



PROJEKT Č. VI20172020068

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

**AUTOMATICKÁ KALIBRACE KAMERY S  
VYUŽITÍM AUTOMATICKY  
DETEKOVANÝCH VÝZNAČNÝCH BODŮ  
NA VOZIDLECH**

**TECHNICKÁ ZPRÁVA**

Vojtěch Bartl, Adam Herout

Vysoké učení Technické v Brně  
Fakulta Informačních Technologií  
Božetěchova 1  
612 66 Brno, Česká republika

Prosinec 2019

## Obsah

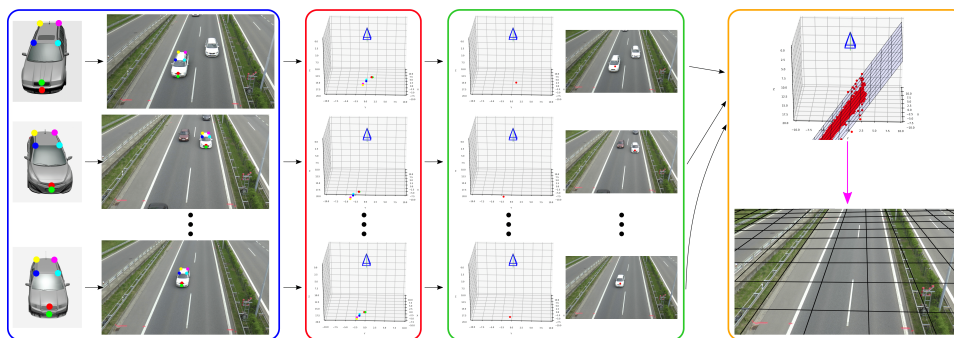
1 Úvod	1
2 Navržený přístup	3
3 Závěr	7

## **Abstrakt**

V této technické zprávě je popsána nově navržená metoda pro automatickou kalibraci kamery na základě pozorování scény a detekce význačných bodů. Díky detekci definovaných význačných bodů na vozidlech a znalosti jejich 3D souřadnic v lokálním souřadném systému objektu lze provést kalibraci kamery. Ta poté slouží k provádění měření ve scéně. Navržená metoda překonává dosud existující přístupy a zároveň klade nižší požadavky na vstupní data.

# 1 Úvod

Kalibrace kamery je aplikovatelnou částí počítačového vidění. Správně zkalibrovaná kamera poskytuje informace potřebné pro provádění měření ve scéně. Pokud tedy máme zkalibrovanou kameru, lze provádět měření např. vzdáleností nebo rychlosti v záznamu z dané kamery. Dříve se tato kalibrace často prováděla manuálně [1], kdy bylo zapotřebí aby technik fyzicky naměřil nějaké vzdálenosti ve scéně a ty se pak využily pro kalibraci kamery [2, 3]. Tento přístup je značně nepraktický a v dnešní době se hledají přístupy založené na automatickém zpracování pomocí počítačového vidění. V některých případech je dokonce nebezpečné se fyzicky dostat do scéně kamery (dálniční dohledová kamera) nebo finančně nákladné (kamera umístěná v místě kam se technik těžko dostane). Proto navrhuje přístup, jak automaticky zkalibrovat kameru jen na základě dlouhodobého pozorování scéně a výskytu automobilů v této scéně. Celkový postup metody je znázorněn na obrázku 1.



Obrázek 1: Náhled navržené metody. Nejprve jsou ve scéně pomocí konvoluční neuronové sítě detekována vozidla [4]. Dále jsou tato vozidla klasifikována na jednotlivé modely [5, 6] a zároveň jsou automaticky detekovány předem zvolené významné body [7] (modře označená část). Pomocí 2D-3D korespondencí je vyřešen algoritmus PnP a tím zjištěna pozice objektu vůči kameře (červeně označená část). Jednotlivá pozorování vozidel poskytují informace o počátku souřadného systému jednotlivých vozidel (zeleně označená část). Více těchto pozorování dává celkové informace o vozovce vůči kameře - tyto informace o vozovce poskytují dostatek informací pro získání kalibračních parametrů kamery (oranžově označená část).

Kalibrace kamery obecně popisuje kde se kamera nachází a jaké má vnitřní parametry (informace o čočce kamery). Pomocí kalibračních parametrů kamery lze promítat body ve 3D souřadnicích  $\mathbf{x} = (x, y, z, 1)^\top$  na jejich 2D pozici na projekční rovině  $\mathbf{x}' = (u, v, 1)^\top$  pomocí rovnice:

$$\lambda \mathbf{x}' = \mathbf{K} [\mathbf{R} | \mathbf{t}] \mathbf{x}. \quad (1)$$

Kalibrační parametry se skládají z matice vnitřních parametrů  $\mathbf{K}$ , rotační

matice  $\mathbf{R}$  a translačního vektorů  $\mathbf{t}$ . Všechny tyto parametry je v procesu kalibrace kamery potřeba odhadnout. Navržená metoda je založena na detekci význačných bodů na vozidlech v obrázku z kamery a znalosti 3D souřadnic těchto význačných bodů v jednotlivých souřadných systémech vozidla. Tyto korespondence 2D-3D pozic slouží k vypočítání algoritmu *PnP* [8, 9, 10, 11, 12], který určuje pozici objektu vůči kameře právě na základě takovýchto 2D-3D korespondencí. Jelikož jsou ale 3D souřadnice v lokálním souřadném systému každého objektu, vypočítané pozice objektů vůči kameře nemají mezi sebou žádnou spojitost. Navržené přístupy řeší právě tento problém a zavádí spojitost mezi jednotlivými pozorováními. Tím lze dosáhnout výsledné kalibrace kamery díky pozorováním vozidel ve scéně. Navržený přístup tak ruší většinu omezení, která měla předchozí řešení automatické kalibrace kamery.

## 2 Navržený přístup

Nejprve je popsán postup jak odvodit vnější parametry kamery se známými hodnotami vnitřních parametrů kamery. Nejprve je vstup transformován na jednotlivé detekce vozidel [13, 14]:

$$\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_N\} \quad (2)$$

a pro každou detekci  $\mathbf{c}_i$  jsou dostupné 2D lokalizace význačných bodů [7]:

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

Pro pozorovaná vozidla  $\mathbf{c}_i$  jsou dostupné 3D souřadnice význačných bodů v lokálním souřadném systému objektu:

$$\hat{\mathcal{K}}^{\mathbf{c}_i} = \{\hat{\mathbf{k}}_1^{\mathbf{c}_i}, \dots, \hat{\mathbf{k}}_K^{\mathbf{c}_i}\}. \quad (4)$$

Detekované 2D ( $\bar{\mathcal{K}}^{\mathbf{c}_i}$ ) a 3D ( $\hat{\mathcal{K}}^{\mathbf{c}_i}$ ) korespondence pro každé vozidlo  $\mathbf{c}_i$  jsou využity pro vyřešení algoritmu  $PnP$  — řešení poskytují vnější parametry kamery (rotační matice  $\mathbf{R}^{\mathbf{c}_i}$  a translační vektor  $\mathbf{t}^{\mathbf{c}_i}$  pro transformaci mezi souřadným systémem objektu a souřadným systémem kamery.

Původní 3D pozice  $\hat{\mathcal{K}}^{\mathbf{c}_i}$  v lokálním souřadném systému objektu tak mohou být transformovány na 3D pozice v souřadném systému kamery:

$$\mathbf{k}_j^{\mathbf{c}_i} = [\mathbf{R}^{\mathbf{c}_i} | \mathbf{t}^{\mathbf{c}_i}] \begin{bmatrix} \hat{\mathbf{k}}_j^{\mathbf{c}_i} \\ 1 \end{bmatrix}. \quad (5)$$

Příklad takové transformace je znázorněn na obrázku 2. Pro každé pozorování tak může být transformován bod z počátku souřadného systému (ležící na vozovce). Z alespoň tří takových pozorování lze sestrojit rovinu, která popisuje vozovku. Takový příklad je ukázán na obrázku 3.

Následně je potřeba vyřešit transformaci ze souřadného systému kamery do světového souřadného systému. Rovina v souřadném systému kamery vlastně popisují parametry kamery. Rovnici popisující rovinu lze napsat jako:

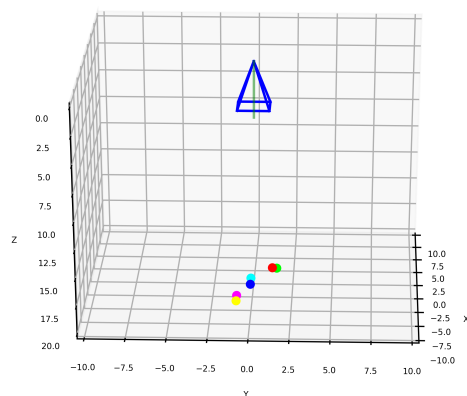
$$ax + by + c = z. \quad (6)$$

Rovinu lze odhadnout jako (metoda nejmenších čtverců):

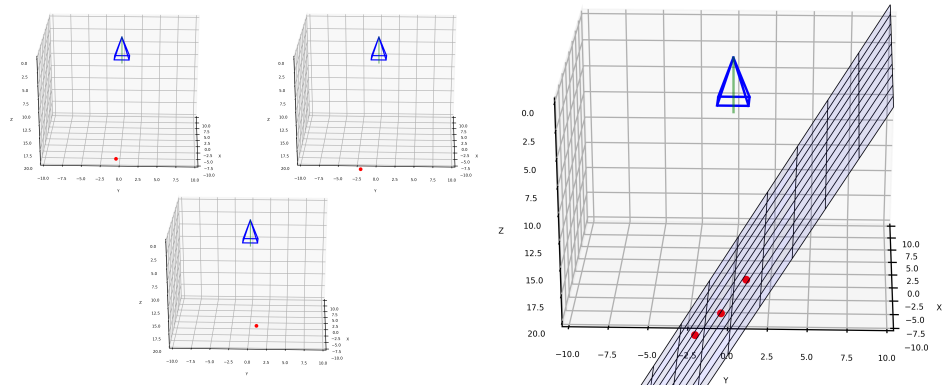
$$\begin{bmatrix} x_0 & y_0 & 1 \\ x_1 & y_1 & 1 \\ & \vdots & \\ x_N & y_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_N \end{bmatrix}, \quad (7)$$

kde  $x_n, y_n, z_n$  jsou 3D souřadnice alespoň tří bodů popisujících počátek souřadného systému vozidel. Snadnější zápis pomocí matic:

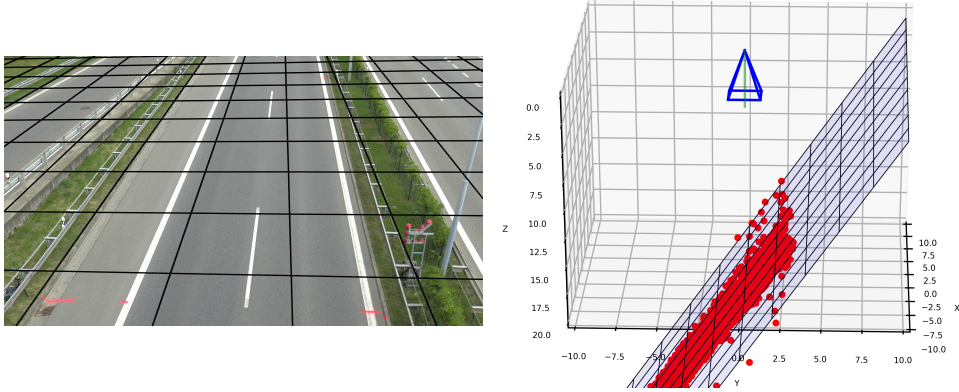
$$\mathbf{A}\mathbf{x} = \mathbf{B}, \quad (8)$$



Obrázek 2: *vlevo*: Detekované 2D význačné body na snímku. *vpravo*: 3D souřadný systém kamery a odpovídající 3D pozice detekovaných význačných bodů.



Obrázek 3: Vypočet roviny v souřadném systému kamery. *vlevo*: Počátky tří různých vozidel transformovaný do souřadného systému kamery. *vpravo*: Rovina v souřadném systému kamery určená pomocí tří (minimální případ) počátků objektů.



Obrázek 4: Ukázka výsledné kalibrace za použití všech pozorovaných vozidel. *vlevo*: Snímek scény s vypočítanou kalibrační mřížkou. *vpravo*: Kalibrační rovina společně s počátečními body jednotlivých pozorování.

kde  $\mathbf{x}$  jsou parametry roviny  $[a, b, c]^T$ . Tyto parametry lze tedy vypočítat:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = A^+ B, \quad (9)$$

kde  $A^+ = (A^T A)^{-1} A^T$  je Moore-Penrose pseudoinverze  $A$ .

Rotační matici  $\mathbf{R}$  lze vypočítat z rotační osy  $\mathbf{u}$  a úhlu  $\theta$ . Rotační úhel  $\theta$  je:

$$\theta = \frac{\arccos(\mathbf{c} \cdot \mathbf{n})}{|\mathbf{c}| \cdot |\mathbf{n}|}. \quad (10)$$

Výsledná rotační matice  $\mathbf{R}$  je vypočítána pomocí *Rodrigues* rovnice:

$$\mathbf{R} = (\cos \theta)I + (\sin \theta)[\mathbf{u}]_{\times} + (1 - \cos \theta)(\mathbf{u} \otimes \mathbf{u}). \quad (11)$$

Dále je potřeba vypočítat ohniskovou vzdálenost, která je vnitřním parametrem kamery. Pro tento výpočet jsou využity 2D detekce význačných bodů ( $\bar{\mathcal{K}}^{c_i}$ ) a jejich 3D pozice ( $\hat{\mathcal{K}}^{c_i}$ ).

Jsou definovány vzdálenosti mezi páry bodů  $a$  a  $b$  jako:

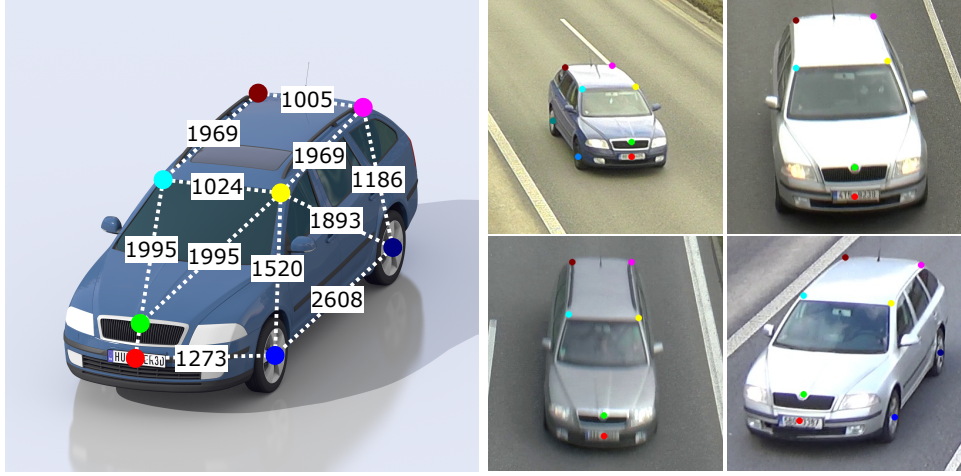
$$\hat{\delta}(\mathbf{c}_i, a, b) = \left| \hat{\mathbf{k}}_a^{c_i}, \hat{\mathbf{k}}_b^{c_i} \right|. \quad (12)$$

Příklad těchto vzdáleností  $\hat{\delta}(\mathbf{c}_i, a, b)$  spolu s detekovanými význačnými body je vidět na obrázku 5.

Odhad ohniskové vzdálenosti vychází z výpočtu vzdáleností naměřených ve scéně, které nejlépe odpovídají těmto známým vzdálenostem. Pokud vycházíme z rovnice (1):

$$\lambda \begin{bmatrix} \bar{\mathbf{k}}_j^{c_i} \\ 1 \end{bmatrix} = \mathbf{K}_{\phi} \left( \mathbf{R}_{\phi} \mathbf{k}_j^{c_i}(\phi) + \mathbf{t}_{\phi} \right), \quad (13)$$





Obrázek 5: *vlevo*: Vzdálenosti  $\hat{\delta}$  (v milimetrech) v 3D modelu *Skoda Octavia mk2*. Je zobrazena jen malá podmnožina všech vzdáleností  $\hat{\mathcal{K}}^{c_i}$ . *vpravo*: Ukázka detekovaných význačných bodů na stejném typu vozidla.

kterou lze přeskádat:

$$\mathbf{R}_\phi^{-1} \mathbf{K}_\phi^{-1} \lambda \begin{bmatrix} \bar{\mathbf{k}}_j^{c_i} \\ 1 \end{bmatrix} = \mathbf{k}_j^{c_i}(\phi) + \mathbf{R}_\phi^{-1} \mathbf{t}_\phi \quad (14)$$

a dále:

$$\mathbf{k}_j^{c_i}(\phi) = \mathbf{R}_\phi^{-1} \left( \mathbf{K}_\phi^{-1} \lambda \begin{bmatrix} \bar{\mathbf{k}}_j^{c_i} \\ 1 \end{bmatrix} - \mathbf{t}_\phi \right). \quad (15)$$

Projekční měřítko  $\lambda$  lze vypočítat z rovnice (14) nastavením  $Z$ -souřadnice ze známého CAD modelu  $\begin{bmatrix} \hat{\mathbf{k}}_j^{c_i} \\ 1 \end{bmatrix}_3$ :

$$\lambda = \frac{\begin{bmatrix} \hat{\mathbf{k}}_j^{c_i} \\ 1 \end{bmatrix}_3 + \begin{bmatrix} \mathbf{R}_\phi^{-1} \mathbf{t}_\phi \\ 0 \end{bmatrix}_3}{\begin{bmatrix} \mathbf{R}_\phi^{-1} \mathbf{K}_\phi^{-1} \begin{bmatrix} \bar{\mathbf{k}}_j^{c_i} \\ 1 \end{bmatrix} \\ 0 \end{bmatrix}_3} \quad (16)$$

Pro každé vozidlo  $c_i$  a každou dvojici viditelných význačných bodů  $a, b$  a jejich pozice ve světovém souřadném systému  $\mathbf{k}_j^{c_i}(\phi)$ , jejich 3D vzdálenost může být spočítána:

$$\delta(c_i, a, b, \phi) = |\mathbf{k}_a^{c_i}(\phi), \mathbf{k}_b^{c_i}(\phi)|. \quad (17)$$

### 3 Závěr

Byl navržen nový postup jak provést automatickou kalibraci libovolné dopravní kamery. Tento postup je založen na detekci 2D význačných bodů na obrazu z kamery a známých 3D souřadnicích těchto bodů ve jejich lokálních souřadných systémech. Aby se dal využít přesný 3D model, je zároveň prováděna klasifikace vozidel dle jejich výrobního modelu. Navržený přístup je tak rozšiřitelný pro nasazení na libovolném místě. Byly provedeny experimenty, které ukázaly zlepšení kalibrace kamery oproti existujícím přístupům. Provedená měření ukázala snížení chyby z 6.88 % na 2.72 %. Navíc bez nutnosti znát vnitřní matici kamery, což požadovaly dosavadní přístupy.

## Reference

- [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,” *Pattern Recognition*, vol. 36, pp. 1155–1164, May 2003.
- [3] Z. Zhang, “Camera calibration with one-dimensional objects,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 892–899, July 2004.
- [4] 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.
- [5] J. Sochor, A. Herout, and J. Havel, “BoxCars: 3D boxes as CNN input for improved fine-grained vehicle recognition,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [6] 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.
- [7] A. Newell, K. Yang, and J. Deng, “Stacked hourglass networks for human pose estimation,” in *European Conference on Computer Vision (ECCV)*, pp. 483–499, Springer, 2016.
- [8] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, pp. 381–395, 06 1981.
- [9] V. Lepetit, F. Moreno-Noguer, and P. Fua, “Epnnp: An accurate  $o(n)$  solution to the pnp problem,” *International Journal of Computer Vision (IJCV)*, vol. 81, pp. 155–166, 2008.
- [10] J. A. Hesch and S. I. Roumeliotis, “A direct least-squares (dls) method for pnp,” in *The IEEE International Conference on Computer Vision (ICCV)*, pp. 383–390, Nov 2011.
- [11] Y. Zheng, Y. Kuang, S. Sugimoto, K. Åström, and M. Okutomi, “Revisiting the pnp problem: A fast, general and optimal solution,” in *The IEEE International Conference on Computer Vision (ICCV)*, pp. 2344–2351, Dec 2013.

- [12] L. Kneip, H. Li, and Y. Seo, “Upnp: An optimal  $o(n)$  solution to the absolute pose problem with universal applicability,” in *Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), (Cham), pp. 127–142, Springer International Publishing, 2014.
- [13] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *The IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, Oct 2017.
- [14] A. Kirillov, R. B. Girshick, K. He, and P. Dollár, “Panoptic feature pyramid networks,” *CoRR*, vol. abs/1901.02446, 2019.