i parkoviště apod.). Algoritmus detekuje projíždějící vozidla, na nich pomocí konvoluční neuronové sítě vyhledává význačné body buď určené člověkem, nebo automaticky navržené strojem. Na základě pozic význačných bodů a určeného přesného typu vozidla (rozpoznání přesného typu je také realizováno konvoluční neuronovou sítí) je z jednoho pozorování určena dílčí informace; takovéto dílčí informace z velkého množství průjezdů (desítky, stovky vozidel) jsou agregovány algoritmem globální optimalizace pro dosažení automatické a co nejpřesnější kalibrace kamery. To, že je kamera zkalibrována, umožní pokročilejší analýzu záběrů z kamery: měření rychlosti a velikosti objektů, automatické pochopení scény (odlišení popředí a pozadí, mapování částí obrazu, kde se obvykle pohybují vozidla, kde chodci, kde se obvykle nic nevyskytuje apod.), efektivnější a přesnější detekce objektů apod.

1 Úvod

Vyvíjené algoritmy mají dosáhnout zkalibrování kamery zcela automatickou cestou – jediná využitá apriorní informace je, že kamera pozoruje (alespoň z části) osobní automobily pohybující se v jedné rovině (po vozovce, po parkovišti, apod.). Přehled navrženého přístupu je znázorněný na obrázku 1.



Obrázek 1: Přehled přístupu k automatické kalibraci z pozorování projíždějících vozidel. Automobily pozorované ve vstupním videu (oranžová) jsou přesně klasifikovány (pro získání přesného výrobce/modelu/modelové varianty) a zpracovány rozpoznavačem význačných bodů (červená). Pro viditelné význačné body $\bar{\mathcal{K}}^{c_i}$ je určena jejich 3D pozice $\hat{\mathcal{K}}^{c_i}$ z CAD modelu (modrá). Vzájemné vzdálenosti známé z 3D pozic $\hat{\delta}(\mathbf{c}_i, a, b)$ jsou porovnány s pozorovanými 3D vzdálenostmi $\delta(\mathbf{c}_i, a, b, \phi)$ a model kamery ϕ je optimalizován (zelená) globální optimalizační metodou pro získání nejlepšího řešení.

Kalibrace kamery je důležitým krokem ve většině aplikací počítačového vidění. V různých scénářích dohledu kamerami, kalibrace včetně měřítka (pro získání pozice bodů v jednotkách světového souřadného systému, tj. v metrech, nikoliv jednotkách obrazu) je důležitá. Zhang [1] zpopularizoval kalibraci vložím vhodného vzoru / markeru známých vlastností; v jeho případě byla použita rovinná šachovnice, ale v principu mohou být použity různé objekty, rovinné i nerovinné [2, ?]. Nicméně, při kamerovém dohledu v reálných scénách, zvláště při velkém počtu zpracovávaných kamer, je extrémně nepohodlné kalibrovat kamery vkládáním markerů do scén a prováděním měření vzdáleností ve scéně (například uprostřed dopravy v jízdách pruzích na dálnici).

The goal of our work is to develop fully automatic calibration algorithms for surveillance, providing the internal camera parameters, camera's rotation and translation with regard to the ground plane, and also the scene's scale so that measurements can be done in the world units (meters). We focus on traffic surveillance, so in this article, we are using vehicles as objects of

know properties (“markers”), but the algorithms presented here work with any other suitable rigid objects.

2 Metoda automatické kalibrace z pozorování dopravy

Všechny snímky vstupního videa jsou zpracovány detektorem vozidel (v našem případě Faster R-CNN [3], naučený na vlastní datové sadě COD20k [4]). Detekovaná vozidla jsou přesně rozpoznána naším dříve navrženým klasifikátorem [5]. Pro nejčastější modely vozidel jsou detekovány význačné body další neuronovou sítí [6]. Navržené řešení je založeno na 2D-3D korelacích a tudíž význačné body musí být co nejpřesnější.

Vstupní video je tedy transformováno do množiny pozorování vozidel:

$$\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_N\} \quad (1)$$

a pro každé vozidlo \mathbf{c}_i je určena množina viditelných význačných bodů:

$$\bar{\mathcal{K}}^{\mathbf{c}_i} = \{\bar{\mathbf{k}}_1^{\mathbf{c}_i}, \dots, \bar{\mathbf{k}}_K^{\mathbf{c}_i}\}. \quad (2)$$

Pouze vozidla nejčastějších typů (v našem případě devíti) jsou zahrnuta do \mathcal{C} ; pro tato vozidla máme dostupný přesný 3D CAD model. Tyto 3D modely jsou ručně zpracovány pro získání přesných 3D pozic význačných bodů. Pro každé pozorované vozidlo \mathbf{c}_i jsou tak známy správné 3D souřadnice v lokálním souřadném systému vozidla:

$$\hat{\mathcal{K}}^{\mathbf{c}_i} = \{\hat{\mathbf{k}}_1^{\mathbf{c}_i}, \dots, \hat{\mathbf{k}}_K^{\mathbf{c}_i}\}. \quad (3)$$

Tyto přesné 3D souřadnice význačných bodů tak definují referenční 3D vzdálenosti (páry význačných bodů jsou označeny indexy a a b):

$$\hat{\delta}(\mathbf{c}_i, a, b) = \left| \hat{\mathbf{k}}_a^{\mathbf{c}_i}, \hat{\mathbf{k}}_b^{\mathbf{c}_i} \right|. \quad (4)$$

Příklad těchto vzdáleností $\hat{\delta}(\mathbf{c}_i, a, b)$, společně s několika příklady detekovaných 2D pozic význačných bodů v reálných snímcích je na obrázku 2.

Kamera je modelována standardním pin-hole modelem:

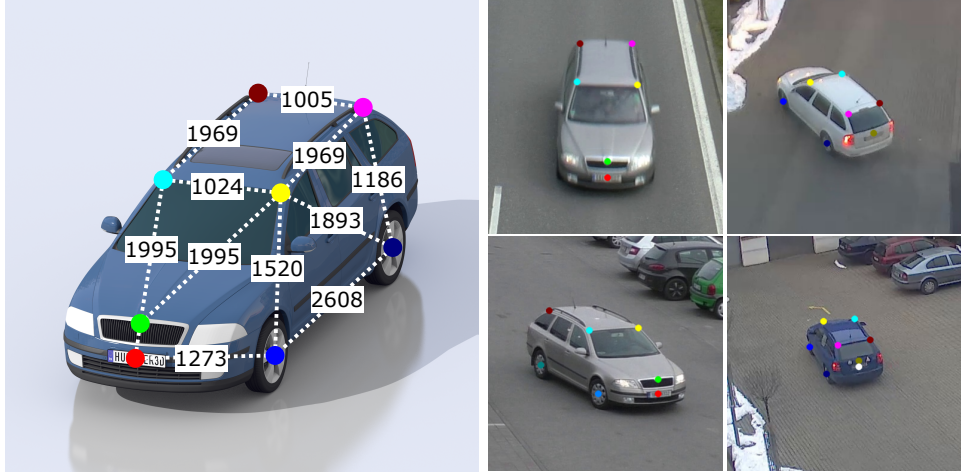
$$\lambda \mathbf{x}' = \mathbf{K} [\mathbf{R} | \mathbf{t}] \mathbf{x}. \quad (5)$$

Na základě tohoto modelu a parametrů kamery ϕ pak

$$\lambda \begin{bmatrix} \bar{\mathbf{k}}_j^{\mathbf{c}_i} \\ 1 \end{bmatrix} = \phi(\mathbf{K}) \left(\phi(\mathbf{R}) \mathbf{k}_j^{\mathbf{c}_i}(\phi) + \phi(\mathbf{t}) \right), \quad (6)$$

což může být upraveno na

$$\phi(\mathbf{R})^{-1} \phi(\mathbf{K})^{-1} \lambda \begin{bmatrix} \bar{\mathbf{k}}_j^{\mathbf{c}_i} \\ 1 \end{bmatrix} = \mathbf{k}_j^{\mathbf{c}_i}(\phi) + \phi(\mathbf{R})^{-1} \phi(\mathbf{t}) \quad (7)$$



Obrázek 2: *vlevo*: Vzdálenosti $\hat{\delta}$ (v milimetrech) na 3D modelu *Škoda Octavia mk2*. Pro přehlednost je ukázána pouze podmnožina celé množiny vzdáleností $\hat{\mathcal{K}}^{c_i}$. *vpravo*: Příklady detekovaných význačných bodů na skutečných obrázcích vozidla téhož typu.

a dále:

$$\mathbf{k}_j^{c_i}(\phi) = \phi_{(\mathbf{R})}^{-1} \left(\phi_{(\mathbf{K})}^{-1} \lambda \begin{bmatrix} \bar{\mathbf{k}}_j^{c_i} \\ 1 \end{bmatrix} - \phi(\mathbf{t}) \right). \quad (8)$$

Projektivní měřítko λ může být vyjádřeno z rovnice (7) s použitím souřadnice Z ze známého 3D CAD modelu $\hat{\mathbf{k}}_j^{c_i}$. Pouze třetí složka všech sloupcových vektorů je použita z (7) (operátor $[\mathbf{x}]_3$ značí extrakci třetího členu vektoru):

$$\lambda = \frac{\begin{bmatrix} \hat{\mathbf{k}}_j^{c_i} \\ 1 \end{bmatrix}_3 + \begin{bmatrix} \phi_{(\mathbf{R})}^{-1} \phi(\mathbf{t}) \\ 1 \end{bmatrix}_3}{\begin{bmatrix} \phi_{(\mathbf{R})}^{-1} \phi_{(\mathbf{K})}^{-1} \begin{bmatrix} \bar{\mathbf{k}}_j^{c_i} \\ 1 \end{bmatrix} \\ 1 \end{bmatrix}_3} \quad (9)$$

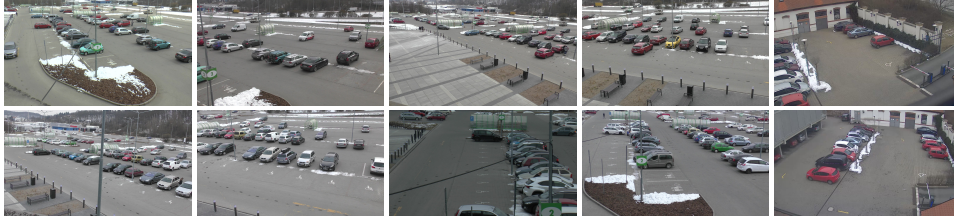
Pro každé vozidlo \mathbf{c}_i a každou dvojici význačných bodů a, b ve světovém souřadném systému $\mathbf{k}_j^{c_i}(\phi)$ je určena jejich 3D vzdálenost jako:

$$\delta(\mathbf{c}_i, a, b, \phi) = |\mathbf{k}_a^{c_i}(\phi), \mathbf{k}_b^{c_i}(\phi)|. \quad (10)$$

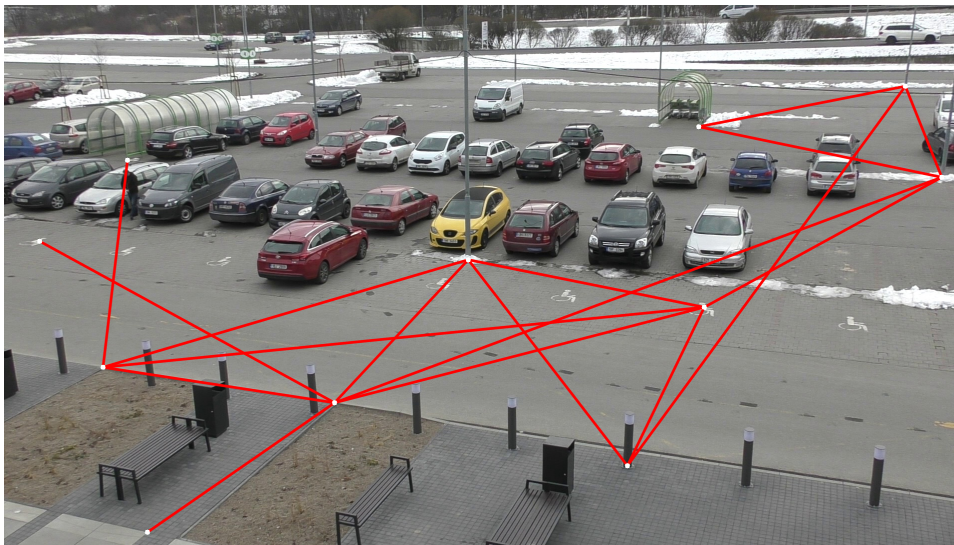
Celková chyba/cena všech pozorování ve videu za předpokladu daných kalibračních parametrů ϕ může být vyjádřena následovně:

$$\varepsilon(\phi) = \sqrt{\frac{1}{cnt} \sum_{\mathbf{c}_i \in \mathcal{C}} \sum_{a, b} \left(\frac{\delta(\mathbf{c}_i, a, b, \phi) - \hat{\delta}(\mathbf{c}_i, a, b)}{\hat{\delta}(\mathbf{c}_i, a, b)} \right)^2}. \quad (11)$$

Proces kalibrace, podobně jako u jiných kalibračních metod, sestává z nalezení takových parametrů kamery ϕ , které minimalizují tuto chybovou funkci. V našich experimentech využíváme metodu *Differential Evolution* [7] pro minimalizaci (11).



Obrázek 3: Příklady obrázků z nové datové sady *BrnoCarPark*. Data zahrnují 11 videí z parkovišť s různou intenzitou pohybu vozidel, obsazeností parkovišť, atd.



Obrázek 4: Příklad anotovaných “ground-truth” vzdáleností v nově pořízené datové sadě *BrnoCarPark*.

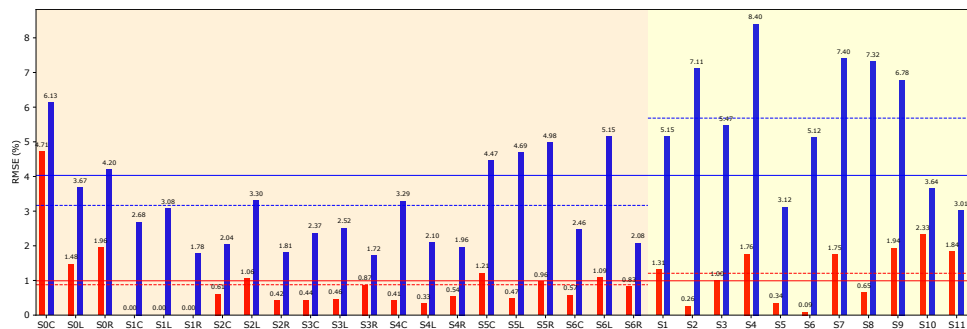
3 Experimentální vyhodnocení

Navrženou metodu pro automatickou kalibraci kamery jsme podrobně vyhodnotili na datové sadě, kterou jsme pořídili dříve (*BrnoCompSpeed*) i na nově pořízené datové sadě z parkovišť, sestavené speciálně pro tento účel (*BrnoCarPark*, obrázky 3 a 4). Výsledky vyhodnocení jsou shrnuty na obrázku 5.

Každá z testovacích scén je vybavena měřeními v reálném světě mezi význačnými body v rovině obrazu (viz obrázek 4 pro příklad):

$$\hat{\mathcal{D}} = \{\hat{d}_1, \dots, \hat{d}_D\}. \quad (12)$$

2D koncové body těchto měření mohou být promítnuty z roviny obrazu do roviny vozovky na základě kalibračních parametrů získaných naší metodou,



Obrázek 5: Přesnost navržené kalibrační metody oproti anotovaným kalibracím. Červené sloupce značí zadanou/anotovanou kalibraci a modré sloupce zobrazují přesnost navrženého algoritmu; červené a modré vodorovné čáry jsou průměry přes dataset a přes všechna měření. Vše je vyhodnoceno na datasetu *BrnoCompSpeed* (levá část) a novém datasetu *BrnoCarPark* (pravá část). Anotovaná kalibrace má průměrnou chybu $RMSE$ 0,99 %, navržená metoda má průměrnou $RMSE$ 4,03 %.

s použitím stejné techniky, která byla popsána v kapitole 2, rovnice (8). Měření mezi body zpětně promítnutými tímto způsobem

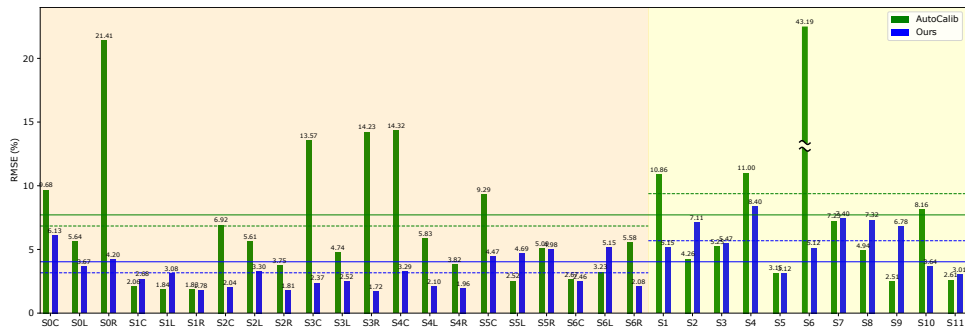
$$\mathcal{D} = \{d_1, \dots, d_D\} \quad (13)$$

mohou být vyhodnocena oproti správným měřením (ground truth) $\hat{\mathcal{D}}$ vyhodnocením *relative root mean square error*:

$$RMSE = \sqrt{\frac{1}{D} \sum_{i=1}^D \left(\frac{d_i - \hat{d}_i}{\hat{d}_i} \right)^2}. \quad (14)$$

I když měření v reálném světě a korespondující anotace 2D bodů ve scéně byly provedeny co nejpřesněji bylo možné, nevyhnutelně došlo k jistým nepřesnostem. Pro vyhodnocení přesnosti *ground-truth* dat jsme provedli kalibraci na základě 2D měření v reálném světě $\hat{\mathcal{D}}$, za použití stejné metodologie popsané v kapitole 2 (souřadnice Z všech bodů je 0, protože se nacházejí v rovině vozovky). Chyba naměřených rozměrů v jednotlivých scénách je tak vykreslena v obrázku 5 jako červené sloupce. Stejný graf také ukazuje chybu (14) naší metody na těchto scénách (vyhodnocená oproti více či méně přesným *ground-truth* měřením).

Porovnali jsme naši navrženou metodu s metodou *AutoCalib* – dosud nejlepším alternativním řešením dostupným v literatuře. Autoři nezveřejnili metodu *AutoCalib*, protože jsme re-implementovali jejich algoritmus na základě jejich článku [8]. Použili jsme stejné význačné body jako v naší metodě a sdělili jsme i použité 3D modely. Předpokládáme, že toto značná vylepšení oproti původní metodě *AutoCalib*, protože tam jsou použity pouze pohledy



Obrázek 6: Comparison of accuracy for the proposed and *AutoCalib* method. Average results are as follows: *BrnoCompSpeed* dataset – *AutoCalib* 6.84 %, ours 3.17 %; *BrnoCarPark* dataset – *AutoCalib* 9.38 %, ours 5.68 %; both datasets – *AutoCalib* 7.71 %, ours 4.03 %.

na zadní část vozidel a množina význačných bodů je tak výrazně omezená. Vedle toho autoři metody *AutoCalib* použili jeden jednotný 3D model reprezentující všechna vozidla typu sedan, určením průměrných pozic význačených bodů a tak nerozlišovali mezi jednotlivými modely vozidel.

Srovnání naší metody s metodou *AutoCalib* je vyobrazeno na obrázku 6. Střední RMSE přes všechny scény bylo sníženo z 7,71 % v případě metody *AutoCalib* na 4,03 % naším přístupem. Je třeba poznamenat, že autoři [8] v článku reportují chybu 8,98 % na jejich data; naše implementace se tedy zdá srovnatelná, nebo i o něco lepší než původní řešení (i když je třeba brát v úvahu, že dataset, na kterém je vyhodnocení prováděno, je odlišný).

V našich experimentech byla použita metoda *Differential Evolution*, jejíž parametry byly nastaveny takto: velikost populace (počet rodičů, NP) – 15 krát více než je počet parametrů (75); pravděpodobnost křížení (CR) – 0,9; technika dither byla použita pro nastavení váhového faktoru F a hodnoty jsou náhodně zvoleny pro každou generaci z intervalu $\langle 0,5, 1.0 \rangle$; metoda pro tvorbu pokusných kandidátů je *DE/best/1/bin* (notace a význam všech parametrů je vysvětlena v článku s představením metody [7]).

Literatura

- [1] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 22, 2000.
- [2] X. Meng and Z. Hu, “A new easy camera calibration technique based on circular points,” vol. 36, pp. 1155–1164, May 2003.
- [3] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in *Advances in Neural Information Processing Systems (NIPS)*, 2015.
- [4] R. Juránek, A. Herout, M. Dubská, and P. Zemčík, “Real-time pose estimation piggybacked on object detection,” in *The IEEE International Conference on Computer Vision (ICCV)*, 2015.
- [5] J. Sochor, J. Špaňhel, and A. Herout, “BoxCars: Improving fine-grained recognition of vehicles using 3D bounding boxes in traffic surveillance.” arXiv:1703.00686, 2017.
- [6] Z. Wang, L. Tang, X. Liu, Z. Yao, S. Yi, J. Shao, J. Yan, S. Wang, H. Li, and X. Wang, “Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification,” in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [7] R. Storn and K. Price, “Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces,” *Journal of Global Optimization*, vol. 11, pp. 341–359, Dec 1997.
- [8] R. Bhardwaj, G. K. Tummala, G. Ramalingam, R. Ramjee, and P. Sinha, “AutoCalib: Automatic traffic camera calibration at scale,” in *The 4th ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys 2017)*, November 2017.