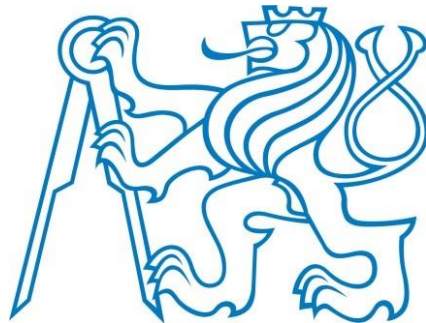


České vysoké učení technické v Praze

Fakulta dopravní

Katedra dopravní telematiky



Studie:

Zvyšování spolehlivosti LPR dat – analýza a algoritmus (návrh a ověření)

Zodpovědný řešitel:

Ing. Eva Hajčiarová

Řešitelský tým:

Ing. Kristýna Navrátilová

2024

Podpořeno z:

161 - 1612408F000

Obsah

Obsah.....	2
1 Úvod	4
1.1 Vymezení použitých pojmů a zkratk	5
2 LPR data	6
3 Metody sběru LPR dat	9
4 Analýza dostupných technologií sběru a zpracování LPR dat	11
4.1 Přístupové a parkovací systémy.....	11
4.2 Sběr a zpracování LPR dat na profilu komunikace	12
4.3 Kamery s funkcí LPR.....	16
4.4 Softwarové řešení	17
5 Zpracování LPR dat a otázky kvality dat.....	19
6 Vymezení oblasti výzkumu	23
7 Příprava referenční datové sady	24
7.1 Zdroj dat	24
7.2 SW rozpoznání RZ ze zdrojových videozáznamů	25
7.3 Manuální kontrola datových sad vůči referenčním videozáznamům, oprava a doplnění	27
7.4 Spárování dat dle Metodiky vyhodnocení směrových průzkumů	31
7.5 Opětovná kontrola spárovaných dat dle referenčních videozáznamů, oprava a doplnění ...	32
7.6 Výsledná referenční datová sada.....	Chyba! Záložka není definována.
8 Analýza SW spolehlivostních parametrů	34
8.1 Popis datové sady pro vyhodnocení spolehlivostních parametrů.....	34
8.2 Zobrazení datové sady pro vyhodnocení spolehlivostních parametrů.....	35
8.3 Vyhodnocení spolehlivostních parametrů.....	41
9 Využití pravděpodobnostní matice	45
9.1 Popis datové sady pro sestavení pravděpodobnostní matice	45
9.2 Matice $p(y x)$ – sestavení a využití	46
9.3 Stanovení meze M a spolehlivosti jejího odhadu, její význam a využití	48
10 Algoritmus	52
10.1 Kroky algoritmu a jejich popis:	53
11 Ověření z dat z Jenerálky dle algoritmu.....	57
11.1 Vstupy využité pro ověření	57
11.2 Průběh ověření	60
11.3 Výsledky ověření.....	61
12 Budoucí rozvoj	64

13	Závěr	65
14	Seznam obrázků.....	66
	Seznam tabulek	67
	Seznam grafů	68
	Reference.....	69

1 Úvod

V současné době pokračuje trend stále narůstající poptávka společnosti po dopravních a přepravních možnostech, kdy doprava představuje samozřejmou součást každodenního života. Řízení a management silniční dopravy hraje spolu s kvalitní infrastrukturou silniční sítě vzhledem k narůstajícímu stupni automobilismu stále důležitější roli, nejen v celkovém kontextu, ale zejména v prostředí měst. Dopravní telematika zaznamenává v posledních desetiletích neustálý vývoj, který je podmíněn pokračujícím technologickým rozvojem. Cílem telematických systémů je zefektivnit řízení dopravy, snížit dopravní zatížení silniční sítě, zkrátit dojezdové doby, zvýšit bezpečnost a plynulost provozu a obecně minimalizovat negativní dopady dopravy celospolečenského charakteru. Nutným předpokladem pro umožnění využití plného potenciálu systémů řízení a managementu dopravy je zajištění kvalitních vstupních dopravně-inženýrských dat, která budou poskytovat dostatečně přesné informace o reálném stavu dopravy, využitelné pro predikce, modely dopravy a adekvátní reakce algoritmů řídicích systému.

Tato studie vychází z potřeb efektivního a spolehlivého předzpracování surových dopravních dat, aby z nich získané informace mohly být dále efektivně využívány. V současné době je na silniční síti instalována celá řada dopravních detektorů a senzorů využívajících různé technologické principy pro sběr dat. Ukazuje se ovšem, že reálná kvalita dopravních dat poskytovaných těmito senzory nemusí odpovídat jejich spolehlivosti deklarované výrobcí. Ke snížení kvality dat poskytovaných senzory v praxi vedou různé faktory jako nedostatečná údržba senzorů a souvisejících technologií, neefektivní správa dat nebo různé faktory vyplývající z prostředí, kdy např. zhoršené povětrnostní podmínky nebo specifické podmínky provozu v konkrétním místě silniční sítě mohou negativně ovlivňovat fungování detektorů. Toto vše vede k potřebě podrobněji se zabývat data managementem a vyhodnocováním kvality dat.

V této studii se autoři zabývají specificky daty ze systémů rozpoznání registračních značek, tzv. LPR daty (license plate recognition). Jedná se o široce dostupný datový zdroj, neboť každé vozidlo vozí viditelně umístěnou jedinečnou registrační značku, jehož využití má dlouholetou tradici a stále skýtá značný potenciál pro řadu dopravních aplikací a telematických systémů. Systémy automatizované detekce a rozpoznání registračních značek s využitím kamer a analýzy obrazu zažívají neustálý technologický vývoj a pokrok, který vede ke stále přesnějším a spolehlivějším výsledkům. Přesto v praxi jsou LPR data stále často zatížena chybami, ať už nepřesně rozpoznávanými texty registračních značek nebo zcela chybějícími záznamy průjezdů vozidel. Autoři se proto zaměřují na fázi předzpracování naměřených surových LPR dat, s cílem minimalizovat vliv uvedených chyb na výsledná zpracovaná a vyhodnocená data, aby byly zajištěny co nejpřesnější informace, které mohou být dále využity pro nejrůznější aplikace řízení a managementu dopravy, ať už v krátkodobém nebo dlouhodobém horizontu.

Hlavním přínosem této studie je navržený algoritmus pro efektivní přípravu LPR dat, který minimalizuje vliv chyb v surových LPR datech, konkrétně umožňuje automatizované spárování chybně rozpoznávaných textů registračních značek naměřených na různých měřicích profilech. Toto je zásadní krok pro dosažení přesné rekonstrukce tras vozidel po zkoumané oblasti, stanovení směrovosti dopravy a např. určení podílu tranzitní dopravy vůči zdrojové a cílové dopravě.

Praktická část prací, kterou popisují kapitoly 8–11 je založena na důkladné analýze současných metod sběru a zpracování LPR dat, na analýze zdrojů chyb a možností jejich eliminace. Analýza byla provedena v oblasti současných odborných výzkumných článků a literatury, kde byly zkoumány zavedené postupy i nové a inovativní výzkumné směry a trendy. Byla provedena rovněž detailní analýza současného stavu technologií dostupných na trhu pro sběr LPR dat – tedy kamer a SW aplikací pro rozpoznání registračních značek. Analytickou část popisují kapitoly 2-7. Navržený algoritmus byl modelově

implementován na datech z vybrané oblasti silniční sítě (data byla získána formou směrového průzkumu). Bylo provedeno ověření algoritmu a stanovení procentuálního zpřesnění spárovaných dat. Popisem implementace a ověření algoritmu se zabývá kapitola 11.

1.1 Vymezení použitých pojmů a zkratek

AI	umělá inteligence (artificial intelligence)
ANPR	automatic number plate recognition
LPR	license plate recognition
OCR	optical character recognition (technologie pro rozpoznávání znaků z obrazu)
RZ	registrační značka
SW	software

2 LPR data

LPR data (license plate recognition data) jsou dopravní data získaná záznamem registračních značek vozidel. Jedná se o historicky ověřený a využívaný zdroj dopravních dat, mezi jehož hlavní přednosti patří široká dostupnost, neboť každé vozidlo má povinnost mít viditelně umístěnou standardizovanou registrační značku. Další předností je možnost jednoznačně identifikovat konkrétní vozidla v konkrétním místě, kde probíhá měření. Tyto faktory zajišťují potenciál velmi přesných informací o stavu dopravy, které lze využít pro management dopravy (1), (2), (3). S pomocí LPR dat je možné například zpětně rekonstruovat trasy a směrovost vozidel ve zkoumané oblasti, stanovovat matice zdrojů a cílů cest, doby jízdy, úsekové rychlosti, dojezdové doby. Všechny tyto dopravně-inženýrské informace jsou získávány z měřicích profilů umístěných ve vybraných řezech komunikací za plného provozu vozidel. (4) Poněkud odlišnou, ale přesto širokou oblast uplatnění LPR dat představují parkovací systémy, které však využívají záznam registračních značek stojících vozidel nebo vozidel vjíždějících na / vyjíždějících z parkovišť. (5) Výzkum popisovaný v tomto dokumentu se nicméně detailněji zaměřuje na spolehlivost LPR dat získaných za provozu vozidel, nikoli ve statickém prostředí, protože právě rychlost jízdy vozidel a horší podmínky umístění kamer při záznamu jedoucích vozidel na komunikacích představují některé ze zásadních faktorů, které přesnost získaných LPR dat snižují.

Dopravně-inženýrské informace získané z LPR dat mají potenciál širokého využití pro on-line i off-line řízení a management dopravy. V podmínkách České republiky i celosvětově se v současné době LPR data běžně využívají pro určování průměrné rychlosti vozidla v měřeném úseku, což je využíváno pro kontrolu dodržování povolené rychlosti a případnou penalizaci. Využívají se i v dalších penalizačních systémech, např. při sledování jízdy na červenou. Další oblast představují například mýtné nebo výše zmíněné parkovací systémy. Oblasti možného využití LPR dat shrnují např. (4), (6). LPR data a dopravně inženýrské informace z nich získané jsou využívány rovněž jako podklad pro strategické rozhodování a plánování v oblasti silniční dopravy, plánování úprav komunikační sítě, organizace dopravy, tvorbu dopravních generelů nebo např. zjišťování dopadů dopravy na životní prostředí. V této oblasti jsou v podmínkách ČR využívány zejména krátkodobé (jednorázové) dopravní průzkumy (7), (8), (9).

Celosvětově se možnostmi využití LPR dat pro řízení dopravy zabývá řada výzkumů a vědeckých studií. Příklady výzkumu v této oblasti jsou následující: zdroj (3) zkoumá využití LPR dat pro optimalizaci řízení dopravy. (10) se zabývá výzkumem směrovosti, specificky vzorci dojíždějících vozidel, (11) pak zpětnou rekonstrukcí tras vozidel. (12) přináší přehled různých směrů výzkumu LPR dat, mezi nimi se například zabývá vztahy zdrojů a cílů cest, jízdami dobami nebo dojezdovými dobami.

Nicméně, přes všechny uvedené přínosy a možnosti využití není v dnešní době potenciál LPR dat plně využit, a to přesto, že např. informace o trasách konkrétních vozidel, směrovosti dopravy v území a podílu tranzitní dopravy je v současnosti poměrně obtížné získat z ostatních běžně využívaných zdrojů dopravních dat. Možností jsou např. FCD data, ale pouze v úsecích, pro která jsou tato data k dispozici. (13) Dále např. data od mobilních operátorů, která ale vykazují omezenou přesnost lokalizace na silniční síti kvůli závislosti na BTS buňkách (14).

Nedostatečné využití potenciálu LPR dat v praxi je způsobeno několika faktory. Možné způsoby využití jsou často omezeny poměrně řídkým pokrytím silniční sítě příslušnými senzory (kamerami). Jednotlivé aplikace, které využívají LPR data, v takovém případě představují spíše dílčí izolované systémy, které jsou navrhovány a využívány pro konkrétní funkce. Řada článků dokládá využitelnost pro analýzu dopravy v makroskopickém měřítku rozsáhlejší oblasti, např.: (1), (10). Nicméně využití pro detailnější analýzy dopravy v malém území je možné pouze s mnohem hustším pokrytím silniční sítě (2), (3).

Dalším z faktorů majících vliv na využití LPR dat je nutnost zajištění ochrany osobních údajů, což je celosvětovým tématem (15), (16) a v Evropské unii je ztělesněno v nařízení GDPR (17). Dalším faktorem je otázka kvality zdrojů LPR dat, kdy přes současný technologický pokrok není vždy vzhledem k reálně využívaným technologiím v provozu, k finančním možnostem apod. dosahováno uspokojivé přesnosti, kompletnosti a tedy věrohodnosti dat. Vědecké zdroje se k této otázce staví dvojím způsobem – některé zdroje pracující s LPR daty automaticky předpokládají jejich přesnost a kompletnost a otázky kvality se nevěnují (nebo pouze velmi okrajově) (1), (18), protože pro jejich účel výzkumu není reálná kvalita zdrojů LPR dat podstatná. Naopak jiné vědecké zdroje přiznávají sníženou reálnou kvalitu LPR dat (3), (19), (20).

V následujících kapitolách je problematika sběru a zpracování LPR dat podrobněji rozvedena, včetně popisu limitujících faktorů a vlivů na kvalitu dat a možných směrů pro zvýšení kvality výstupů z tohoto datového zdroje.

Vymezení pojmů

Studie pracuje s některými pojmy, které nemají v odborné komunitě ve vztahu k LPR datům ustálenou definici. Jednoznačné vymezení těchto pojmů pro účely této práce je uvedeno zde:

Měřicí profil

Měřicím profilem se rozumí virtuální řez pozemní komunikací, kde probíhá měření LPR dat.

Surová data

Jedná se o LPR data v elektronické podobě ve formátu, který je snadno softwarově zpracovatelný, s využitím tabulkového procesoru a podobných aplikací. Příklady možného formátu dat jsou: .csv, .xls, .txt. Surová data jsou neanonymizovaná a musí obsahovat všechny důležité informace pro místně a časově jednoznačnou identifikaci průjezdu vozidla a pro následné zpracování a vyhodnocení. Jedná se o ID záznamu, čas průjezdu vozidla (s přesností na sekundy), RZ vozidla (v plném tvaru), kategorii vozidla a identifikaci měřicího profilu, ve kterém byl záznam pořízen.

Možnou podobu surových dat zobrazuje následující tabulka.

Tabulka 1 Struktura surových dat; zdroj: (DP), upraveno

ID	Čas průjezdu	RZ	Typ vozidla	Měřicí profil
1	čas 1 (HH:MM:SS)	RZ 1	typ vozidla 1	ID měřicího profilu
2	čas 2 (HH:MM:SS)	RZ 2	typ vozidla 2	ID měřicího profilu
3	čas 3 (HH:MM:SS)	RZ 3	typ vozidla 3	ID měřicího profilu
4	čas 4 (HH:MM:SS)	RZ 4	typ vozidla 4	ID měřicího profilu
...

Sběr a zpracování dat

Pod pojem „sběr dat“ jsou zahrnuty všechny procesy a činnosti, které vedou k získání surových dat v podobě definované výše. Jde tedy především o samotný záznam registračních značek v měřicích profilech na silniční síti. Případně jde rovněž o off-line dodatečné rozpoznání registračních značek z videozáznamu nebo o převod zaznamenaných registračních značek z jiných médií a datových formátů do elektronického datového formátu umožňujících SW zpracování v tabulkovém procesoru (viz

specifikováno výše). Pro zpracování videozáznamů za účelem rozpoznání obrazu se z logiky věci využívá rovněž pojem „zpracování dat“

Pojmem „zpracování dat“ se však v této Studii rozumí zpracování surových dat do podoby, ze které je možné přímo vyhodnocovat dopravně-inženýrské charakteristiky dopravy a následně pak interpretovat získané informace a dále je využívat pro on-line i off-line aplikace řízení a modelování dopravy. Vstup do zpracování dat tvoří surová data v podobě definované výše. Samotné zpracování dat probíhá elektronicky, v ideálním případě automatizovaně s minimální nutností zásahu lidské obsluhy. Zahrnuje především činnosti a procesy týkající se tzv. párování dat, sestavení cest vozidel, anonymizace dat a řešení kvality dat.

Kvalita dat

Pro účel této studie byly pojmy související s kvalitou dat definovány v souladu s (21). Kvalita dat je chápána jako pojem zastřešující několik dimenzí a určující, nakolik datová sada odpovídá reálnému stavu. Tento výzkum pracuje zejména s dimenzemi kompletnosti a přesnosti. Kompletnost se zabývá úplností datové sady ve srovnání s realitou, tedy, zda některá data v datové sadě zcela nechybí. Přesnost se zabývá chybami a nejasnostmi v existujících (tedy nechybějících) datech.

3 Metody sběru LPR dat

Obecně, předpokladem úspěšného využití LPR dat je, že jsou pro záznam registračních značek využity takové postupy a záznamové technologie, které poskytnou dostatečně spolehlivé výsledky pro účel využití dat. Tato kapitola uvádí základní přehled možných metod sběru LPR dat, spolu s jejich základními výhodami a nevýhodami a vlivem na spolehlivost dat.

Na problematiku sběru LPR dat lze pohlížet z několika různých pohledů, zde bude nejdříve představen sběr dat za běžného provozu vozidel na silniční síti, následně bude krátce uvedena oblast statické dopravy, tedy parkovacích systémů.

Sběr LPR dat za provozu vozidel

Záznam registračních značek jedoucích vozidel je využíván v intravilánu i extravilánu a použité metody sběru dat by měly být natolik robustní, aby poskytovaly spolehlivé záznamy všech vozidel projíždějících příslušnými měřicími profily. Měřicím profilem se rozumí virtuální řez pozemní komunikací, kde probíhá měření. Rozmístění měřicích profilů na silniční síti a jejich počet jsou důležité faktory z pohledu následného zpracování a možností vyhodnocení dat. Znalost rozmístění měřicích profilů vůči reálné topologii silniční sítě je zásadní i pro vyhodnocení spolehlivosti dat. Samotný sběr LPR dat může být prováděn kontinuálně nebo jednorázově (případně periodicky).

V případě kontinuálního sběru dat dochází k dlouhodobému záznamu registračních značek v trvale instalovaných měřicích profilech. (22) Sběr dat probíhá celodenně po celou dobu existence měřicího profilu. Vzhledem k této povaze sběru dat jsou využívány automatizované metody měření pomocí detektorů dlouhodobě instalovaných na infrastruktuře. V České republice je tento způsob využíván zejména u dílčích (zejména penalizačních) samostatně fungujících systémů, jak již bylo popsáno v předchozí kapitole. Data z těchto systémů standardně nejsou veřejně dostupná pro další využití mimo tyto systémy, důvody jsou např. GDPR, technologická uzavřenost jednotlivých systémů nebo smluvní nastavení vztahů mezi zainteresovanými stranami.

Vzhledem k nedostatečné dostupnosti kontinuálně sbíraných LPR dat ze sítě detektorů v praxi dochází rovněž k jednorázovým nebo krátkodobě periodicky opakovaným měřením, formou dopravních průzkumů. Tato měření jsou specificky zaměřená na určitou konkrétní zájmovou lokalitu a LPR data jsou sbírána za konkrétním účelem vyhodnocení, zejména s důrazem na získání informací o směrovosti vozidel. (23)

Co se týče konkrétních metod sběru LPR dat, v současnosti je nejběžněji využívána technologie videodetekce, která spočívá ve snímání obrazového záznamu kamerami a pomocí technik softwarového zpracování obrazu dochází k rozpoznání registračních značek zaznamenaných vozidel. (24) Podrobněji je technologie videodetekce rozvedena v následující kapitole. Historicky starším způsobem sběru LPR dat, který se uplatňoval u dopravních průzkumů, je manuální záznam RZ proškolenými pracovníky. (23) S rostoucím počtem vozidel na silnicích a s rozvojem technologií rozpoznání obrazu však tento způsob přestává být dostatečně spolehlivý, a to včetně varianty hlasového záznamu na diktafon. Dnes se již využívá minimálně, přestože ho aktuální Metodika pro provádění a vyhodnocení směrových průzkumů automobilové dopravy stále umožňuje (23). Převažuje záznam RZ vozidel pomocí kamerové technologie převažuje, a to i při jednorázových měřeních.

Sběr LPR dat v oblasti statické dopravy

Do oblasti statické dopravy neboli dopravy v klidu patří parkovací systémy. Jak již bylo zmíněno v předchozí kapitole, v této oblasti se záznam registračních značek již v současnosti využívá poměrně standardně, a to nejen v podmínkách České republiky. Registrační značka je využívána jako jednoznačný identifikátor vozidla v souvislosti s platbou parkovacího poplatku, případně v souvislosti s povolením vjezdu vozidla na parkoviště (25). V případě parkování v uličním prostoru řidič vozidla standardně zadá RZ do systému prostřednictvím parkovacího automatu nebo mobilní aplikace a RZ je tímto automaticky vložena do databáze vozidel s platným parkovacím oprávněním. Pro kontrolu legálnosti parkování je v tomto případě využíván kontrolní záznam registračních značek, často pomocí kontrolního vozidla s kamerovým záznamem. (5) V případě parkování na uzavřených parkovištích je záznam registračních značek využíván při vjezdu/výjezdu vozidla z parkoviště, RZ je detekována pomocí kamerové technologie (25). LPR data zde slouží primárně opět pro kontrolu oprávněnosti parkování, dále však mohou být využita pro stanovení obsazenosti parkovišť, případně obrátkovosti. V této studii se oblastí statické dopravy dále zabýváme pouze velmi okrajově, a to z důvodu odlišných podmínek, za kterých sběr dat probíhá (stojící vozidla) i z důvodu odlišných vyhodnocovaných dopravně-inženýrských veličin.

V dalším textu se Studie bude podrobněji zabývat technologiemi automatizovaného sběru dat, zejména pomocí videozáznamu a následného softwarového zpracování obrazu, protože se v současnosti jedná o nejrozšířenější metodu sběru LPR dat za provozu vozidel. Dále pak se Studie bude zabývat přístupy k řešení kvality LPR dat.

4 Analýza dostupných technologií sběru a zpracování LPR dat

V současné době existuje na trhu celá škála kamer poskytujících obrazový záznam různé kvality, od levných nahrávacích zařízení určených spíše pro laické potřeby široké veřejnosti nesouvisející s dopravním zaměřením, přes dohledové a bezpečnostní průmyslové kamery po kamery se specifickými funkcemi, určené přímo pro poskytování dostatečně kvalitního záznamu pro rozpoznávání registračních značek při běžných rychlostech jízdy vozidel. Výzkum za účelem získání spolehlivých LPR dat se v souvislosti s využívanou kamerovou technologií v současnosti ubírá dvěma základními směry. Prvním je zvyšování kvality samotných nahrávacích zařízení (kamer), aby docházelo ke sběru dostatečně kvalitního videozáznamu pro eliminaci chyb při rozpoznávání registračních značek. Tento přístup však v praxi souvisí s (řádově) vyššími finančními náklady na pořízení kamer. Druhým směrem je vývoj robustních SW algoritmů pro detekci a rozpoznání registračních značek z méně kvalitního videozáznamu. Tento přístup předpokládá, že zdrojem videozáznamu budou levnější varianty záznamových zařízení, např. dohledové kamery běžně instalované ve městech, které obvykle nemají nevhodnější úhel záběru.

V současné době oba výzkumné směry již dosáhly značných pokroků a řada zdrojů vykazuje úspěšnost rozpoznání registračních značek přes 95 % (při vhodných povětrnostních a viditelnostních podmínkách se udává přesnost nabízených řešení kolem 98 %). Zejména v případě komerčních produktů však spolehlivost deklarovaná výrobcem nemusí odpovídat reálné spolehlivosti rozpoznání registračních značek v běžném provozu, protože deklarovaná spolehlivost bývá určena za specifických testovacích podmínek. Jeden ze zdrojů se věnuje testování podmínek, za kterých dojde či již nedojde k rozpoznání jednotlivých znaků RZ. Ze vzdálenosti necelých 2 m jsou jednotlivé znaky rozpoznatelné z bočního úhlu do cca 70°. Dalším testováním bylo rozlišení obrazu, kdy byly znaky rozpoznány do rozlišení 30 pixelů. Rozpoznání za nižšího množství osvětlení až tmy i mezi velkým množstvím polepů v blízkosti RZ se ukázalo být velmi spolehlivé. (26), (27), (28), (29)

Kamera je v dnešní době již vnímána spíše jako senzor, kterým je možné získávat více typů dopravních dat najednou. Tato skutečnost spolu s možností instalace bez narušení infrastruktury, snadnou manipulací a dostupnou cenou z kamer činí stále využívanější technologii nejen v oblasti dopravy / mobility, ale také např. bezpečnosti.

Tato kapitola rozebírá technologie, které jsou aktuálně na trhu dostupné a přináší přehledné shrnutí zjištěných poznatků. Rozsáhlým průzkumem trhu bylo zjištěno, že většina nabízených řešení pro sběr a následné zpracování LPR dat je velmi podobná, má obdobné parametry a způsoby využití. Obvykle platí, čím vyšší požadovaná přesnost, tím vyšší pořizovací náklady. Kapitola je rozdělena do třech větších celků, prvním jsou přístupové nebo parkovací systémy, druhou skupinu tvoří rozpoznávání RZ na profilech komunikací a poslední část popisuje softwarové řešení. Metoda optického rozpoznávání znaků (OCR) je technologie využívající umělou inteligenci (AI), která mimo jiné umožňuje identifikaci čísel a textu na RZ vozidel. Další možností je využívání metod (algoritmů) hlubokého učení. (30), (31)

4.1 Přístupové a parkovací systémy

Výrobci a dodavatelé na svých stránkách většinou blíže popisují způsob sběru dat u nabízených přístupových (někdy nazývaných též vjezdové) či parkovacích systémů než u jiných aplikace využívající LPR. Z technologického hlediska lze označit čtení RZ v tomto případě za snazší řešení, než je tomu v běžném provozu na silničních komunikacích. Hlavními důvody jsou zejména: definovaná oblast, kde

se bude s největší pravděpodobností RZ nacházet, nízké rychlosti nebo rozpoznávání znaků při úplném zastavení vozidla a lepší světelné podmínky.

Moderní management parkování na parkovacích plochách s omezenou možností vjezdu a výjezdu založený na rozpoznávání RZ snižuje délku zdržení na vjezdu či výjezdu a přispívá k vyšší efektivitě i bezpečnosti. Obecně výrobci nabízí řešení pro: areály s omezeným přístupem, firemní parkovací garáže, parkoviště s řízeným vjezdem a výjezdem (např. záchytná parkoviště), garáže obchodních center či parkovací domy. Mnoho řešení nabízí zároveň správu několika desítek až stovek uložených záznamů zaměstnanců, kterým je vjezd umožněn. U obchodních center či jiných veřejně dostupných parkovišť je načtená RZ propojena s parkovacím lístkem a u výjezdu po pozitivním výsledku kontroly platnosti automaticky umožní odjezd danému vozidlu. Zároveň bývá zveřejňován údaj o aktuálním obsazenosti dané parkovací plochy / volné kapacitě, kterou systém získává automaticky díky zadanému maximálnímu počtu vozidel a počítání aktuálně parkujících (na vjezdu a výjezdu). U těchto řešení se zpravidla předpokládá rychlost vozidel kolem 20 km/h, některé nabízené kamery umožňují čtení znaků až do 50 km/h (někdy uvádí výrobci i 70 km/h), rozpoznání jednotlivých znaků trvá řádově několik setin, max. desetiny sekundy. Standardem je čtení RZ z evropských států i dalších zemí, některé řešení nabízejí více než 160 států světa. (30), (32), (33), (34), (35), (36), (37), (38)

4.2 Sběr a zpracování LPR dat na profilu komunikace

Rozpoznávání RZ u vozidel v pohybu se nejčastěji využívá na příčném řezu komunikací, kde je snazší nasměrovat záběr kamery na oblast, kde se RZ vozidel zpravidla nachází a pravděpodobně bude nacházet. Může se jednat o systémy, které mapují dráhu vozidla a určují směrovost, nebo o systémy měřící rychlost vozidla, ať už z informačních důvodů nebo radary přímo napojené na penalizační systém.

Dříve se jednalo především o serverová řešení v podobě fyzického serveru umístěného např. v zabezpečené části budovy, kam byl záběr z kamery (sekvence snímků) přenášen do datového centra a následně zpracován. Zmenšující se velikost procesorů kamer a čím dál výkonnější výpočetní jednotky umožnily zpracování obrazu přímo v místě pořízení. A to buď v samotné kameře nebo pomocí výpočetní jednotky umístěné např. v technologickém rozvaděči poblíž kamery. Nevýhodou tohoto řešení jsou vyšší nároky na nosnost a prostor na objektu, kde bude kamera instalována, např. na stožáru veřejného osvětlení. Toto řešení je výhodné také z hlediska ochrany osobních údajů, jelikož citlivé záznamy neopustí místo jejich pořízení a přenášena jsou pouze agregovaná data. Stále se však uvádí, že přesnost rozpoznávání může být při využití serveru nebo externí výpočetní jednotky vyšší než při zpracování dat přímo v kameře. (39), (40), (41), (42)

Existují dva základní způsoby rozpoznání RZ v obrazu, a to buď přímo v kameře (Edge Computing) nebo přenos záznamu na vzdálený server (Cloud/Server-Based Computing), níže jsou uvedeny jejich charakteristiky vč. hlavních výhod a nevýhod obou řešení.

Zpracování obrazu a rozpoznání RZ přímo v kameře (Edge Computing)

Kamera obsahuje vlastní integrovaný SW nebo čip, který umožňuje zpracování obrazu přímo na místě. Výstupem rozpoznání RZ mohou být:

- **Data:** Textové informace o rozpoznávaných RZ (např. "ABC1234").
- **Obrazové výstupy:** Fotografie vozidla, často se zvýrazněnou RZ.
- **Metadata:** Další informace, jako je čas, datum, místo nebo rychlost vozidla.

Napojení kamery na server s vyhodnocovacím SW (Cloud / Server-Based Computing)

Kamera je připojena k centrálnímu serveru (nebo serverovému clusteru), kde probíhá zpracování obrazu a rozpoznání RZ. Server obsahuje software pro analýzu obrazu (např. OpenALPR, PlateRecognizer aj.). Kamera pak standardně přenáší:

- **Videostream:** Reálný přenos obrazu ve vysokém rozlišení.
- **Snímky:** V některých případech mohou kamery přenášet snímky spouštěné např. pohybem.

Tabulka 2 Výhody a nevýhody obou popsaných řešení

Zpracování obrazu a rozpoznání RZ přímo v kameře (Edge Computing)	
<p>Výhody:</p> <ul style="list-style-type: none"> • Nižší požadavky na přenosovou kapacitu: Není nutné přenášet kompletní videostream, pouze výsledná data a obrázky. • Menší latence: Data jsou zpracována okamžitě v kameře. • Decentralizace: Kamerový systém je soběstačnější a méně závislý na výkonném centrálním serveru. • Škálovatelnost: S rostoucím počtem kamer není nutné navyšovat výpočetní kapacitu serveru. 	<p>Nevýhody:</p> <ul style="list-style-type: none"> • Omezený výkon: Rozpoznávání může být méně přesné nebo pomalejší. • Komplexnější údržba: Aktualizace softwaru na jednotlivých kamerách může být složitější. • Vyšší cena kamer: Kamery s integrovaným zpracováním obrazu jsou dražší.
Napojení kamery na server s vyhodnocovacím SW (Cloud / Server-Based Computing)	
<p>Výhody:</p> <ul style="list-style-type: none"> • Vysoký výkon a flexibilita: Dostatečný výpočetní výkon serverů pro komplexní analýzu, včetně rozpoznání více typů RZ nebo pokročilých algoritmů. • Snadná aktualizace: Software na serveru lze aktualizovat centrálně a jednoduše. • Využití běžných kamer: Lze využít levnější kamery. 	<p>Nevýhody:</p> <ul style="list-style-type: none"> • Vyšší nároky na přenosovou kapacitu: Přenos videa nebo snímků v reálném čase vyžaduje vysokorychlostní síťové připojení. • Latence: Zpracování může být pomalejší z důvodu přenosu a následného zpracování na centrálním serveru. • Závislost na konektivitě: V případě výpadku nedojde k přenosu dat.

Oba uvedené přístupy mohou být kombinovány dle specifických požadavků, může se jednat např. o:

- **Hybridní model:** Kamery zpracovávají data lokálně, ale v případě potřeby přenášet videostream na server pro pokročilejší analýzu.
- **Distribuované systémy:** Využití edge zařízení pro primární detekci a serveru pro pokročilou analýzu nebo ukládání dat.

Volba přístupu závisí na mnoha faktorech, jako jsou finanční náklady, požadavky na přesnost, velikost systému, dostupná infrastruktura (např. síťová kapacita, pokrytí signálem), účel využití (např.

monitoring dopravy, parkovací systémy, bezpečnost) a doba, po kterou bude zařízení na daném místě provozováno.

Kromě standardních dodavatelských řešení, kdy je LPR funkcí kamery a které jsou popsány v samostatné dílčí podkapitole dále v dokumentu, existují na trhu i finančně méně náročné možnosti, které nabízí především zpracování záznamu z mobilního telefonu či jiného zařízení. Níže jsou uvedeny konkrétní příklady řešení dohledané na webu (<https://www.google.com/>). (43), (44), (45)

Mobilní aplikace s integrovaným rozpoznáváním RZ

Existují mobilní aplikace s funkcí rozpoznávání RZ přímo v zařízení. Např. aplikace **LprVerifier** společnosti MACH SYSTEMS s.r.o. nabízí umožňuje automatizované ověření RZ v terénu pomocí fotoaparátu mobilního telefonu nebo tabletu. Aplikace vyfotí RZ vozidla, automaticky ji rozpozná a zobrazí uživateli informace načtené ze serverové databáze.

ALPR Apps (např. Plate Recognizer Mobile) jsou mobilní aplikace umožňující rychlé rozpoznávání RZ přímo na zařízení. Některé vyžadují připojení k internetu, některé fungují i off-line.

OCR-based Apps jsou aplikace využívající optické rozpoznávání znaků pro extrakci textu z obrázků a videí.

Výhody:

- Rychlá a jednoduchá implementace.
- Není nutný přenos dat do cloudu (u offline aplikací).

Nevýhody:

- Omezený výkon při zpracování rozsáhlých záznamů.
- Přesnost může být ovlivněna kvalitou obrazu z mobilu.

Cloudové služby pro analýzu záznamů

Některé společnosti nabízejí cloudové služby, které umožňují vzdálené monitorování videa a rozpoznávání RZ. Lze je využít na analýzu a rozpoznání RZ v záznamech z mobilního telefonu a podobných zařízení typu běžných fotoaparátů a kamer. Např. **AxxonNet** je bezplatná cloudová služba pro správu lokálních VMS (Open Virtual Memory System), která umožňuje pohodlné a bezpečné vzdálené monitorování videa a centralizovanou správu sledovacích seznamů pro rozpoznávání obličejů a RZ v reálném čase.

Plate Recognizer (API a web) nabízí jednoduché API pro nahrání videí nebo obrázků. Podporuje různé formáty RZ z celého světa.

OpenALPR Cloud je platforma pro zpracování a rozpoznávání RZ na serverech poskytovatele, která disponuje vysokou přesností díky pokročilým modelům AI.

Google Vision AI nebo AWS Recognition zahrnují obecné služby pro analýzu obrazu, které mohou být použity i pro OCR a rozpoznávání RZ.

Výhody:

- Škálovatelnost: vhodné pro zpracování velkého množství dat.
- Vysoká přesnost díky pokročilým algoritmům.
- Snadná integrace přes API.

Nevýhody:

- Vyžaduje přenos dat do cloudu, což může být nevhodné z hlediska bezpečnosti a soukromí.
- Nutné připojení k internetu.

Open-Source řešení pro zpracování videa na PC

Pro zpracování záznamů z mobilu a obdobných zařízení na PC lze využít open-source řešení.

OpenALPR (Open Source) je nástroj pro rozpoznávání RZ s podporou více platforem (Linux, Windows, Mac), který dokáže zpracovávat obrázky i videa.

EasyOCR + vlastní modely představuje open-source OCR knihovnu, kterou lze upravit pro rozpoznávání RZ.

YOLO + Tesseract kombinuje detekci objektů pomocí YOLO a OCR knihovny Tesseract umožňuje efektivní rozpoznávání RZ.

Výhody:

- Bezplatné řešení s možností přizpůsobení.
- Vysoká kontrola nad procesem zpracování.

Nevýhody:

- Potřeba technických znalostí pro instalaci a konfiguraci.
- Náročnější na hardware, zejména při práci s delšími videy.

Mobilní frameworky pro vývoj vlastních aplikací

Další možností je vývoj vlastní aplikace na základě existujících frameworků a knihoven.

TensorFlow Lite + ML Kit představují frameworky pro vývoj aplikací s rozpoznáváním obrazu přímo na mobilních zařízeních. Lze integrovat předtrénované modely pro OCR nebo trénovat vlastní modely na detekci RZ.

PyTorch Mobile umožňuje implementaci AI modelů na mobilních zařízeních.

OpenCV (Android / iOS) je populární knihovna pro zpracování obrazu, která podporuje funkce OCR.

Výhody:

- Přizpůsobitelnost a možnost optimalizace.
- Funkčnost i off-line.

Nevýhody:

- Vyšší časová a znalostní náročnost vývoje.

Specializované desktopové nástroje pro dávkovou analýzu

Existují také specializované desktopové aplikace, které umožňují hromadné rozpoznávání RZ v záznamech pořízených např. mobilním telefonem nebo obdobnými běžnými zařízeními jako fotoaparáty či kamery. Např. aplikace **LprFolderRecognition** od společnosti MACH SYSTEMS s.r.o. poskytuje hromadné rozpoznávání RZ v obrazových souborech uložených ve vybraném adresáři.

Vigilant PlateSearch (by Motorola) představuje profesionální software pro zpracování RZ z videí a obrázků.

TrafficVision je specializovaný nástroj pro analýzu dopravních záznamů, včetně detekce a rozpoznání RZ.

Výhody:

- Optimalizováno pro analýzu dopravních scén.
- Možnost dávkového zpracování.

Nevýhody:

- Vyšší náklady na licenci.
- Vhodné spíše pro větší projekty.

Shrnutí

Pro rozpoznávání RZ na záznamech z mobilu a podobných běžných zařízení jako fotoaparátů nebo kamer je na výběr od jednoduchých mobilních aplikací až po robustní cloudové služby nebo open-source řešení. Při výběru vhodného softwaru je důležité zvážit konkrétní potřeby, jako je požadovaná přesnost, rychlost zpracování, dostupný hardware a rozpočet. Volba závisí na konkrétních požadavcích, zjednodušeně lze rozlišit následovně:

- **Pro rychlé a jednoduché použití:** mobilní aplikace nebo cloudové služby.
- **Pro flexibilitu a nízké náklady:** open-source řešení jako OpenALPR.
- **Pro profesionální a přesné výsledky:** komerční SW nebo vývoj vlastních nástrojů.

4.3 Kamery s funkcí LPR

Provedená rozsáhlá rešerše ukázala, že na trhu je mnoho výrobců, kteří nabízí kamery s integrovanou funkcí rozpoznávání RZ či kompletní systém zahrnující HW i SW. V českém prostředí je několik větších renomovaných dodavatelů dodávajících většinu rozsáhlých systémů zajišťujících monitoring v dopravní oblasti. Tato kapitola zobecňuje zjištěné informace a poskytuje tak přehled o možnostech trhu. (31), (46), (47), (48), (49), (50), (51), (52), (53), (54), (55), (56), (57), (58), (59), (60), (61), (62)

Někteří výrobci nabízí primárně přístupové systémy – komplexní správu vjezdů a výjezdů vozidel, ale i kamery s rozpoznáváním RZ do rychlosti mnohdy až 200 km/h. Řešení nabízí rozlišení různých skupin uživatelů s definovaným oprávněním i nastavením cen, vytváření reportů a statistik, je univerzální a přizpůsobitelné s možností integrace dalších modulů či systémů. Pomocí optického rozpoznávání znaků (OCR) a videokamer zachycuje ANPR snímky registračních značek a převádí je na alfanumerické znaky za méně než sekundu. Uživatelé je pak mohou porovnat s databází registrovaných vozidel a spustit příslušné akce (zvednutí závoje nebo spuštění alarmu). Kamery ANPR společnosti Hikvision identifikují parametry vozidel, jako je barva, typ a výrobce, a také směr jízdy. Tyto údaje spolu s poznávacími značkami umožňují rozsáhlé a flexibilní aplikace.

Obecně kamery díky rozpoznávání RZ umožňují vyhledávání odcizených či hledaných vozidel, upozornění na neoprávněný vjezd atp. Zpravidla jsou modulární, rozšiřitelné, kompatibilní, fungují bezobslužně a disponují intuitivním ovládáním. Přesnost rozpoznání RZ je uváděna přes 96 % při vhodných viditelnostních podmínkách. Výrobci garantují přesnost rozpoznávání RZ nejčastěji do rychlosti 100 / 120 / 130 / 200 km/h. Napájecí zdroj je vyžadován 12 V DC, POE nebo 110-220 V. Provozní teplota bývá od -40 °C do +60 °C, definované zorné pole a úhel záběru cca 50°-40° ~15°- 8°.

Standardem jsou kamery v anti vandal provedení / s ochranou proti vandalismu. Kamery mají stupeň krytí IP67, což je spolehlivý stupeň krytí pro venkovní použití i za zhoršených povětrnostních podmínek (vodotěsnost, vniknutí prachu atp.) a stupeň mechanické odolnosti IK10 (ochrana proti nárazu). Některé kamery disponují audio vstupem / výstupem, mají zabudovaný mikrofon či reproduktor, nabízí tak jednosměrnou nebo obousměrnou komunikaci. Používán je infračervený (IR) přísvit (zpravidla LED) s dosahem 20–100 m pro noční provoz a monitoring při zhoršených viditelnostních podmínkách. V dnešní době již nabízené produkty zpravidla podporují více státních příslušností (mnohdy desítky). Instalační výška bývá kolem 6 m, buď vedle jízdního pruhu nebo nad jízdní pruh, IR přísvit může být dodáván zvlášť. Některá řešení lze propojit s mobilní aplikací, SW rozpoznává model, značku, barvu

vozidla a většinou dokáže i klasifikovat příslušnou kategorii, rozlišení kategorií se liší, obecně se jedná o osobní vozidla, dodávky a nákladní vozidla, pěší a případně další kategorie (např. cyklisti, skútry atp.). Některé kamery mohou mít zabudovaný procesor s hlubokým učením, jedno z nalezených řešení mělo např. tyto parametry: rozpoznání 2 jízdních pruhů (obvyklé řešení), přísvit do vzdálenosti až 30 m, napájení DC 12 V / PoE (IEEE 802.3bt), stupeň krytí IP68, odolnost vůči vandalismu IK10, rozpoznání vozidel do rychlosti až 100 km/h, export rozpoznávaných RZ, podpora iPhone, iPad, Android & 3GPP, webové rozhraní v 31 jazycích.

K dispozici jsou také variantní řešení – varianta bez použití serveru nebo s použitím serveru. Samostatné řešení LPR detekce kamerou, ke kterému není třeba server. Umožňuje spolupráci více kamer najednou (např. 4), nebo jednotlivě neomezeně. Kamery publikují RZ pomocí API kanálu, nebo jsou dostupné na webu. IP technologie umožňuje spolupráci několika kamer najednou. Ty snímají potřebná data a přenášejí je do uživatelského webového rozhraní. Řešení založené na serveru, kde dochází k detekci RZ, umožňuje dosáhnout větší přesnosti detekce. Komunikační rozhraní pomocí TCP, XML, API – realtime komunikační rozhraní pro snadnou integraci. Serverové řešení umožňuje lépe škálovat systém. Dohled a administrace probíhá z jednoho místa, podporuje hybridní řešení LPR běžící na serveru i kamerách a nabízí možnost úpravy speciálních znakových sad.

4.4 Softwarové řešení

Obvykle je nabízeno celé řešení vč. HW, většinou lze nakombinovat, existuje i možnost koupě pouze SW řešení nebo služby tzv. na klíč, kdy je video nahráno / posláno zprostředkovateli takové služby a zákazník obdrží výslednou sadu rozpoznávaných RZ nebo i jiného zpracování / vyhodnocení dopravních parametrů z videa. Vybraná řešení vč. konkrétních příkladů jsou uvedena výše, tato kapitola se zaměřuje na profesionální SW umožňující spolehlivé rozpoznání RZ. (63), (64), (65), (66), (67), (68), (69), (70)

Kritériem úspěšného rozpoznání je viditelnost, minimální snímkovací frekvence / počet snímků za sekundu (min. 15), oslnění a rozlišení obrazu. Příkladem SW na rozpoznání RZ je platforma ATEAS Security dostupná v 5 základních edicích, 4 z nich nabízí funkci rozpoznávání RZ, umožňuje také propojení více kamer. Službu zpracování analýzy videa, kdy je video nahráno a uživatel poté obdrží výsledek zpracování nabízí např. společnost VisionCraft. Společnost GoodVision poskytuje u svého SW možnost nahrání videa nebo napojení kamer živě. Axxon One Enterprise nabízí licenci pro rozpoznání RZ vozidel a neomezenou databázi vozidel, jejich SW Road AR zaručuje rychlý provoz. Jedná se o doplňkovou licenci softwaru, která živě umožňuje rozpoznávat RZ vozidel z bezpečnostní kamery a odesílat oznámení (UI, Text, Email, API), pokud je detekována konkrétní RZ z databáze. Modul je možné použít pro otevírání brány, či spuštění jiného makra pro automatizaci.

Příkladem spolehlivého SW, který je možné využít na Fakultě dopravní, je SW Eyedea, který nabízí cloudovou i desktopovou verzi. Disponuje přesností rozpoznání RZ přes 99 %, vysokou rychlostí rozpoznání a dokáže číst i víceřádkové značky (např. u motocyklů). Rychlost a přesnost je dosažena díky využití natrénované umělé inteligenci. Nejprve je třeba nahrát fotografie či video, v tomto souboru jsou poté rozpoznávány jednotlivé znaky v oblasti, kde se dle programu nachází RZ. Aktuálně SW rozpozná většinu státních příslušností a zobrazuje spolehlivost, s jakou je konkrétní RZ rozpoznána. Výhodou je možnost využití pro video z profilu komunikace, vjezdu či výjezdu i RZ parkujících vozidel. (71)

Je mnoho dalších systémů pro rozpoznání RZ vozidel s obdobnými funkcemi, např. LOOK (spol. HISYS) – SW na rozpoznávání v reálném čase při neomezené rychlosti vozidel. SW na analýzu videa / dopravních dat i kamera s rozpoznáváním RZ (spol. PROFComms), dopravní kamera s integrovanou detekcí RZ do 250 km/h, 5 Mpx přehledová kamera, IR přísvit, IP68, 24 VDC. Posledním konkrétním příkladem je SW pro přístupové systémy (spol. GeoVision) s hlubokým učením, porovnává záznamy v managementu a

zdroje z připojeného videa, který podporuje až 100 000 vozidel, dokáže jednomu uživateli přiřadit více vozů, podávat informace o volných kapacitách i vzniklých mimořádných událostech.

Existují také open source řešení, která jsou veřejně dohledatelná přes web, většina z nich však nabízí pouze omezené funkce a plná funkčnost je uživateli dostupná až v placené verzi.

5 Zpracování LPR dat a otázky kvality dat

Kvalitu softwarového procesu rozpoznání registračních značek ovlivňuje celá řada vnějších faktorů, jak již bylo uvedeno v předchozí kapitole, v přehledu současných dostupných technologií. Tyto faktory jsou dlouhodobé, dlouhodobě dokumentované i v odborné literatuře a vývojem je snaha dosáhnout minimalizace jejich vlivu. Pro shrnutí – kvalita SW rozpoznávaných RZ souvisí jak s technickými parametry využitého záznamového zařízení (zejm. rozlišení obrazového záznamu, fps a obecně poruchovost) (4), (12), tak se situačními aspekty v místě měřicích profilů (12), (26). Situační aspekty zahrnují zejména úhel kamery vzhledem k projíždějícím vozidlům, dobu, po kterou vozidla zůstávají v záběru a jejich rychlost, umístění a kvalita (čistota) samotných registračních značek vozidel, úroveň kvality dopravy v místě měřicího profilu (přítomnost kolon způsobuje časté vzájemné zákryty vozidel a zakrytí registračních značek). Nezanedbatelný vliv mají rovněž světelné podmínky a počasí (12), (26) – v případě noci je pro úspěšné rozpoznání registračních značek potřeba využít např. infračervený přísvit, negativně se může projevit oslnění sluncem, hranice světla a stínu, zhoršená viditelnost vlivem deště, mlhy nebo sněžení.

V praxi sběru LPR dat v podmínkách (nejen) České republiky je metodika, technologie měření a vyhodnocení dat nezanedbatelně ovlivněna faktory jako finanční rozpočet a další požadavky zadavatele průzkumu, stejně jako reálnou praxí zpracovatele (72). Často jsou využívány přenosné kamery, které jsou buď dočasně instalovány na infrastrukturu (73) nebo mohou být umístěny na běžné stativy a po celou dobu měření jsou pod fyzickým dohledem proškolených pracovníků (23). Takto koncipované měřicí profily však vykazují vyšší chybovost a výpadky záznamů – běžná spolehlivost rozpoznání registračních značek může klesat například k 80 % (74), což má negativní vliv na výsledky.

Předpokladem úspěšného praktického využití LPR dat je proto jejich vhodné zpracování a vyhodnocení, s využitím postupů, které zamezí vnášení nových chyb do dat a minimalizují vliv existujících chyb na interpretaci výsledků (12). Specifickým úskalím při vyhodnocování směrových dat je vyhodnocení směrovosti vozidel (11), (75). Chybně rozpoznanou registrační značku není bez aplikace korekčních metod možné spárovat se záznamy stejného vozidla v jiných měřicích profilech. Dochází k chybnému určení trasy vozidla a toto má vliv na celkovou kvalitu výsledků vyhodnocení. Metodami korekce chybějících nebo chybných LPR dat se zabývají články: (11), (19) (75), (76). Pokud však nejsou k dispozici „ground-truth“ data, možnosti korekce chyb jsou pouze omezené. (19), (76).

Uvedené odborné zdroje zabývající se korekcí chyb v datech způsobených chybně rozpoznávanými RZ, budou v následujících odstavcích podrobněji popsány. Lze je rozdělit na 3 obecné kategorie vzhledem k tomu, z jakého úhlu pohledu ke korekci LPR dat přistupují:

- 1) Korekce chybějících záznamů RZ (ve smyslu doplnění informace o intenzitách)
- 2) Korekce neúplných trajektorií vozidel na základě charakteristik pohybu ostatních vozidel v síti
- 3) Korekce párování chybně rozpoznávaných registračních značek

Do první kategorie lze zařadit následující odborný zdroj:

Missing data detection and imputation for urban ANPR system using an iterative tensor decomposition approach

Han Zhang , Peng Chen , Jianfeng Zheng, Jingqing Zhu , Guizhen Yu , Yunpeng Wang , Henry X. Liu (2019) (22)

Autoři se zabývají chybějícími daty z LPR systémů. Neřeší tedy chyby v rozpoznání znaků RZ, ale řeší zcela chybějící záznamy o průjezdech vozidel, jejich detekci a opravu naměřených počtů intenzit. Za tímto účelem vyvinuli algoritmus iterativní tenzorové dekompozice. Prakticky byla navržená metoda testována na robustní reálné sadě dat, z 237 městských LPR lokací (v Číně) sbírajících data v průběhu prosince 2017, každý den od 6:00 do 22:00, která jsou agregovaná v 30minutových intervalech. Testovány byly případy náhodně chybějících krátkých úseků dat a extrémních případů výpadků, kdy chyběly celé dny. Pro testování byla vybrána data z LPR stanic s kompletním průběhem, tato data byla následně uměle přeškálována a byly do nich vneseny chybějící data, aby došlo k vytvoření testovacích datových sad. Navržená metoda se ukázala být jako přesná v detekci a opravě dat a podává lepší výsledky než stávající tenzorové metody.

Do druhé kategorie spadají následující odborné zdroje:

Vehicle Trajectory Reconstruction from Automatic License Plate Reader Data

Yu, H., Yang, S., Wu, Z., Ma, X. (2018) (11)

Autoři se zabývali rekonstrukcí trajektorií vozidel po zkoumané oblasti, přičemž záznamy o průjezdech vozidel byly získávány z LPR sítě kamer instalovaných na 108 křižovatkách v městské oblasti Ruian v Číně. Výzkum byl zaměřen na doplnění chybějících záznamů průjezdů vozidel za účelem získání kompletních trajektorií. Schéma představené metody je následující: Zdrojová LPR data, topologie silniční sítě -> sestavení trajektorií vozidel -> identifikace nekompletních trajektorií (kvůli chybějícím záznamům) -> vyhledání všech možných trajektorií pro doplnění chybějících záznamů -> algoritmus pro výběr nejvhodnějších trajektorií -> kompletace nejvhodnějších trajektorií. Pro vyhledání všech možných trajektorií byl využit algoritmus DFS (depth-first search), pro určení nejvhodnější trajektorie byl využit algoritmus TOPSIS (technique for order preference by similarity to an ideal solution). Metoda byla ověřována na datové sadě 2000 kompletních trajektorií, ze kterých byly náhodným výběrem odstraněny některé záznamy, takže bylo dosaženo 30 %, 60 % a 80 % chybějících informací. Výsledky prokázaly, že metoda je slibná, nicméně se zvyšujícím se podílem chybějících informací se zhoršuje její přesnost. Mez, kdy metoda dosahuje přesnosti 80 %, byla nalezena při 50 % chybějících informací.

Vehicle Identity Recovery for Automatic Number Plate Recognition Data via Heterogeneous Network Embedding

Yixian Chen, Zhaocheng He (2020) (77)

Autoři se zabývají řešením záznamů z LPR systémů, které obsahují všechny charakteristiky o průjezdu vozidla v definovaném čase a měřicím bodě, avšak z důvodu nerozpoznání chybí textový řetěz registrační značky. V navržené metodě využívají seskupování záznamů vozidel na základě časoprostorových charakteristik jejich pohybu v síti a párují nerozpoznané záznamy se známými RZ vozidel. Vycházejí z metod a algoritmů pro práci s heterogenními informačními sítěmi. Navržená metoda byla ověřena na datové sadě z LPR senzorů umístěných na 9 křižovatkách ve městě Xuancheng v Číně, data byla měřena v průběhu 4 dopoledních hodin, 7:00 – 11:00. Z naměřených dat bylo náhodně

odstraněno 5 % a 10 % záznamů RZ a na takto vytvořených datových sadách byla vyhodnocována úspěšnost metody. Výsledky ověřily účinnost navržené metody.

Origin-destination pattern estimation based on trajectory reconstruction using automatic license plate recognition data

Rao, W., Wu, Y., Xia, J., Ou, J., Kluger, R. (2018) (75)

Autoři navrhli offline metodu pro odhad vzorců zdrojů a cílů cest vozidel z LPR dat v rozsáhlé městské síti. Konkrétně se zabývali odhadem pravděpodobnosti trajektorie vozidla ve zkoumané oblasti, pokud z naměřených LPR je možné určit trajektorii pouze částečně. Pro výběr z množiny možných trajektorií, kterou určovali na základě zjištěných kompletních trajektorií vozidel v určitém časovém intervalu, využívali částicový filtr. Navržená metoda byla implementována na dopravní síti ve městě Kunshan v Číně, přičemž byla využita data ze 344 LPR senzorů instalovaných na příjezdech ke křižovatkám, data byla měřena v průběhu pracovních dní po dobu 2 týdnů. Soubor dat obsahoval přibližně 175 tis. záznamů za hodinu, přičemž bylo zaznamenáno přibližně 88% celkového počtu vozidel. Výsledky byly ověřeny s využitím simulačního dopravního modelu, neboť zdrojová data neobsahovala přesnou referenční sadu. Výsledky prokázaly, že navrhovaná metoda efektivně využívá data ze senzorů ALPR k dosažení vysoce přesné rekonstrukce trajektorie a odhadu historických vzorů OD a lze ji použít v reálných situacích řízení dopravy. Zároveň bylo stanoveno, že pro využitelnost navržené metody je třeba, aby bylo zaznamenáno alespoň 60 % celkového počtu průjezdů vozidel.

Do třetí kategorie lze zařadit následující odborné zdroje:

Online license plate matching procedures using license-plate recognition machines and new weighted edit distance

Oliveira-Neto, F. M., Han, L. D., Jeong, M. K. (2011) (76)

Oliveira-Neto, F. M., Han, L. D., Jeong, M. K. (2011) Online license plate matching procedures using license-plate recognition machines and new weighted edit distance. *Transportation Research Part C: Emerging Technologies*, 21(1), 306-320. doi: 10.1016/j.trc.2011.11.003.

Autoři se zabývají párováním chybně rozpoznávaných registračních značek vozidel na dvou měřicích profilech LPR systému. Svůj přístup založili na váhových funkcích, konkrétně GED (generalized edit distance) a CED (constraint edit distance). Navrhli novou váhovou funkci založenou na pravděpodobnostním modelu, přičemž zohlednili fakt, že při rozpoznávání textů v LPR systémech nedochází k reverzním chybám (tzn. k přehazování znaků v textovém řetězci). Dále autoři využívají údajů o čase záznamu, které jsou v LPR systémech k dispozici. Autoři byli svojí metodou schopni dosáhnout 97% úspěšnosti při párování záznamů vozidel s 0% falešně spárovaných záznamů.

Speed profile estimation using license plate recognition data

Mo, B., Li, R., Zhan, X., (2017) (19)

Autoři řeší chybně rozpoznaná LPR data v kontextu určení rychlostních profilů vozidel ve dvou po sobě následujících měřicích profilech za účelem tvorby emisního modelu. Pro spárování chybně rozpoznávaných záznamů RZ využívají zjednodušenou metodu představenou v (76). Situace, kterou autoři

řeší, je specifická tím, že se jedná o mezikřižovatkový úsek, kdy LPR senzory jsou umístěny za stopčárou křižovatek. Autoři tedy za účel určení realistické doby jízdy jednotlivých vozidel využívají další podmínky vyplývající z tohoto uspořádání, jako korekci chybně určeného času z důvodu zastavení vozidla před křižovatkou a znalost signálního plánu křižovatky. Metodu představenou v (76) zjednodušují na problém párování textových řetězců o stejné délce s tím, že záznam registrační značky na druhém měřicím profilu považují za reálně správný záznam vozidla.

6 Vymezení oblasti výzkumu

V souladu s provedenou analýzou současného technologického stavu metod sběru a zpracování LPR dat, jejich chybovosti a možností eliminace chyb se tato studie zaměřuje na LPR data získaná SW rozpoznáním videozáznamů průjezdů vozidel naměřených především široce dostupnými záznamovými zařízeními, jako jsou např. mobilní telefony. Tento zdroj dat se vyznačuje vyšším potenciálem chybovosti než profesionální kamery vyvíjené pro účely detekce a rozpoznání registračních značek, přesto je však v současné praxi stále hojně využíván, pro účely krátkodobých dopravních průzkumů, zejména vzhledem k nízkým nákladům. Studie se zaměřuje především na adaptaci a rozšíření metody představené v (76) a zahrnutí této metody do komplexního algoritmu pro efektivní přípravu LPR, který bude umožňovat dosažení jejich potřebné spolehlivosti. Vybraná metoda bude obohacena především o využití informace o pravděpodobnosti správného rozpoznání textového řetězce RZ, kterou poskytují některé SW pro rozpoznání RZ, např. SW Anpr GUI 2.1.1 od společnosti Eyedea Recognition s.r.o., který je v této studii využíván.

Metoda představená v (76) byla zvolena z následujících důvodů:

Především metoda pracuje přímo s konkrétními záznamy registračních značek, které jsou k dispozici z LPR systému, ale naměřené s potenciální chybou. Metoda tedy využívá potenciál informací, které o průjezdech vozidel jsou k dispozici. Na rozdíl od toho řada nalezených metod (11), (75), (77) zabývajících se rekonstrukcí trajektorií vozidel v území tyto informace nevyužívá, předpokládá, že záznamy o průjezdech vozidel zcela chybí a trajektorie rekonstruuje na základě odhadu a modelů s využitím trajektorií ostatních spárovaných vozidel v oblasti a s využitím znalostí o topologii silniční sítě.

Dále, vybraná metoda (76) byla vyvinuta na datech ze dvou měřicích profilů, ovšem skýtá potenciál pro rozšíření do rozsáhlejších sítí LPR senzorů. Rozšíření této metody je předmětem výzkumu popisovaného v této studii, nicméně jako modelový příklad je brána standardní situace dopravního průzkumu v podmínkách České republiky, tedy oblast vymezená především měřicími profily na jejích hranicích (vjezdy, výjezdy) s přítomností vnitřních měřicích profilů umístěných na klíčových místech silniční sítě. Nalezené metody zabývajících se rekonstrukcí trajektorií vozidel (11), (75) byly na rozdíl od toho většinou vytvářeny pro rozsáhlou síť měřicích profilů, přičemž silniční síť jimi byla hustě pokryta (např. na většině křižovatek) a není zřejmá adaptovatelnost těchto metod pro oblasti s řidším pokrytím měřicími profily.

Přepokladem pro úspěšnou implementaci vybrané metody a návrh efektivního algoritmu byla tvorba referenční přesné sady LPR dat, která sloužila pro nastavení vstupů algoritmu. Tato datová sada byla podrobně analyzována vzhledem k chybovosti. Popisem přípravy datové sady se zabývá kapitola 7.

Navržený algoritmus doplňuje a rozšiřuje metodiku vyhodnocení směrových dopravních průzkumů, která byla vytvořena a ověřena v rámci diplomové práce (78). Navržený algoritmus se proto zabývá čistě postupem pro zvýšení kvality zpracovávaných LPR dat za účelem dosažení přesnějších výsledků. V souladu s vybranou metodou (76) se algoritmus zabývá především fází párování dat. Popisem algoritmu a implementací využívaných metod a postupů se zabývají kapitoly 8-10.

V závěru výzkumu je navržený algoritmus modelově implementován na datech z nově provedeného směrového průzkumu a je provedeno jeho ověření a stanovení procentuálního zpřesnění spárovaných dat. Popisem implementace a ověření algoritmu se zabývá kapitola 11.

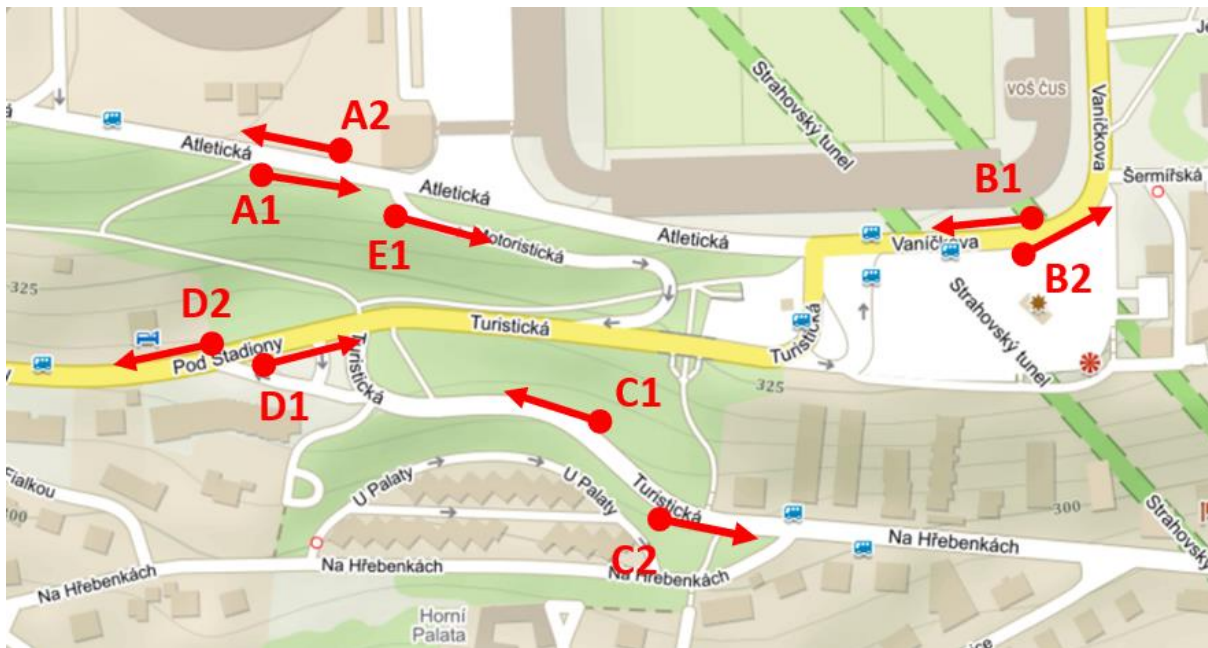
7 Příprava referenční datové sady

Cílem bylo připravit zcela správnou (referenční) sadu LPR dat, která bude dále využívána pro analýzu SW určených spolehlivostních parametrů a pro aplikaci metody zvýšení spolehlivosti dat pomocí pravděpodobnostní matice. Za „zcela správná“ jsou data považována ve chvíli, kdy přesně odpovídají reálnému stavu dopravy, a to v parametrech: RZ vozidla, čas průjezdu vozidla a měřící profil, na kterém bylo vozidlo zaznamenáno. Příprava referenční datové sady proběhla v několika krocích:

1. SW rozpoznání RZ ze zdrojových videozáznamů,
2. manuální kontrola datových sad vůči zdrojovým videozáznamům, oprava a doplnění,
3. spárování dat dle metodiky pro vyhodnocení směrových průzkumů,
4. opětovná kontrola spárovaných dat dle referenčních videozáznamů, oprava a doplnění.

7.1 Zdroj dat

Referenční datová sada byla vytvářena dle originálních videozáznamů z proběhlého průzkumu na pražském Strahově.



Obrázek 1 Schéma rozložení měřících profilů směrového průzkumu Strahov

Následující tabulka zobrazuje časové intervaly, ve kterých jsou videozáznamy z jednotlivých měřících profilů k dispozici. Tyto časové intervaly určují rovněž dobu, po kterou je k dispozici referenční datová sada. Poslední řádek tabulky zobrazuje pro každé stanoviště časový úsek, po který je videozáznam k dispozici bez výpadků. (Dvousekundový výpadek při automatickém uložení a opětovném spuštění videozáznamu nemá negativní vliv na kompletnost referenční datové sady.)

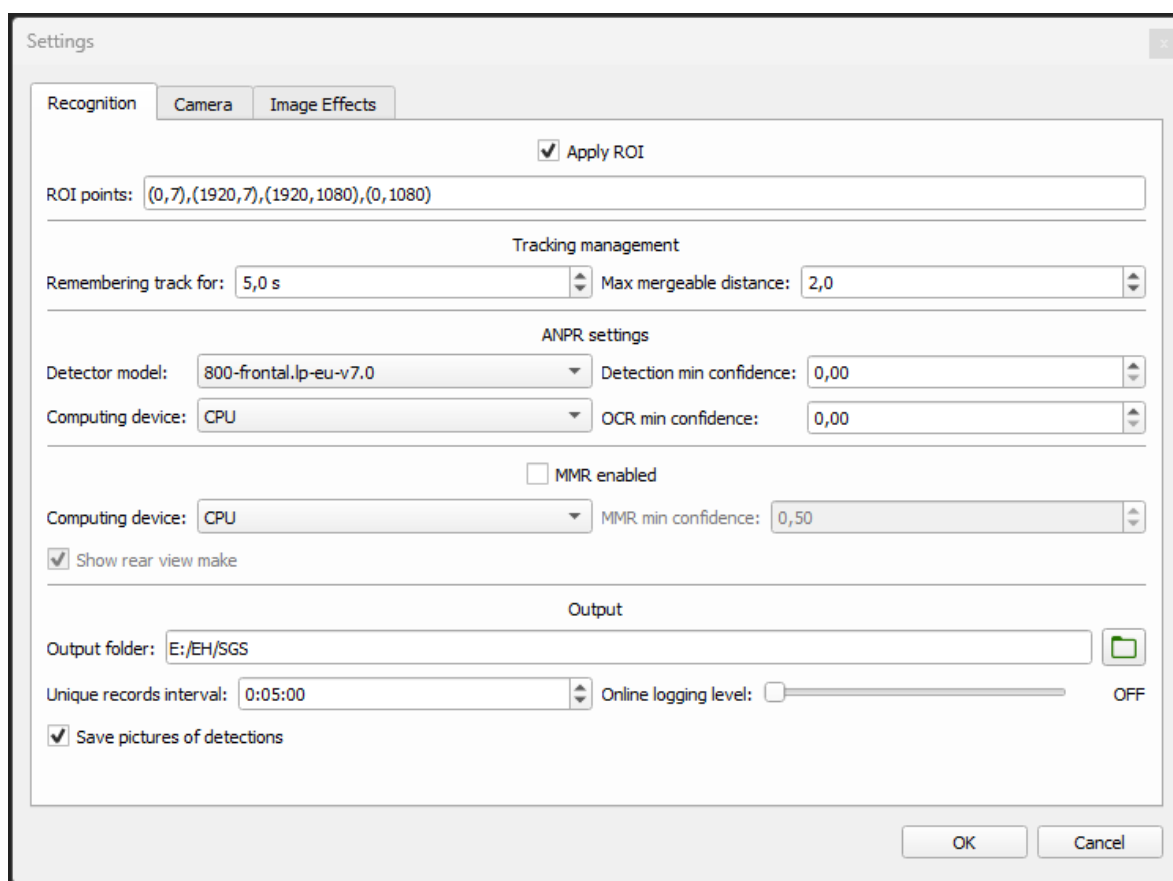
Tabulka 3 Dostupné časové intervaly videozáznamů, tzn. referenční datové sady (doba měření byla 8:00-9:15)

	Měřící profil								
	A1	A2	B1	B2	C1	C2	D1	D2	E1
Časové intervaly	8:00:05-8:01:04	7:58:27-8:58:27	7:59:49-8:59:49	8:00:44-9:00:44	7:56:54-8:56:54	7:58:36-8:58:36	8:00:36-9:00:36	7:59:29-8:02:04	7:55:37-8:49:30

	8:01:32- 9:01:32	8:58:29- 9:15:12	8:59:51- 9:15:22	9:00:46- 9:15:47	8:56:57- 9:14:58	8:58:38- 9:15:01	9:00:39- 9:15:06	8:02:19- 8:02:25	8:50:19- 9:15:35
	9:01:34- 9:15:13							8:04:32- 9:04:32	
								9:04:35- 9:19:36	
Časový úsek bez výpadků měření	8:01:32- 9:15:13	7:58:27- 9:15:12	7:59:49- 9:15:22	8:00:44- 9:15:47	7:56:54- 9:14:58	7:58:36- 9:15:01	8:00:36- 9:15:06	8:04:32- 9:19:36	7:55:37- 9:15:35 (při využití A1)

7.2 SW rozpoznání RZ ze zdrojových videozáznamů

K SW rozpoznání RZ z referenčních videozáznamů byl využit SW Anpr GUI 2.1.1 od společnosti Eyedea Recognition s.r.o. (dále jen SW Eyedea). Software byl využit s následujícím nastavením (viz Obrázek 1). Z důvodu technické chyby nebyl k dispozici modul MMR, ale na přípravu referenční datové sady toho nemělo vliv.



Obrázek 2 Nastavení SW Eyedea

Výstupem ze SW Eyedea byla sada csv datových souborů, jeden soubor pro každý referenční videozáznam. Pro snazší další práci s daty byly sloučeny soubory ze stejných měřicích profilů, tzn. vznikla sada 9 souborů s daty, ve formátu xlsx pro snadné zpracování v MS Excel. Všechny soubory mají

stejnou strukturu. V řádcích jsou uloženy jednotlivé záznamy o průjezdech vozidel, 1 řádek odpovídá jednomu záznamu. Ve sloupcích jsou uloženy parametry (informace), které ke každému záznamu poskytuje SW Eyedea. Přehled počtů SW rozpoznávaných záznamů a seznam parametrů poskytovaných SW Eyedea zobrazuje následující tabulka:

Tabulka 4 Výstup ze SW Eyedea

Počty záznamů rozpoznané SW Eyedea									
Měřicí profil	A1	A2	B1	B2	C1	C2	D1	D2	E1
Počet záznamů	518	361	1555	1446	1243	884	897	667	249
Celkem	7820								
Parametry vyhodnocované SW Eyedea									
Název	Popis ^{*)}								
ID	Interní identifikace stopy								
Licence plate	Text rozpoznané RZ								
Country	Mezinárodní kód státu uvedený na RZ								
LPM confidence	Faktor spolehlivosti detekce RZ (rozsah 0-1)								
OCR confidence	Faktor spolehlivosti rozpoznání textu RZ (rozsah 0-1)								
Frame count	Počet snímků s detekovanou RZ								
ROI	Souřadnice ROI (oblasti zájmu)								
Direction	Směr jízdy vozidla (Incoming/Outgoing)								
Start time	Počáteční čas detekce stopy (relativně od začátku videozáznamu)								
End time	Koncový čas detekce stopy (relativně od začátku videozáznamu)								
Best detection X-coordinate	Vodorovná souřadnice středu RZ s nejvyšší spolehlivostí OCR								
Best detection Y-coordinate	Svislá souřadnice středu RZ s nejvyšší spolehlivostí OCR								
Best detection frame number	Číselné označení (pořadí) snímku s nejvyšší spolehlivostí OCR								
Best detection frame timestamp	Časové razítko detekce s nejvyšší spolehlivostí OCR (relativně od začátku videozáznamu)								
Category	Kategorie vozidla								
Category confidence	Faktor spolehlivosti rozpoznání „Category“ (rozsah 0-1)								
Color	Barva vozidla								
Color confidence	Faktor spolehlivosti rozpoznání „Color“ (rozsah 0-1)								
Make	Model vozidla								
Make confidence	Faktor spolehlivosti rozpoznání „View“ (rozsah 0-1)								
View	Směr vozidla dle viditelného profilu (Frontal/Rear)								
View confidence	Faktor spolehlivosti rozpoznání „View“ (rozsah 0-1)								
Source file name	Název souboru zdrojového videa								
Image file name	Název souboru uloženého snímku vozidla								
Legenda: tmavě zelené – parametry využití v další práci světle zelené – kontrolované parametry pro možné budoucí využití šedé – z důvodu nefunkčnosti modulu MMR prázdné									
*) popis parametrů byl vytvořen dle návodu pro novější verzi SW Eyedea (2.3.1), z důvodu nedostupnosti manuálu pro použitou verzi									

7.3 Manuální kontrola datových sad vůči referenčním videozáznamům, oprava a doplnění

Výše popsané soubory s daty, které tvořily výstupy ze SW Eyedea, byly zkontrolovány dle originálních videozáznamů. V případě nesrovnalostí byly v maximální míře opraveny, případně doplněny. Pokud nebyla možná plná manuální oprava (doplnění) z důvodu snížené kvality referenčních videozáznamů, byla k příslušnému záznamu přidána tato informace a alternativní upřesnění. Zkontrolovaná, opravená a doplněná datová sada byla následně využita pro přípravu datové sady pro vyhodnocení spolehlivostních parametrů a vstupní datové sady pro vyhodnocení dle metodiky pro vyhodnocení směrového průzkumu.

Konkrétní způsob kontroly byl následující:

Do datové sady bylo doplněno několik dalších parametrů (viz níže Tabulka 5). Tyto parametry indikují výsledek provedené kontroly nebo v případě nalezených nepřesných nebo chybějících hodnot u kontrolovaných parametrů uchovávají správnou informaci, případně uchovávají doplněnou informaci, která byla doplněna pro další práci s datovou sadou.

Tabulka 5 Parametry doplňované na základě kontroly

Parametry doplňované na základě kontroly	
Název	Popis
Správně celkem	Souhrnná indikace správnosti celého záznamu, bližší popis chyb, případně jejich pravděpodobných zdrojů
Správně RZ	Indikace správnosti parametru Licence plate (ano/ne)
Správná RZ	Manuální záznam správné RZ
Správně Country	Indikace správnosti parametru Country (ano/ne)
Správně Direction	Indikace správnosti parametru Direction (ano/ne)
Správně čas	Indikace správnosti parametru Best detection frame timestamp (ano/ne, pozn. na základě principu fungování SW Eyedea považován automaticky za správný, jako chybný byl označen v případě chybějícího záznamu)
Správný čas	Manuální záznam správného času od začátku videa
Referenční čas	Čas začátku referenčního videozáznamu
Reálný čas	Vypočítaný reálný čas záznamu vozidla
Druh Vozidla	Druh vozidla (doplňován manuálně vzhledem k nefunkčnosti modulu MMR)

V průběhu kontroly docházelo v zásadě k doplňování informací o kvalitě dat ke každému ze záznamů. Rovněž docházelo k doplňování chybějících záznamů a k doplňování správných verzí chybných záznamů RZ. Nedocházelo k odstraňování žádného ze záznamů. Kontrolu datové sady lze rozdělit do několika bloků:

Kontrola parametru času

Vzhledem k tomu, že parametr Best detection frame timestamp uchovává časovou informaci relativně od začátku videozáznamu, byly ke všem záznamům doplněny parametry Referenční čas a Reálný čas za účelem přepočtu času. Po přepočtení časových razítek záznamů na skutečný čas průjezdů vozidel v době průzkumu dojde k časové synchronizaci všech záznamů ze všech měřicích profilů v datové sadě.

Samotný softwarově určený parametr Best detection frame timestamp byl při kontrole jednotlivých záznamů považován za správný. Nebyl tedy kontrolován a pouze probíhalo doplnění relativních časových razítek záznamů u zcela chybějících záznamů.

Doplnění druhu vozidla

Vzhledem k nefunkčnosti modulu MMR byl během kontroly ke všem záznamům manuálně doplňován parametr *Druh vozidla*.

Kontrola RZ (parametr Licence plate)

Kontrola SW rozpoznávaných RZ vůči referenčnímu videozáznamu tvořila zásadní část procesu kontroly datové sady. Každý SW vytvořený záznam v datové sadě byl zejména v parametru Licence plate porovnáván s referenčním videozáznamem. Výsledky porovnání (číselně a procentuálně viz Tabulka 4) lze shrnout následovně:

- *vozidlo v opačném směru* – zaznamenaná RZ náležela vozidlu v opačném směru jízdy, než který náležel příslušnému měřicímu profilu. V tomto případě nebyla prioritou další kontrola správnosti záznamu, záznam byl pouze označen pro snadnou budoucí identifikaci.

Následující body se tedy vztahují pouze k záznamům vozidel ve směru náležícímu k příslušnému měřicímu profilu.

- *správný záznam RZ* – přesný a úplný záznam RZ
- *potenciálně nepřesné RZ* – jedná se o záznamy RZ, u kterých z důvodu nižší kvality referenčního videozáznamu nebylo možné jednoznačně ověřit jejich přesnost a kompletnost
- *nepřesně rozpoznané RZ* – dle referenčního videozáznamu bylo jednoznačně stanoveno, že RZ je softwarově rozpoznána s chybou. V tomto případě byla manuálně zaznamenána správná podoba RZ a původní chybný záznam byl ponechán z důvodu další práce s daty. Pokud z referenčního videozáznamu nebylo možné jednoznačně určit správnou podobu celé RZ, byla tato informace zaznamenána a případně byla zaznamenána doplňující informace ke specifikaci vozidla.
- *záznam neexistující RZ* – jedná se o chybný záznam jiného textu než RZ,
- *opakovaný záznam stejné RZ* – jedná se o opakovaný záznam jednoho průjezdu vozidla v rozmezí nízkých jednotek sekund, tento druh záznamu představuje chybu, ať už byla RZ rozpoznána chybně nebo správně,
- *chybějící záznam průjezdu vozidla* – v tomto případě došlo k manuálnímu doplnění záznamu o průjezdu vozidla – došlo k doplnění údajů RZ, druh vozidla a čas průjezdu. Pokud RZ ve videozáznamu nebyla vizuálně čitelná, nemohlo dojít k jejímu doplnění, namísto RZ byla zaznamenána barva, příp. jiné specifikum vozidla.

Tabulka 6 Shrnutí výsledků kontroly datových sad, ve vztahu k parametru RZ

Měřicí profil	Jednotka	Počet záznamů, procentuální vyjádření							
		Správné	Opačný směr jízdy	Chybné				Potenciálně nepřesná RZ (Nemožné zkontrolovat)	Celkem průjezdů vozidel dle ref. videozáznamu
				Neexistující RZ	Opakovaná detekce stejného vozidla	Nepřesně rozpoznaná RZ	Chybějící záznam		
A1	-	408	93	4	0	11	5	2	426
	%	95,77	-	-	3,76			0,47	100,00
A2	-	195	129	1	0	35	13	1	244
	%	79,92	-	-	19,67			0,41	100,00
B1	-	827	704	20	0	4	5	0	836
	%	98,92	-	-	1,08			0,00	100,00
B2	-	576	846	13	3	6	11	2	598
	%	96,32	-	-	3,34			0,33	100,00
C1	-	186	658	391	1	7	7	0	201
	%	92,54	-	-	7,46			0,00	100,00
C2	-	656	165	20	24	19	0	0	699
	%	93,85	-	-	6,15			0,00	100,00
D1	-	479	407	6	0	5	4	0	488
	%	98,16	-	-	1,84			0,00	100,00
D2	-	420	236	6	1	4	5	0	430
	%	97,67	-	-	2,33			0,00	100,00
E1	-	234	-	6	2	4	11	3	254
	%	92,13	-	-	6,69			1,18	100,00
Celkem	-	3981	3238	467	31	95	61	8	4176
	%	95,33	-	-	4,48			0,19	100,00

Kontrola parametrů Country a Direction

Pro možnou další práci s daty byla zároveň s předchozími typy kontrol prováděna kontrola správnosti parametrů *Country* a *Direction*. Tyto parametry nejsou podstatné pro současný výzkum. Např. co se týče parametru směru jízdy vozidla, pro maximální přesnost byly záznamy tříděny dle směru jízdy manuálně. U parametrů *Country* a *Direction* byla pouze zaznamenávána indikace o jejich správném nebo chybném určení.

Dodatečná kontrola a doplnění datové sady z profilu E1

Pro další kontrolu a doplnění dat k profilu E1 byl vzhledem k reálnému rozmístění měřicích profilů v území využito videozáznamů z profilu A1. Toto doplnění bylo provedeno až ve vztahu k přípravě vstupních dat pro vyhodnocení dle metodiky pro vyhodnocení směrových průzkumů. Z toho důvodu není uvedeno v Tabulce 4.

Vstupní datová sada pro vyhodnocení dle metodiky vyhodnocení směrových průzkumů

Prvním krokem k vytvoření vstupní datové sady pro vyhodnocení dle metodiky pro vyhodnocení směrového průzkumu byla volba časového intervalu, ve kterém bude vyhodnocení probíhat. Byla zvolena 1 hodina a konkrétní časový interval byl zvolen na základě Tabulky 1 tak, aby datové sady ze všech měřicích profilů byly kompletní, bez výpadků.

Vybraný časový interval: 8:10-9:10

Do vstupní datové sady byly řazeny záznamy ze zkontrolované datové sady (viz Tabulka 4) v definovaném časovém intervalu, a to tak, aby ve vstupní datové sadě byly zaznamenány všechny (a pouze tyto) reálné průjezdy vozidel. Konkrétní postup vzhledem ke kvalitativním kategoriím záznamů z Tabulky 4 je následující:

- *vozidlo v opačném směru* – nezařazeny,
- *správný záznam RZ* – zařazeny, bez poznámky,
- *potenciálně nepřesné RZ* – zařazeny, s poznámkou „potencialne nepresna RZ“,
- *nepřesně rozpoznané RZ* – zařazeny (v opravené verzi), s poznámkou ve specifických případech,
- *záznam neexistující RZ* – nezařazeny,
- *opakovaný záznam stejné RZ* – nezařazeny,
- *chybějící záznam průjezdu vozidla* – zařazeny (v doplněné verzi) – s poznámkou ve specifických případech.

Struktura vstupní datové sady je následující – jednotlivé záznamy (průjezdy vozidel) jsou uvedeny v řádcích, ke každému záznamu jsou pak ve sloupcích uvedeny následující podrobnější informace:

- *ID* – jednoznačný identifikátor záznamu (číslo řádku),
- *Licence plate* – RZ, případně známá část RZ nebo prázdné pole v případě, že RZ chybí,
- *Upřesnění* – v případě potenciálně nepřesné nebo nekompletní RZ obsahuje doplňkovou informaci upřesňující vozidlo dle jiných znaků (např. barva), tato informace je určena k manuálnímu čtení, nemá přesně danou strukturu,
- *Poznámka* – v případě potenciálně nepřesné, nepřesné nebo nekompletní RZ obsahuje tuto informaci, může nabývat hodnot:
 - o bez RZ – RZ zcela chybí, manuálně nemohla být určena z referenčního videozáznamu (z důvodu nekvalitního videozáznamu, nekvalitní RZ nebo zcela chybějící RZ na vozidle),
 - o nekompletní RZ – přesně je určena pouze část textu RZ, zbytek textu RZ chybí,
 - o potenciálně nepřesná RZ – z referenčního videozáznamu nebylo možné určit, zda je RZ přesná, nebo ne,
- *Druh vozidla* – druh vozidla v rozlišení na kategorie OA, DOD, LNA, NA, MOTO, BUS
- *Reálný čas* – skutečný čas průjezdu vozidla ve formátu HH:MM:SS
- *LPM confidence* – SW určená hodnota spolehlivosti detekce RZ, neexistuje pro manuálně opravované a doplňované RZ
- *OCR confidence* – SW určená hodnota spolehlivosti rozpoznání RZ, neexistuje pro manuálně opravované a doplňované RZ
- *Image file name* – název souboru SW uloženého snímku vozidla
- *IDProfil* – označení měřicího profilu

Tabulka 7 Kvalitativní přehled dat ve vstupní datové sadě pro vyhodnocení dle metodiky pro vyhodnocení směrových průzkumů

Měřicí profil	Jednotka	Počet záznamů, procentuální vyjádření				
		Správné	Chybné			Celkem
			Chybějící RZ ("Bez RZ")	Chybí část RZ ("Nekompletní RZ")	Potenciálně nepřesná RZ	
A1	-	346	4	2	2	354
	%	97,74	2,26			100,00
A2	-	170	9	1	1	181
	%	93,92	6,08			100,00
B1	-	706	0	0	0	706
	%	100,00	0,00			100,00
B2	-	471	1	2	0	474
	%	99,37	0,63			100,00
C1	-	150	6	1	0	157
	%	95,54	4,46			100,00
C2	-	576	0	0	0	576
	%	100,00	0,00			100,00
D1	-	384	0	0	0	384
	%	100,00	0,00			100,00
D2	-	341	0	1	0	342
	%	99,71	0,29			100,00
E1	-	199	1	0	0	200
	%	99,50	0,50			100,00
Celkem	-	3343	21	7	3	3374
	%	99,08	0,92			100,00

7.4 Spárování dat dle Metodiky vyhodnocení směrových průzkumů

V tomto kroku byla využita Metodika vyhodnocení směrových dopravních průzkumů (78), představené v diplomové práci autorky. Konkrétně byla vstupní datová sada zpracována v jazyce R dle následujících dílčích kroků Metodiky:

- Sloučení dat ze všech měřicích profilů do 1 souboru,
- Párování dat,
- Sestavení Matice spárovaných dat,

- Statistické zhodnocení doby jízdy vozidel mezi po sobě následujícími profily a stanovení limitní doby jízdy,
- Sestavení cest vozidel,
- Tvorba O-D matice.

Získané cesty vozidel představují ucelené informace o reálném pohybu vozidel ve zkoumané oblasti v době měření. Sada cest je sestavována tak, že každá získaná cesta odpovídá posloupnosti záznamů RZ vozidla ve zkoumané oblasti po dobu, kdy vozidlo nepřerušilo jízdu z důvodu cíle své cesty. Zda se vozidlo v relaci dvou návazných měřicích profilů pohybovalo bez přerušení jízdy z důvodu cíle cesty (tzn. nikoliv z důvodu vynucených zastavení vozidla podmínkami dopravního proudu a dopravní infrastruktury) je dáno stanovenými limitními dobami jízdy.

Datová sada cest, Matice spárovaných dat a O-D matice, které jsou výsledky procesu spárování dat dle Metodiky, přinášejí důležité poznatky o vstupních datech a o chování vozidel v dané oblasti z pohledu jejich směrovosti. Tyto poznatky, zejména poměr spárovaných a nespárovaných dat a výskyt cest s potenciálně nelogickými trasami, byly využity v dalším kroku – při opětovné kontrole dat.

7.5 Opětovná kontrola spárovaných dat dle referenčních videozáznamů, oprava a doplnění

Opětovná kontrola dat probíhala dle vytvořené datové sady cest. Kontrolovány byly záznamy z cest, jejichž výskyt vzhledem k povaze zkoumané oblasti (uzavřená oblast s malým množstvím parkovacích ploch) mohl představovat chybu. Konkrétně byly kontrolovány:

- nespárované záznamy,
- spárované záznamy, u nichž byl v rámci jedné cesty identifikován pouze jeden měřicí profil,
- spárované záznamy tvořící cestu v neočekávané relaci vzhledem k topologii silniční sítě (v případě využívané datové sady konkrétně relace A1-C2 a A1-D2, kde byl silný předpoklad, že mezi těmito měřicími profily budou vozidla zaznamenána ještě na profilu E1 (viz mapa na Obrázku 1).

Kontrola probíhala formou manuálního dohledávání v datové sadě cest potenciálně spárovatelných záznamů. Ke každému kontrolovanému záznamu byly hledány potenciálně spárovatelné záznamy, které byly identifikovány pomocí podobnosti RZ a logickému výskytu vzhledem k měřicímu profilu a času záznamu. Takto nalezené potenciálně spárovatelné záznamy byly ověřovány dle zdrojových videozáznamů.

Výsledkem opětovné kontroly bylo objevení celkem 101 chyb ve vstupní datové sadě, které v datech zůstaly navzdory dříve již provedené kontrole. Týkaly se zejména chybného rozpoznání registračních značek z referenčních videozáznamů. Tabulka 8 níže zobrazuje v přehledné formě reflexi nově stanovené kvality původní vstupní datové sady z Tabulky 7.

Všechny nově identifikované chyby se s pomocí zdrojových záznamů podařilo opravit. Oprava proběhla ve vstupní datové sadě, byla tedy vytvořena nová referenční vstupní datová sada, která byla již považována za maximálně správnou. Při opravě nebyl kladen důraz na 100% přesné určení reálného textu RZ u vozidel, u nichž ani na jednom z měřicích profilů, kde bylo vozidlo zaznamenáno, kvalita zdrojového videozáznamu neumožňovala přesný text RZ určit. U těchto vozidel byl kladen důraz na přiřazení všech záznamů vozidla v době měření stejnému textu RZ. Toto umožní stanovení správných

cest vozidel po zkoumané oblasti, třebaže RZ, která je vozidlu v průběhu vyhodnocení přiřazena, nemusí odpovídat jejímu reálnému tvaru.

Tabulka 8 Chybovost původní zdrojové datové sady po opravě dle datové sady cest

Měřicí profil	Jednotka	Počet záznamů, procentuální vyjádření		
		Celkový počet chyb po kontrole dle sady Cest	Celkové % "Správně" po kontrole dle sady Cest	Celkem
A1	-	18	337	355
	%	5,07	94,93	100,00
A2	-	52	134	186
	%	27,96	72,04	100,00
B1	-	4	702	706
	%	0,57	99,43	100,00
B2	-	5	468	473
	%	1,06	98,94	100,00
C1	-	13	143	156
	%	8,33	91,67	100,00
C2	-	2	574	576
	%	0,35	99,65	100,00
D1	-	1	383	384
	%	0,26	99,74	100,00
D2	-	2	341	343
	%	0,58	99,42	100,00
E1	-	4	197	201
	%	1,99	98,01	100,00
Celkem	-	101	3279	3380
	%	2,99	97,01	100,00

8 Analýza SW spolehlivostních parametrů

Účelem bylo navrhnout postup, jak analyzovat a prakticky využít parametry související se spolehlivostí SW rozpoznání registrační značky z obrazového záznamu. Tyto parametry mohou být poskytovány rozpoznávacími SW. Cílem bylo s využitím konkrétních dat poskytovaných SW Eyedea navrhnout obecný postup, který by byl přenositelný na celou řadu obdobných SW.

SW Eyedea poskytuje k rozpoznávanému textovému řetězci *Licence plate* dva spolehlivostní parametry:

- *LPMconfidence* – Faktor spolehlivosti detekce RZ (rozsah 0-1)
- *OCRconfidence* – Faktor spolehlivosti rozpoznání textu RZ (rozsah 0-1)

Oba parametry nabývají spojitých číselných hodnot v rozsahu 0–1, což lze převést na procenta jako 0%–100%. Zároveň je třeba si uvědomit, že ani jeden z parametrů sám o sobě neposkytuje informaci o tom, zda rozpoznávaný textový řetězec *Licence plate* odpovídá realitě – tzn. zda reprezentuje skutečný průjezd vozidla se správně rozpoznanou registrační značkou. Parametr *LPMconfidence* pouze určuje pravděpodobnost, s jakou je rozpoznávaný záznam skutečně součástí registrační značky. Parametr *OCRconfidence* pak určuje, zda byl správně rozpoznán textový řetězec.

Cílem bylo s využitím referenční datové sady tyto parametry analyzovat a navrhnout vhodný postup ke stanovení mezních hodnot obou parametrů, které s předem danou pravděpodobností určí, že SW poskytnutý záznam RZ lze považovat za odpovídající realitě (reprezentující reálný průjezd vozidla a správný tvar skutečné registrační značky).

8.1 Popis datové sady pro vyhodnocení spolehlivostních parametrů

Do datové sady pro vyhodnocení parametrů spolehlivosti byly zařazeny všechny záznamy z výsledné referenční vstupní datové sady (viz Tabulka 8), pro které platí:

- jsou k dispozici SW hodnoty parametrů *LPMconfidence* a *OCRconfidence*
- v průběhu kontroly bylo jednoznačně určeno, zda byla SW vyhodnocená hodnota parametru *Licence plate* správná, nebo chybná.

Při řazení do datové sady pro vyhodnocení parametrů spolehlivosti byla data rozřazena do dvou skupin, tedy do skupiny „správně“ a skupiny „špatně“. Ve skupině „správně“ se nachází správně rozpoznané záznamy RZ. Ve skupině „špatně“ se nachází chybně určené záznamy RZ, ať už nepřesně rozpoznané RZ nebo případně opakované detekce projíždějících vozidel.

Do datových sad nebyla použita data vozidel s potenciálně nepřesnými RZ, vozidla v opačném směru, protože pro tato vozidla nebyla prováděna důsledná kontrola správnosti záznamů RZ. Rovněž nebyly využity záznamy nedetekovaných průjezdů vozidel, a to z důvodu absence parametrů *LPMconfidence* a *OCRconfidence* a jejich ne vždy správně čitelné RZ. Dále nebyly využity detekce různých textů nesouvisejících s RZ.

Přehledný souhrn je zde:

- *vozidlo v opačném směru* – nezařazeny,
- *správný záznam RZ* – zařazeny, do skupiny „správně“,
- *potenciálně nepřesné RZ* – nezařazeny,

- nepřesně rozpoznané RZ – zařazeny, do skupiny „špatně“,
- záznam neexistující RZ – nezařazeny,
- opakovaný záznam stejné RZ – zařazeny, do skupiny „špatně“
- chybný záznam průjezdu vozidla – nezařazeny

Následující Tabulka 9 zobrazuje souhrn počtů zařazených záznamů, dle jednotlivých měřicích profilů a celkem.

Tabulka 9 Popis počtu záznamů

	Měřicí profil								
	A1	A2	B1	B2	C1	C2	D1	D2	E1
Počet záznamů využitých pro vyhodnocení parametrů spolehlivosti	419	229	831	578	193	699	484	425	241
Počet záznamů ve skupině „správně“	402	174	829	567	184	656	475	421	231
Počet záznamů ve skupině „špatně“	17	55	2	11	9	43	9	4	10
Celkový počet záznamů využitých pro vyhodnocení parametru spolehlivost	4108								
Celkový počet „Správně“	3948								
Celkový počet ve skupině „špatně“	160								

Datová sada pro vyhodnocení SW spolehlivostních parametrů byla připravena ve struktuře - jednotlivé záznamy (průjezdy vozidel) jsou uvedeny v řádcích, ke každému záznamu jsou pak ve sloupcích uvedeny následující podrobnější informace:

- *LP_R* – rozpoznání tvar RZ
- *LP_correct* – referenční (správný) tvar RZ
- *LPM confidence* – SW určená hodnota spolehlivosti detekce RZ
- *OCR confidence* – SW určená hodnota spolehlivosti rozpoznání RZ
- *ID profil* – označení měřicího profilu
- *ref* – referenční hodnota shody rozpoznání a referenčního tvaru RZ

8.2 Zobrazení datové sady pro vyhodnocení spolehlivostních parametrů

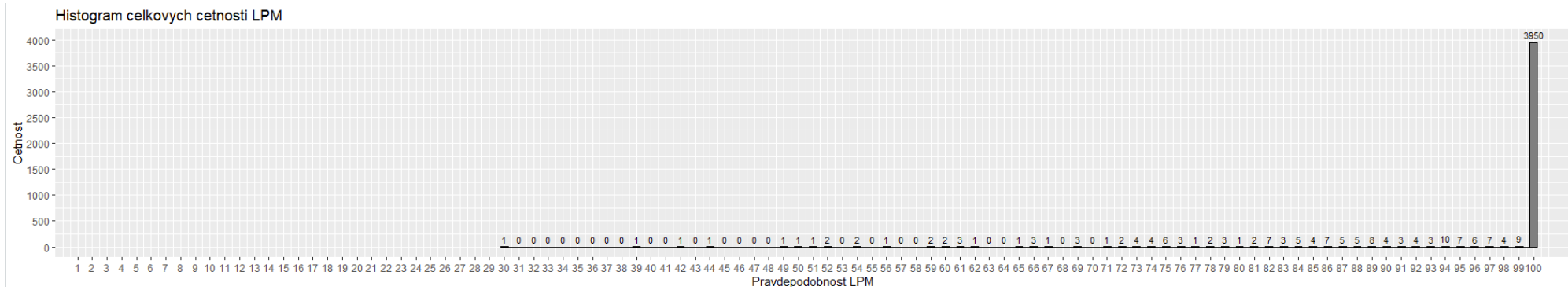
Pro bližší představu o obsahu dat, zejména o počtech záznamů vzhledem k vyhodnocovaným SW parametrům, byla datová sada znázorněna v sérii histogramů. Histogramy (grafy) zobrazují četnosti záznamů vzhledem k hodnotám parametrů *LPMconfidence* a *OCRconfidence*. Pro přehlednější práci a zobrazení byly hodnoty parametrů převedeny na škálu 0–100 a zaokrouhleny na celá čísla.

Grafy 1 a 2 zobrazují celkové počty záznamů, grafy 3 a 4 zobrazují počty záznamů, které byly SW rozpoznány správně (s referenční hodnotou *ref* == TRUE), grafy 5 a 6 zobrazují počty záznamů, které byly SW rozpoznány chybně (s referenční hodnotou *ref* == FALSE).

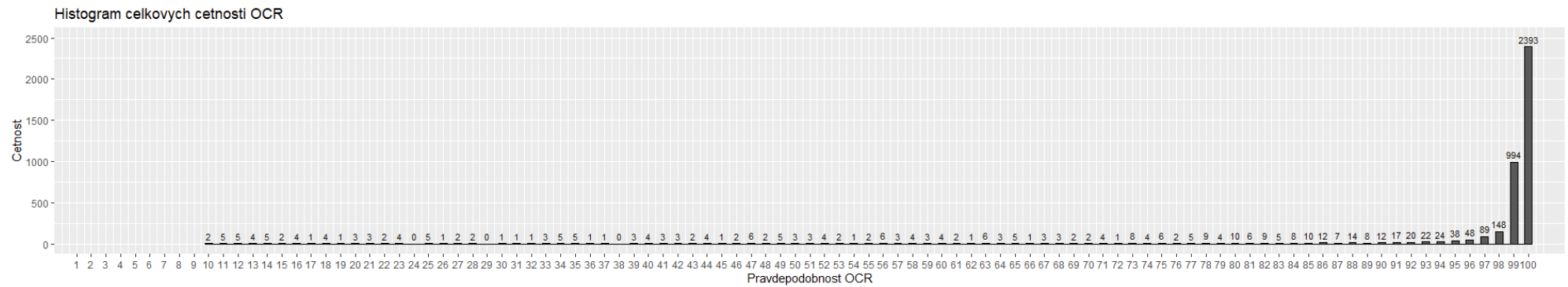
Z histogramů lze vidět, že naprostá většina záznamů se nachází v intervalu zaokrouhlených hodnot parametrů *LPMconfidence* 100 a *OCRconfidence* 99 a 100.

Pro zobrazení referenčních záznamů v závislosti na obou vyhodnocovaných spolehlivostních parametrech najednou byl vytvořen bodový graf 7. Jednotlivé body reprezentují jednotlivé záznamy z datové sady a jejich barva vyjadřuje referenční hodnotu, zda byly SW rozpoznány správně (modrá barva) či chybně (červená barva).

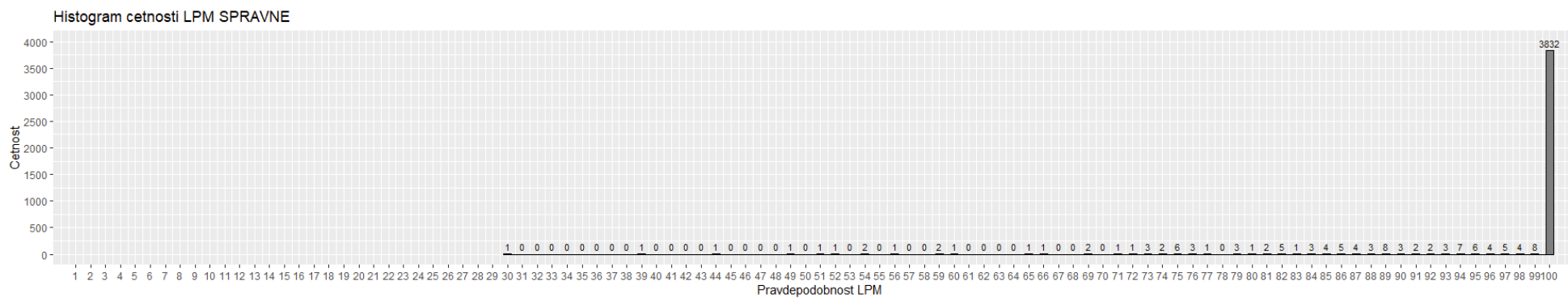
Všechny grafy vyjadřují velmi nerovnoměrné rozložení referenčních záznamů vzhledem k hodnotám *LPMconfidence* a *OCRconfidence*. Toto je důsledkem způsobu získání referenční datové sady, kdy šlo o směrový průzkum, při kterém bylo cílem pořídit co nejkvalitnější videozáznamy pro následné zpracování dat. Je proto logické, že nejvíce záznamů se vyskytuje s vysokými hodnotami spolehlivostních parametrů.



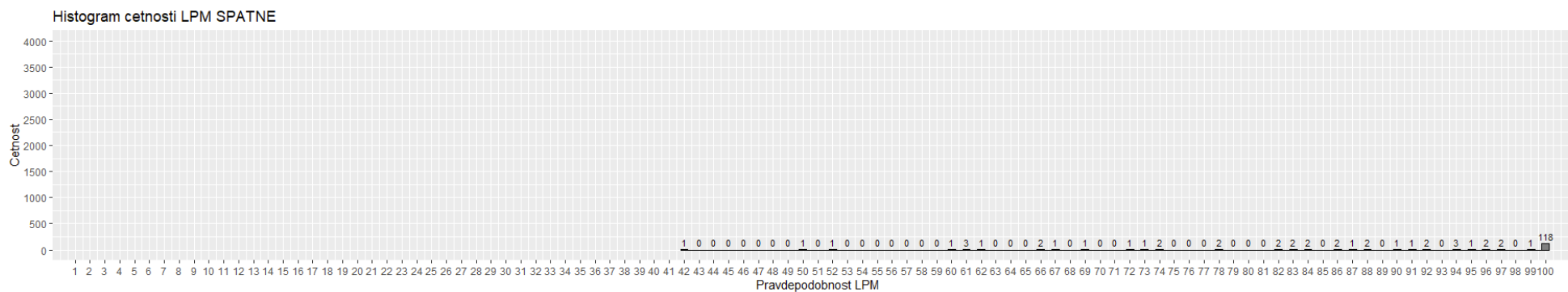
Graf 1 Histogram celkových četností záznamů dle parametru LPMconfidence



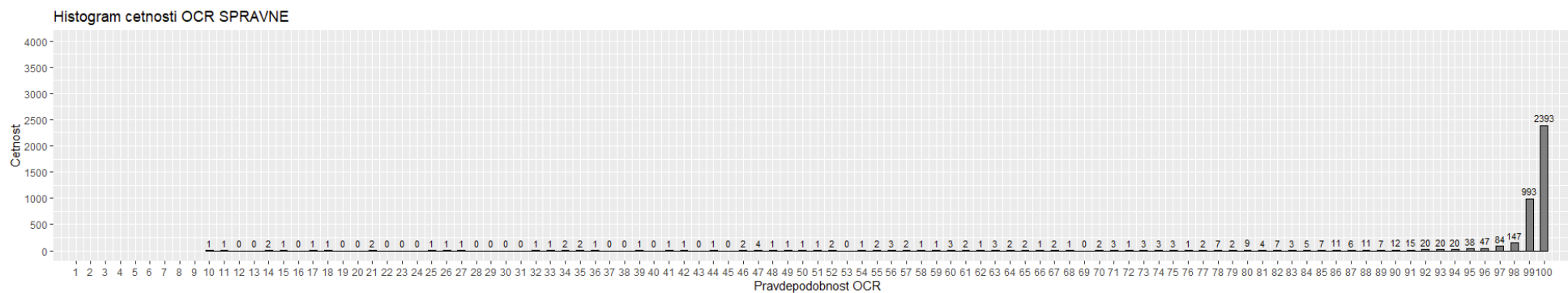
Graf 2 Histogram celkových četností záznamů dle parametru OCRconfidence



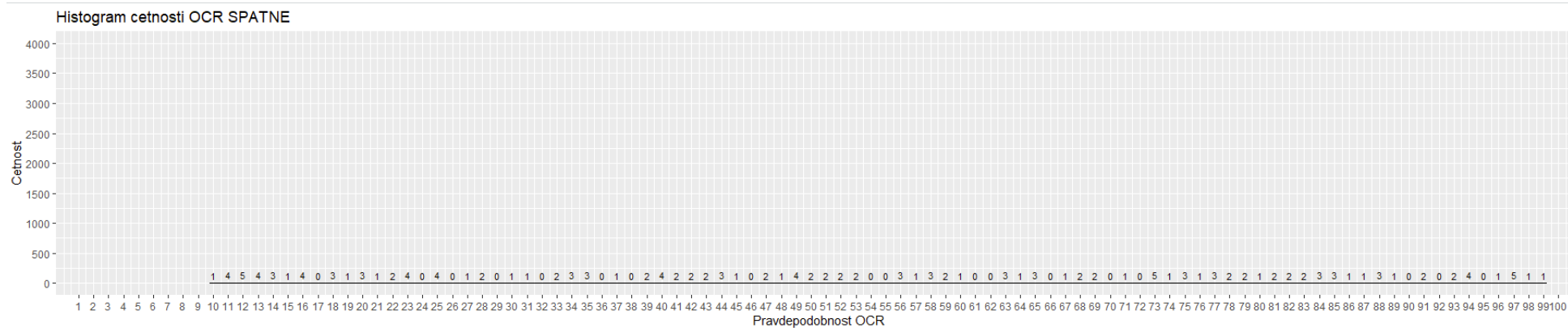
Graf 3 Histogram četností správných záznamů dle parametru LPMconfidence



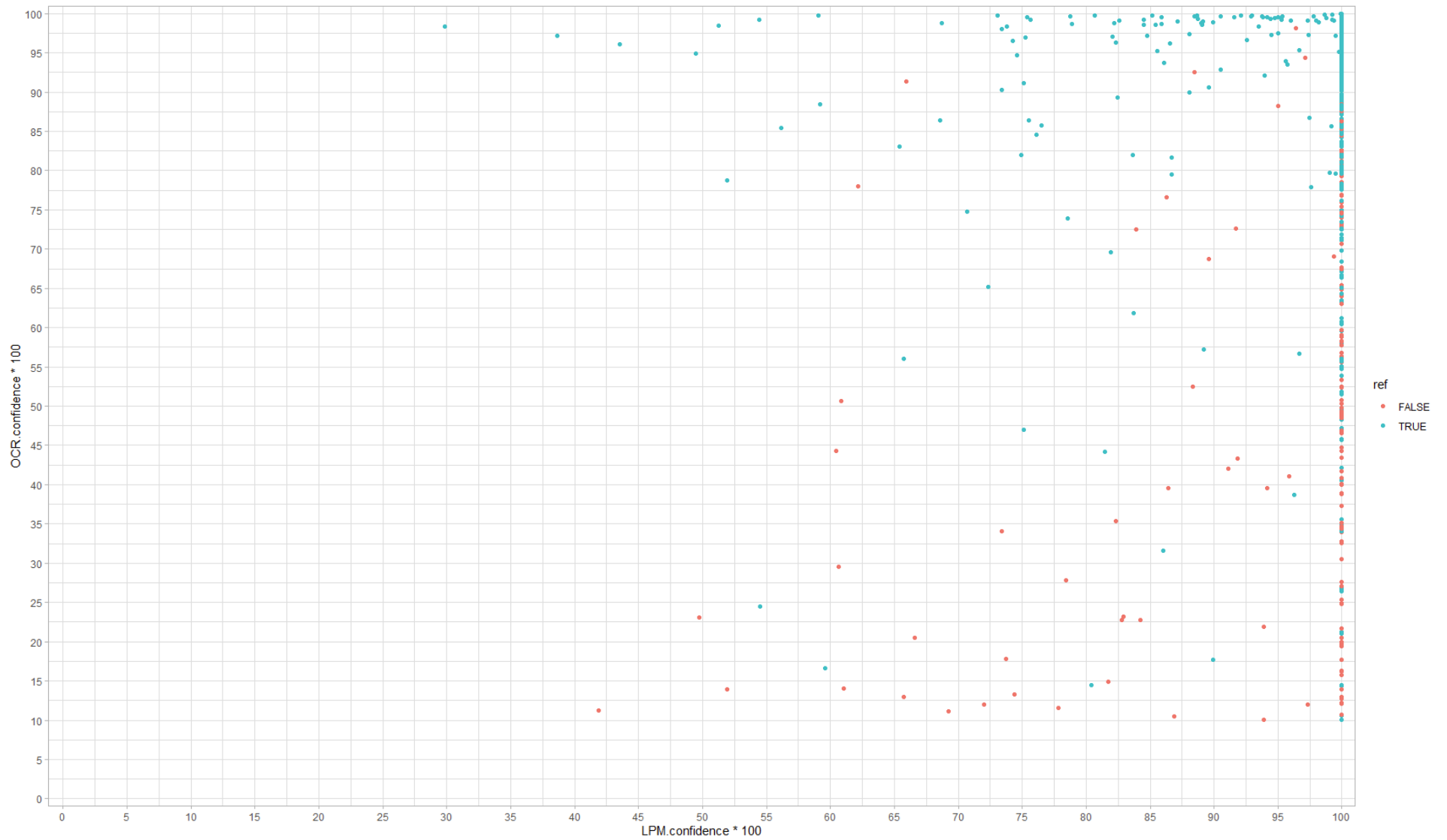
Graf 4 Histogram četností chybných záznamů dle parametru LPMconfidence



Graf 5 Histogram četností správných záznamů dle parametru OCRconfidence



Graf 6 Histogram četností chybných záznamů dle parametru OCRconfidence



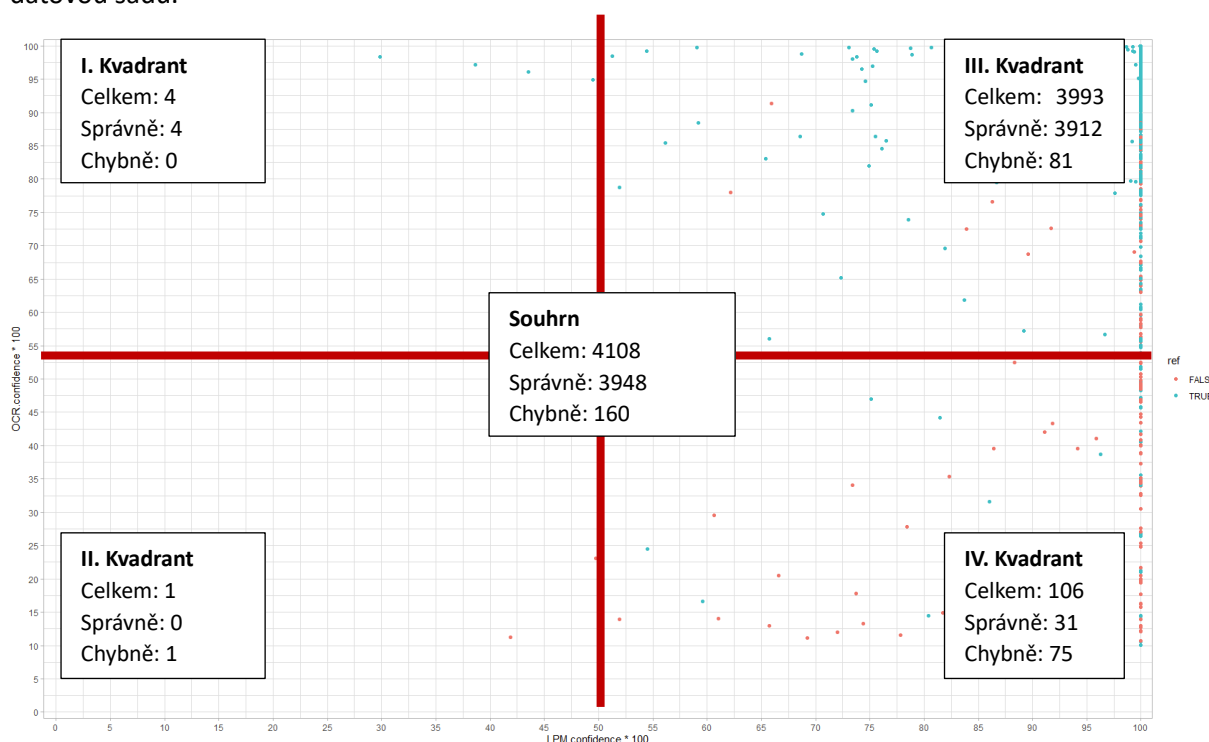
Graf 7 Grafické zobrazení správně/chybně rozpoznaných záznamů ve vztahu k oběma parametrům LPMconfidence a OCRconfidence

8.3 Vyhodnocení spolehlivostních parametrů

Tři základní otázky, které byly v průběhu analýzy SW spolehlivostních parametrů řešeny, jsou následující:

1. Máme-li nový záznam registrační značky „RZ“, kterému byl SW přiřazeny hodnoty pravděpodobnosti $LPMconfidence = X$ a $OCRconfidence = Y$ ($X, Y \in <0;1>$). Jaká je pravděpodobnost, že RZ je správně? Tzn. jedná se o reálný průjezd vozidla se správně rozpoznanou registrační značkou. Tzn. $p(RZ = „správně“) = ?$ (Nebo obráceně: $p(RZ = „chybně“) = ?$, přičemž platí: $p(RZ = „správně“) = 1 - p(RZ = „chybně“)$)
2. Jaká je pravděpodobnost správnosti výše uvedeného odhadu? Tzn. $p(p(RZ = „správně“) = „správně“) = ?$ Neboli, jak je výše uvedený odhad spolehlivý?
3. Pro jaký interval hodnot $LPMconfidence$ a $OCRconfidence$ je odhad spolehlivosti stanoven?

Tyto tři otázky lze ilustrovat následujícím příkladem. Graf 7 je rozdělen na čtyři kvadranty, přičemž hranice mezi kvadranty jsou tvořeny v hodnotách $LPMconfidence = 0,5$ (50 %) a $OCRconfidence = 0,5$ (50 %). Celkové četnosti záznamů v jednotlivých kvadrantech jsou uvedeny v příslušných textových polích, stejně tak je uvedeno jejich rozdělení dle referenčních hodnot na SW správně a chybně rozpoznané záznamy. Zároveň ve středovém textovém poli jsou uvedeny souhrnné hodnoty pro celou datovou sadu.



Obrázek 3 Vyhodnocení spolehlivostních parametrů – ilustrativní příklad

Vypočítané hodnoty pravděpodobností dle výše uvedených základních otázek, jsou k vidění v tabulce níže:

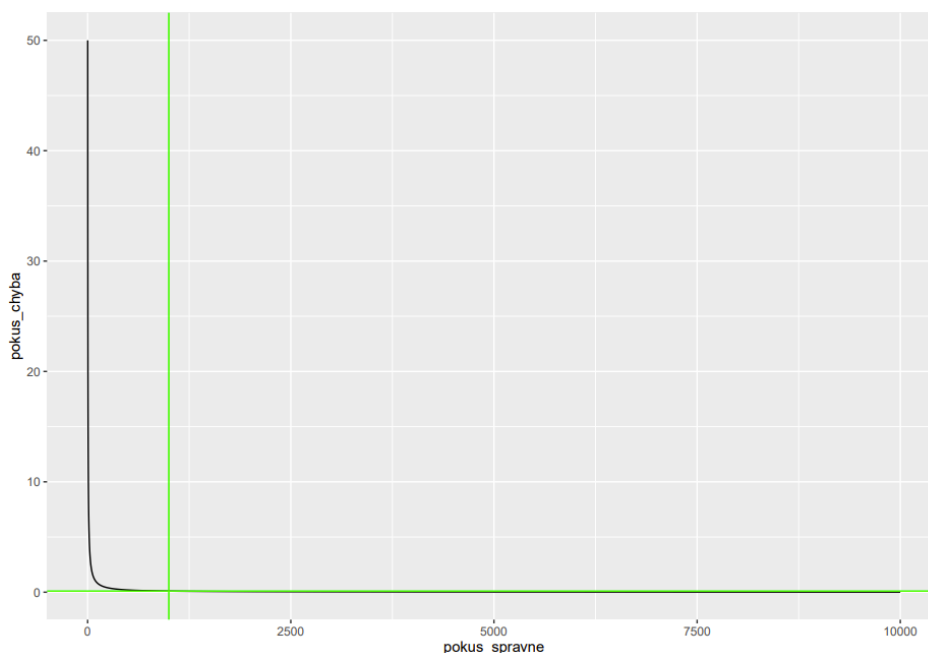
Tabulka 10 Vyhodnocení spolehlivostních parametrů – ilustrativní příklad

	p(RZ = „správně“)	p(RZ = „chybně“)	Spolehlivost odhadu
I. Kvadrant (interval LPM <0;50>, OCR (50;100>	100,00 %	0,00 %	nízká
II. Kvadrant (interval LPM <0;50>, OCR (0;50>	0,00 %	100,00 %	nejnižší
III. Kvadrant (interval LPM <50;100>, OCR (50;100>	97,97 %	2,03 %	velmi vysoká
IV. Kvadrant (interval LPM <50;100>, OCR (0;50>	29,25 %	70,75 %	vysoká
Souhrn (interval LPM <0;100>, OCR (0;100>	96,11 %	3,89 %	nejvyšší

Výše uvedený příklad lze interpretovat tak, že pokud nový záznam registrační značky RZ1 bude mít hodnoty spolehlivostních parametrů $LPMconfidence = 0,65$ (65 %) a $OCRconfidence = 0,80$ (80 %), lze říci, že na souhrnné úrovni je tento záznam s pravděpodobností 96,11 % správný a tento odhad je velmi spolehlivý. Konkrétněji lze specifikovat, že na úrovni kvadrantů je RZ1 s pravděpodobností 97,97 % správně, a tento odhad je opět velmi spolehlivý. Na rozdíl od toho, pokud nový záznam registrační značky RZ2 bude mít hodnoty spolehlivostních parametrů $LPMconfidence = 0,35$ (35 %) a $OCRconfidence = 0,20$ (20 %), lze říci, že na souhrnné úrovni je tento záznam opět s pravděpodobností 96,11 % správný a tento odhad je velmi spolehlivý. Konkrétněji lze pak ale specifikovat, že na úrovni kvadrantů je RZ1 s pravděpodobností 100 % chybný, ale tento odhad je velmi nespolehlivý.

Za účelem konkrétního určení spolehlivosti odhadu bylo využito faktu, že odhad je pouze natolik spolehlivý, kolik je k dispozici dat, ze kterých byl odhad určen. Obecně lze říci, že množství dat, které je potřeba k tomu, aby z nich bylo možné stanovit dostatečně spolehlivý odhad, je dáno účelem, pro který budou data dále využívána. Tedy, neexistuje obecně platné číslo určující potřebné množství dat na stanovení dostatečně spolehlivého odhadu. Hranici je vždy potřeba hledat na základě účelu celého problému – k čemu budou data dále využita a jak spolehlivá data to cílové využití potřebuje.

Zároveň však lze říci, že čím větší je trénovací množina dat, tím je odhad vyvozovaný na základě této množiny spolehlivější. Přičemž spolehlivost tohoto odhadu neroste lineárně, ale nejdříve prudce narůstá a od určitého okamžiku se naopak již téměř nezvyšuje. Tento vývoj ilustruje následující graf 8, kde na ose x je počet trénovacích dat, na ose y je pak procentuální pravděpodobnost chyby odhadu stanoveného na základě trénovacích dat.



Graf 8 Vývoj pravděpodobnosti chyby odhadu v závislosti na velikosti trénovací množiny

Pro sestavení grafu 8 byla využita funkce $y = 1/(x+1)*100$. V tomto případě platí, že když $x \geq 100 \Rightarrow y < 1$, tzn. chyba odhadu je nižší než 1 %. Zároveň, když $x \geq 1000 \Rightarrow y < 0,1$, tzn. chyba odhadu je nižší než 0,1 %. Uvedený předpis vychází z předpokladu, že odhad je stanovován na základě x známých hodnot. Tedy první další hodnota je již neznámá a může vnést do stanoveného odhadu chybu.

Ve vztahu k trénovací množině záznamů registračních značek vozidel lze předpis $y = 1/(x+1)*100$ interpretovat následovně. Pokud trénovací množina obsahuje 1000 záznamů a na základě této množiny je stanovena pravděpodobnost $p(RZ = \text{„správně“}) = 80 \%$, tak pravděpodobnost chyby tohoto odhadu je 0,09 %. Pokud na základě takto natrénovaných hodnot bude vyhodnocována nová sada záznamů, 80 % z nich bude přiřazena hodnota „správně“, tedy budou vyhodnoceny jako reálný průjezd vozidel se správně rozpoznanou registrační značkou softwarem. 20 % záznamů bude přiřazena hodnota „chybně“. 0,09% je pravděpodobnost, že poměr 80:20 byl stanoven chybně.

Tabulka níže zobrazuje vyčíslené hodnoty spolehlivosti odhadu z tabulky 11 výše. Je třeba brát v úvahu, že všechny hodnoty uvedené v tabulce platí vždy pro celý vymezený interval hodnot parametrů *LPMconfidence* a *OCRconfidence*.

Tabulka 11 Vyhodnocení spolehlivostních parametrů – ilustrativní příklad, závěr

	$p(RZ = \text{„správně“})$	$p(RZ = \text{„chybně“})$	Spolehlivost odhadu
I. Kvadrant (interval LPM <0;50>, OCR (50;100>))	100,00 %	0,00 %	80 %
II. Kvadrant (interval LPM <0;50>, OCR (0;50>))	0,00 %	100,00 %	50 %
III. Kvadrant (interval LPM <50;100>, OCR (50;100>))	97,97 %	2,03 %	99,97 %
IV. Kvadrant (interval LPM <50;100>, OCR (0;50>))	29,25 %	70,75 %	99,07 %
Souhrn (interval LPM <0;100>, OCR (0;100>))	96,11 %	3,89 %	99,98 %

Vzhledem k nerovnoměrnému rozložení dat v trénovací množině záznamů vzhledem k hodnotám parametrů *LPMconfidence* a *OCRconfidence* (viz graf 7) bylo třeba stanovit, jaká spolehlivost odhadu je požadována a tedy, kolik dat v trénovací množině je potřeba na dostatečně spolehlivé určení pravděpodobnosti $p(RZ = \text{„správně“})$.

Určováním konkrétních potřebných spolehlivostí vyhodnocených LPR dat vzhledem k jejich návaznému využití, se výzkum v tomto roce nezabýval (jedná se o činnost naplánovanou do následujícího roku). Proto zde byla modelově zvolena hodnota požadované spolehlivosti 99,9 %. Toto odpovídá chybnému určení 1 z 1000 záznamů. Tato mez je v grafu 8 vyznačena zelenými přímkami.

Prakticky toto znamená, že jsou hledány minimální hodnoty *LPMconfidence* (X) a *OCRconfidence* (Y), pro které platí, že: pokud záznam registrační značky $RZ \in (X;100\%) \wedge (Y;100) \Rightarrow p(RZ = \text{„správně“}) \geq 99,9\%$. Obecně lze říct, že nemusí jít o celočíselné mezní hodnoty X, Y, ale o mezní křivku danou funkcí $f(y,x)$.

Zároveň požadavek na spolehlivost 99,9 % určuje, že spolehlivost odhadu poměru správně a chybně rozpoznávaných RZ, který byl stanoven na 99,9:0,1, musí být rovněž větší nebo rovna 99,9 %.

Tyto požadavky určují, že je třeba mít minimálně 1000 trénovacích záznamů. Vzhledem k rozložení četností záznamů v datové sadě (viz grafy 1-7, nicméně grafy zobrazují zaokrouhlené hodnoty) je tato podmínka splněna pouze pro hodnoty parametru *LPMconfidence* = 100 %. Podrobněji byl analyzován tedy pouze parametr *OCRconfidence*, a to ve dvou krocích. V prvním kroku byl vzhledem k tomu, že *OCRconfidence* je spojitá veličina, stanoven nejmenší interval tohoto parametru, kdy počet záznamů v trénovací datové sadě přesáhl 1000 a zároveň poměr správných a chybných záznamů byl nanejvýš 99,9:0,1. Interval byl hledán od maximální hodnoty *OCRconfidence* směrem k nižším hodnotám a byl stanoven na $\langle 99,75\% ; 100\% \rangle$. Počet záznamů RZ v tomto intervalu byl 1047 a všechny byly „správné“. Následně byla mez *OCRconfidence* snižována, vždy o hodnotu 0,25 %, přičemž počet záznamů v intervalu $\langle Y; 100\% \rangle$ narůstal a zároveň se zvyšoval počet záznamů s příznakem „chybné“. Za mez parametru *OCRconfidence* byla určena minimální hodnota Y, při které poměr „správných“ a „chybných“ záznamů nepřesáhl 99,9:0,1. Tímto postupem byla určena mez parametru *OCRconfidence* jako 97%.

Závěrem vyhodnocení spolehlivostních parametrů pro konkrétní trénovací datovou sadu rozpoznanou SW Eyedea jsou meze spolehlivostních parametrů *LPMconfidence* = 100 % a *OCRconfidence* jako 97 %. Tyto meze zajišťují, že pokud záznam registrační značky RZ má hodnoty *LPMconfidence* = 100 % a zároveň *OCRconfidence* $\geq 97\%$, pak $p(RZ = \text{„správně“}) = 99,91\%$. Hodnota spolehlivosti tohoto odhadu je (při počtu 3527 záznamů) 99,97 %.

9 Využití pravděpodobnostní matice

Cílem je vytvořit nástroj, který umožní porovnávat textové řetězce registračních značek z různých záznamů a bez znalosti reálného (správného) tvaru registrační značky určit, zda porovnávaná dvojice záznamů patří ke stejnému vozidlu či nikoliv – tedy, zdá má být spárována.

Pro dosažení uvedeného cíle byl využit postup prezentovaný v článku [zdroj - článek], který představil váhovou funkci založenou na pravděpodobnostním modelu. Tato funkce umožňuje párovat záznamy registračních značek získaných z LPR systému o dvou měřicích profilech. Výzkum prezentovaný v [zdroj – článek] zohledňuje při práci s editační vzdáleností dvou řetězců znaků fakt, že při rozpoznávání textů v LPR systémech nedochází k reverzním chybám (tzn. k přehazování znaků v textovém řetězci). Dále využívá údajů o čase záznamu, které jsou v LPR systémech k dispozici. Autoři v [zdroj - článek] byli svojí metodou schopni dosáhnout 97% úspěšnosti při párování záznamů vozidel s 0% falešně spárovaných záznamů.

Tento výzkum se věnuje aplikaci, aktualizaci a rozšíření výše uvedené metody na LPR systém s libovolným počtem měřicích profilů.

V textu této kapitoly se pracuje s následujícími zkratkami:

- T ... reálná podoba registrační značky (z anglického true value)
- t_i ... reálná podoba jednotlivého znaku registrační značky
- X ... SW rozpoznaná podoba registrační značky na měřicím profilu X (libovolném měřicím profilu LPR systému)
- x_i ... SW rozpoznaná podoba jednotlivého znaku registrační značky na měřicím profilu X (libovolném měřicím profilu LPR systému)
- Y ... SW rozpoznaná podoba registrační značky na měřicím profilu Y (libovolném měřicím profilu LPR systému, různém od X)
- y_i ... SW rozpoznaná podoba jednotlivého znaku registrační značky na měřicím profilu Y (libovolném měřicím profilu LPR systému, různém od X)

9.1 Popis datové sady pro sestavení pravděpodobnostní matice

Do datové sady pro vyhodnocení parametrů spolehlivosti byly zařazeny všechny záznamy z výsledné referenční vstupní datové sady (viz Tabulka 8), pro které platí:

- záznamy byly SW rozpoznané,
- v průběhu kontroly bylo jednoznačně určeno, zda byla SW vyhodnocená hodnota parametru *Licence plate* správná, nebo chybná.

Přehledný souhrn je zde:

- *vozidlo v opačném směru* – nezařazeny,
- *správný záznam RZ* – zařazeny,
- *potenciálně nepřesné RZ* – nezařazeny,
- *nepřesně rozpoznané RZ* – zařazeny,
- *záznam neexistující RZ* – nezařazeny,
- *opakovaný záznam stejné RZ* – zařazeny,
- *chybějící záznam průjezdu vozidla* – nezařazeny.

Při řazení do datové sady pro sestavení pravděpodobnostní matice nebyly zařazeny záznamy, u kterých při SW rozpoznání došlo k vynechání nebo naopak doplnění některého znaku, tedy počet znaků SW rozpoznávaného záznamu a referenčního reálného tvaru RZ se neshodoval. Případně byly zařazeny jen ty části záznamů, u nichž se počet znaků s referenčním tvarem RZ shodoval. Vzhledem k tomu, že tento problém se týkal minima záznamů z výše uvedených zařazených skupin, nepředstavovalo to významné zanedbání.

Zároveň bylo přistoupeno k plošnému smazání znaků pomlček „-“ ze všech záznamů RZ. K tomuto opatření bylo přistoupeno vzhledem k nespolehlivé detekci pomlček rozpoznávacím SW Eyedea, což vedlo k rozdílnému počtu znaků v sadě rozpoznávaných RZ oproti referenčním tvarům RZ.

Datová sada pro vyhodnocení SW spolehlivostních parametrů byla připravena ve struktuře - jednotlivé záznamy (průjezdy vozidel) jsou uvedeny v řádcích, ke každému záznamu jsou pak ve sloupcích uvedeny následující informace:

- *LP_R* – rozpoznávaný tvar RZ,
- *LP_correct* – referenční (správný) tvar RZ.
- *ID profil* – označení měřicího profilu
- *ref* – referenční hodnota shody rozpoznávaného a referenčního tvaru RZ

9.2 Matice $p(y|x)$ – sestavení a využití

Centrem metody je matice podmíněných pravděpodobností, která pro jednotlivé možné znaky registračních značek ukazuje pravděpodobnosti, s jakou byl znak t_i , který byl na měřicím profilu X rozpoznán ve tvaru x_i , rozpoznán na měřicím profilu Y ve tvaru y_i . Přitom reálná podoba znaku t_i není známá.

Pro sestavení pravděpodobnostní matice byly všechny registrační značky v datové sadě rozděleny na jednotlivé znaky, aby tyto znaky byly pak vzájemně porovnávány.

Sestavovaná pravděpodobnostní matice má rozměry 37x37. Obsahuje všechna písmena anglické abecedy, která odpovídá písmenům českých registračních značek. Zároveň obsahuje číslovky 0–9. Postup sestavování respektuje postup uvedený v článku XY, nicméně nezařazoval prázdný znak, který v postupu uvedeném v [zdroj - článek] symbolizoval vynechání nebo naopak doplnění znaku při rozpoznávání SW. Zároveň byl zohledněn fakt, že na všech měřicích profilech je pro měření videozáznamů a následné rozpoznání registračních značek využívána stejná technologie, stejné HW nahrávací zařízení a stejný rozpoznávaný SW.

Postup byl následující:

1. Matice počtů správně a špatně rozpoznávaných znaků
2. Matice $p(t|x)$
3. Matice $p(x|t)$
4. Matice $p(y|x)$

Matice počtů správně a špatně rozpoznávaných znaků je tvaru 37x37, přičemž řádky reprezentují SW rozpoznávaný znak a sloupce reprezentují správný (reálný) tvar znaku daný referenční datovou sadou. Čísla v buňkách matice reprezentují prosté počty znaků v datové sadě. Na závěr byly do matice doplněny jednotlivých řádků a sloupců.

Matice $p(t|x)$ je stejně jako v předchozím případě je tvaru 37x37, přičemž řádky reprezentují SW rozpoznávaný znak a sloupce reprezentují správný (reálný) tvar znaku daný referenční datovou sadou. Čísla v buňkách matice reprezentují podmíněnou pravděpodobnost, že znak registrační značky rozpoznávaný jako x_i má reálnou hodnotu t_i . Součty hodnot v jednotlivých řádcích matice jsou rovny jedné.

Matice $p(x|t)$ je stejně jako v předchozím případě je tvaru 37x37, přičemž řádky reprezentují SW rozpoznávaný znak a sloupce reprezentují správný (reálný) tvar znaku daný referenční datovou sadou. Čísla v buňkách matice reprezentují podmíněnou pravděpodobnost, že znak s reálnou hodnotou t_i bude SW rozpoznán jako x_i . Součty hodnot v jednotlivých sloupcích matice jsou rovny jedné.

Finální matice $p(y|x)$ je opět maticí ve tvaru 37x37, přičemž řádky reprezentují SW rozpoznávaný znak na měřicím profilu X a sloupce reprezentují SW rozpoznávaný znak na měřicím profilu Y. Čísla v buňkách matice reprezentují podmíněnou pravděpodobnost, že znak, který byl na měřicím profilu X rozpoznán s hodnotou x_i , bude na měřicím profilu Y rozpoznán s hodnotou y_i . Po sestavení této matice je možné určovat podmíněnou pravděpodobnost, nakolik 2 rozpoznané tvary registrační značky na libovolných měřicích profilech patří ke stejnému vozidlu.

Matice $p(y|x)$ je dána předpisem [zdroj - článek] $p(y|x) = \sum_t p(y|t) \cdot p(t|x)$. Vzhledem k tomu, že na všech měřicích profilech je pro měření videozáznamů a následné rozpoznání registračních značek využívána stejná technologie, stejné HW nahrávací zařízení a stejný rozpoznávaný SW. Uvedený předpis lze tedy zjednodušit na $p(y|x) = \sum_t p(x|t) \cdot p(t|x)$. Matice $p(y|x)$ byla tedy vypočítána jako skalární součin dříve stanovených matic $p(x|t)$ a $p(t|x)$. Součty hodnot v jednotlivých řádcích matice jsou rovny jedné. Finální podoba matice $p(y|x)$, sestavená dle datové sady, je k vidění na následujícím obrázku 3.

Obrázek 4 Matice $p(y|x)$

Konkrétně lze využít demonstrovat na následujícím příkladu: Na měřicím profilu X byl naměřen záznam registrační značky RZ1 ve tvaru 6C97541. Na měřicím profilu Y byl naměřen záznam registrační značky RZ2 ve tvaru 8C95741. Skutečné tvary těchto registračních značek (T1, T2) jsou neznámé. S využitím matice $p(y|x)$ jsou porovnány příslušné dvojice znaků obou RZ a je jim přiřazena příslušná

pravděpodobnost. Výsledkem je řetězec hodnot pravděpodobností {0,00798; 0,97989; 0,98139; 0,97215; 0,99526; 0,98945; 0,97928}. Součet těchto pravděpodobností, příslušných jednotlivým znakům, udává celkovou pravděpodobnost $p(X \rightarrow Y)$, tedy, že se jedná o registrační značku stejného vozidla. V uvedeném příkladu je to hodnota 5,90541. Po znormování dle počtu znaků RZ, zde po vydělení 7, již výsledná hodnota leží v klasickém intervalu pravděpodobnosti $\langle 0;1 \rangle$, konkrétně dosahuje hodnoty 0,84.

Při reálném využití metody je prvním krokem výběr vhodné množiny potenciálně spárovatelných záznamů k řešenému záznamu registrační značky. Tato množina zahrnuje vhodně zvolené záznamy registračních značek z N relevantních měřících profilů v relevantním časovém intervalu. Konkrétní specifikací množiny potenciálně spárovatelných záznamů se věnuje kapitola 10.

Všechny potenciálně spárovatelné záznamy jsou pak s využitím matice $p(y|x)$ spárovány s řešeným záznamem a každé takto vzniklé dvojici záznamů je přiřazena pravděpodobnost $p(X \rightarrow Y)$. Pro každou relevantní relaci měřících profilů jsou pak jako správně spárované dvojice záznamů určeny ty dvojice, pro které platí, že $p(X \rightarrow Y) \geq M$, kde M je vhodně stanovená mezní hodnota. Určením mezní hodnoty M se zabývá následující kapitola 9.3.

9.3 Stanovení meze M a spolehlivosti jejího odhadu, její význam a využití

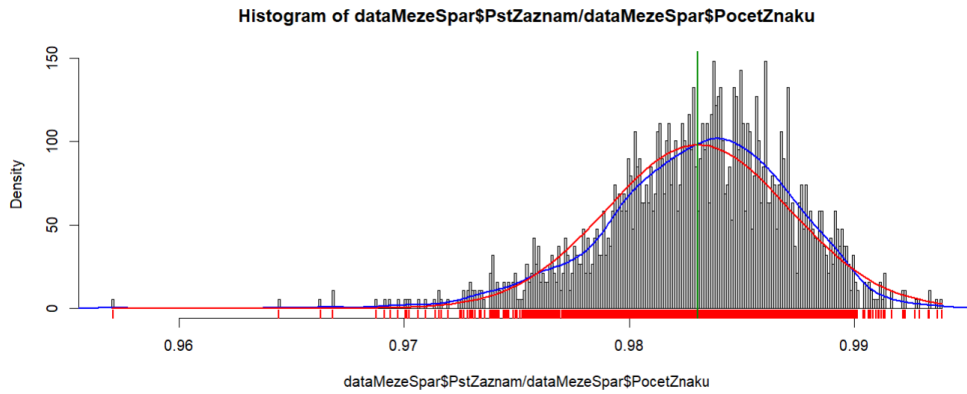
Mez M představuje důležitou hodnotu z intervalu $\langle 0;1 \rangle$, pro kterou platí, že pokud $p(X \rightarrow Y) \geq M \rightarrow$ spárovaná dvojice záznamů představuje záznamy registrační značky stejného vozidla. Taková dvojice záznamů skutečně patří k sobě a je správně spárovaná. Pokud naopak platí: $p(X \rightarrow Y) < M$, předmětná dvojice záznamů reálně představuje záznamy dvou registračních značek dvou různých vozidel a nemá být spárovaná. Pokud by byla spárovaná, šlo by o falešné spárování.

Odhad spolehlivosti meze M určuje procentuální podíl záznamů, které budou při aplikaci meze M spárovány (nebo naopak nespárován) chybně.

Pro stanovení meze M byly z datové sady vytvořeny následující 3 dílčí datové sady:

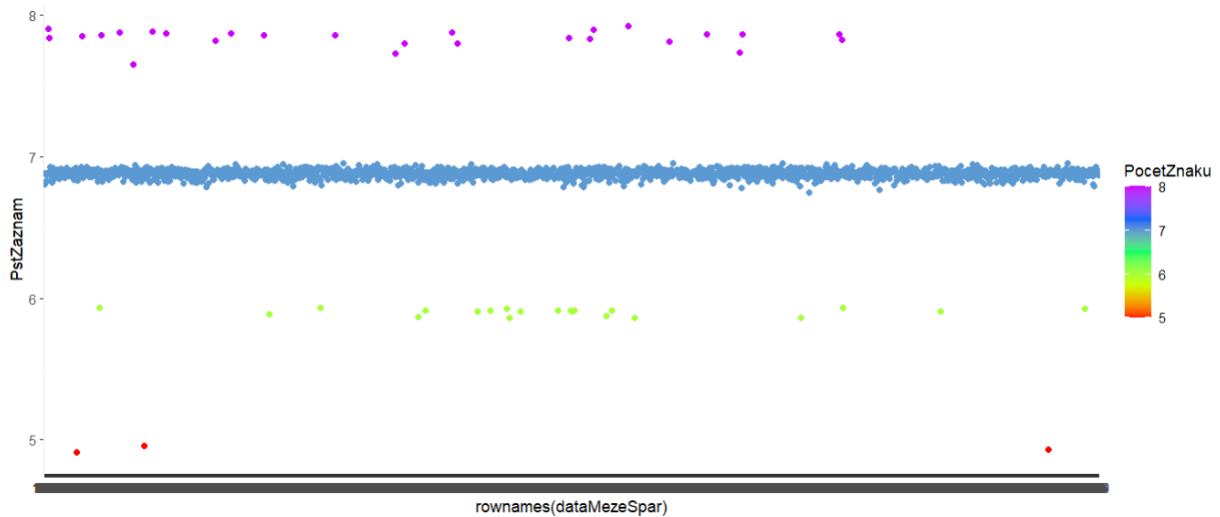
- Zelená datová sada – jedná se o reálně stejné a správně rozpoznané záznamy,
- Červená datová sada – jedná se o reálně různé a správně rozpoznané záznamy,
- Modrá datová sada – jedná se o reálně stejné a chybně rozpoznané záznamy.

Zelená datová sada představuje trénovací množinu záznamů, který reálně patří ke stejnému vozidlu a po spárování jsou správně spárovány. Byla vytvořena ze všech referenčních (správných) hodnot registračních značek ze zdrojové datové sady, tedy byl využit sloupec LP_correct, s tím, že byly odstraněny duplicity. Počet záznamů v této datové sadě byl 1890. Každý ze záznamů byl s využitím matice $p(y|x)$ spárován sám se sebou. Výsledné hodnoty pravděpodobností byly znormovány dle počtu znaků v registrační značce a zobrazeny formou histogramu a grafu hustoty pravděpodobnosti. Výsledek lze vidět na grafu 9 níže. Modrá křivka znázorňuje hustotu pravděpodobnosti, červená pak ilustračně normální rozdělení. Minimální hodnota normované pravděpodobnosti je 0,957.

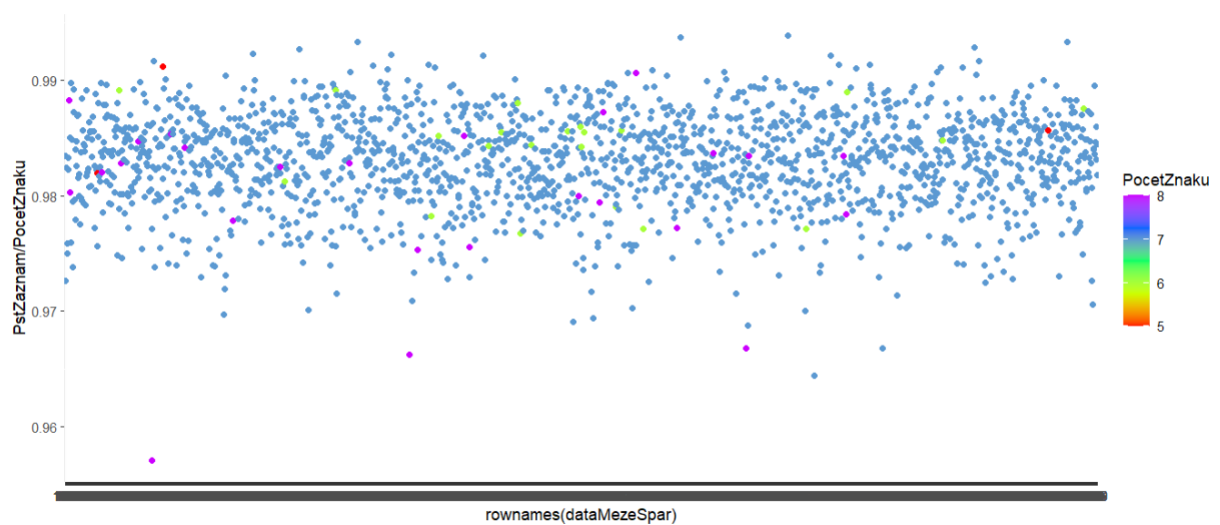


Graf 9 Hustota pravděpodobnosti zelené datové sady

Následující dva grafy demonstrují nutnost normování výsledků dle počtu znaků RZ:



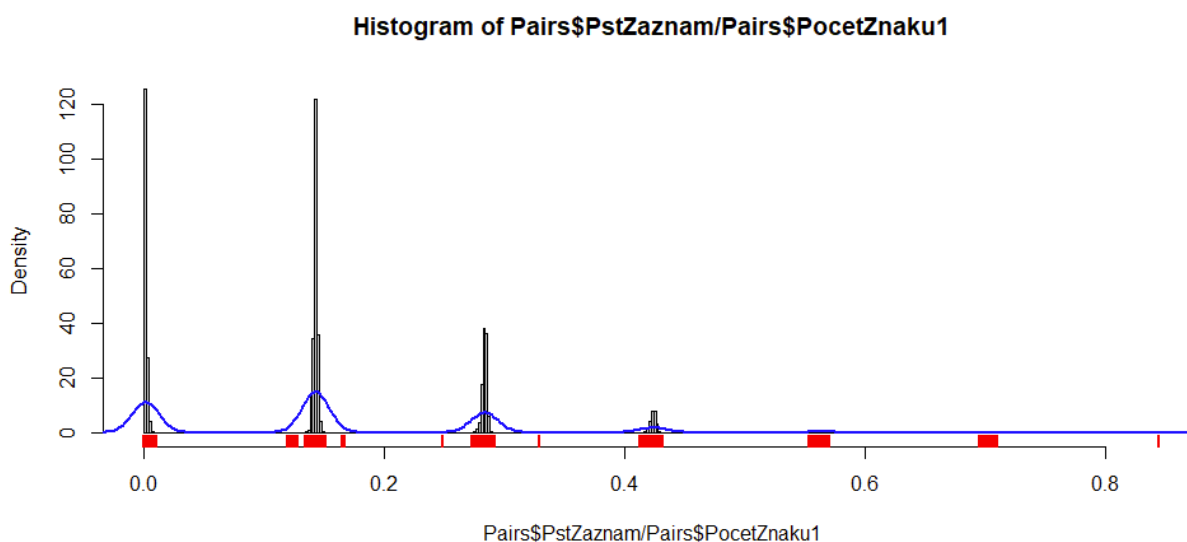
Graf 10 Pravděpodobnost shody RZ, bez normování



Graf 11 Pravděpodobnost shody RZ, normovaný graf

Červená datová sada představuje trénovací množinu záznamů, které reálně patří ke dvěma různým vozidlům a po spárování jsou falešně spárovány. Byla vytvořena ze všech referenčních (správných)

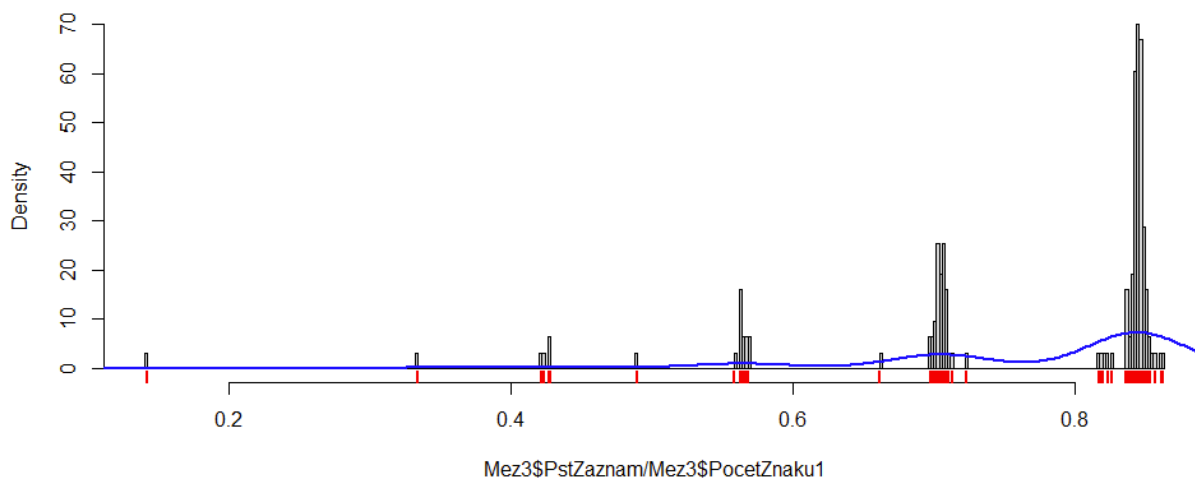
hodnot registračních značek ze zdrojové datové sady, tedy byl využit sloupec LP_correct, s odstraněním duplicit. Z těchto záznamů bylo náhodně vytvořeno 100 000 dvojic, s tím, že žádná dvojice neobsahovala dva shodné záznamy a oba záznamy ve dvojici měly stejný počet znaků. Počet záznamů v červené datové sadě byl tedy 100 000; vysoký počet záznamů byl zvolen z důvodu zisku co nejrelevantnějšího výsledku. Teoreticky by z 1890 záznamů bylo možné vytvořit až cca 3,5 mil. unikátních dvojic, nicméně další práce s takto vysokým počtem by byla výpočetně náročná. Každá dvojice záznamů byla s využitím matice $p(y|x)$ spárována a výsledné hodnoty pravděpodobností byly znormovány dle počtu znaků v registrační značce a zobrazeny formou histogramu a grafu hustoty pravděpodobnosti. Výsledek lze vidět na grafu 10 níže, kde modrá křivka znázorňuje hustotu pravděpodobnosti. Maximální hodnota normované pravděpodobnosti je 0,845.



Graf 12 Hustota pravděpodobnosti červené datové sady

Modrá datová sada představuje trénovací množinu záznamů, který reálně patří ke stejnému vozidlu, ale SW byl jeden ze záznamů rozpoznán s chybou, tedy textové řetězce obou záznamů nejsou shodné (ale mají stejnou délku). Takové záznamy po spárování budou správně spárovány a jejich spárování je cílem aplikované metody. Modrá datová sada byla výběrem těch řádků ze zdrojové datové sady, ve kterých se hodnota SW rozpoznané RZ nerovná referenční, tedy hodnota ve sloupci LP_R byla odlišná od hodnoty ve sloupci LP_correct. Jde o mírné zjednodušení, protože byla brána referenční hodnota namísto druhé reálně rozpoznané hodnoty registrační značky stejného vozidla. Protože však většinou platí, že alespoň na jednom z měřicích profilů je registrační značka rozpoznána správně, jedná se o zanedbatelné zanedbání. Počet dvojic záznamů v modré datové sadě byl 157. Každá dvojice záznamů byla s využitím matice $p(y|x)$ spárována a výsledné hodnoty pravděpodobností byly znormovány dle počtu znaků v registrační značce a zobrazeny formou histogramu a grafu hustoty pravděpodobnosti. Výsledek lze vidět na grafu 11 níže. Modrá křivka znázorňuje hustotu pravděpodobnosti. Minimální hodnota normované pravděpodobnosti je 0,142, maximální hodnota normované pravděpodobnosti je 0,863.

Histogram of Mez3\$PstZaznam/Mez3\$PocetZnaku1



Graf 13 Hustota pravděpodobnosti modré datové sady

Na základě získaných rozložení pravděpodobností v zelené, červené a modré datové sadě, je možné stanovit mezní hodnotu M . Mezní hodnotu je třeba stanovit tak, aby s co nejmenší chybou platilo, že když $p(X \rightarrow Y) \geq M$, pak jsou potenciálně spárovatelné záznamy správně spárované. Naopak, když $p(X \rightarrow Y) < M$, pak potenciálně spárovatelné záznamy nesmí být spárovány, protože by šlo o falešné spárování (patří reálně různým vozidlům). Z toho vyplývá, že by mělo platit M je menší než pravděpodobnosti stanovené pro zelenou a modrou datovou sadu. Naopak, M je větší než pravděpodobnosti stanovené pro červenou datovou sadu. Vzhledem k tomu, že pravděpodobnosti červené a modré sady nelze oddělit jednou hranicí, protože se překrývají, je třeba stanovit M tak, aby s jejím využitím bylo minimum potenciálně spárovatelných záznamů vyhodnoceno chybně a aby bylo dosaženo požadované spolehlivosti výsledných vyhodnocených LPR dat. Určováním konkrétních potřebných spolehlivostí vyhodnocených LPR dat vzhledem k jejich návaznému využití, se výzkum v tomto roce nezabýval (jedná se o činnost naplánovanou do následujícího roku). Z tohoto důvodu byla prozatím mez M stanovena tak, aby 99,9 % záznamů, které nemají být spárovány, spárovány nebyly. Byl zohledněn fakt, že červená datová sada je dostatečně robustní pro poskytnutí spolehlivých informací o rozložení hodnot pravděpodobností pro dvojice záznamů patřící různým vozidlům. **Konkrétně, mez M byla stanovena na hodnotu 0,57.**

10 Algoritmus

Na základě předchozího výzkumu byl sestaven algoritmus pro zvýšení kvality LPR dat, konkrétně pro spárování záznamů registračních značek, které rozpoznávacím SW nebyly rozpoznány jako shodné textové řetězce. Cílem algoritmu je snížení chybovosti výsledných sestavených cest a tras vozidel po zkoumané oblasti.

Ve vztahu ke kvalitativním skupinám surových záznamů (po SW rozpoznání), které byly definovány v tabulce 6 v kapitole 7.3, algoritmus řeší vybrané kvalitativní skupiny, a to následovně:

- *správně rozpoznané záznamy* – algoritmus řeší nespárované záznamy z této kategorie, nespáruje je a označí jako „správné“ – tyto záznamy vyhovují mezi SW psti ($LPMconfidence = 100\%$, $OCRconfidence \geq 97\%$),
- *opakovaná detekce RZ stejného vozidla* – algoritmus řeší seskupením všech záznamů stejné RZ na stejném měřicím profilu (pokud mají stejný počet znaků),
- *nepřesně rozpoznané RZ* – algoritmus řeší; pokud vozidlo projelo přes více měřicích profilů, bude spárováno, pokud ne, bude považováno za cílové/zdrojové, s možným upozorněním na chybu záznamu (dle hodnot SW stanovených parametrů spolehlivosti)

Algoritmus naopak neřeší následující kvalitativní skupiny surových záznamů:

- *opačný směr jízdy* – algoritmus předpokládá, že v surových datech je správně určen směr jízdy; informaci o směru jízdy poskytují rozpoznávací SW, její spolehlivost by bylo možné řešit v rámci návazného výzkumu,
- *neexistující RZ* – opakované záznamy jiného textu v rámci obrazu v průběhu celého měření (např. plakát apod.) – pokud jde o stacionární text v obraze, měl by být ošetřen už v průběhu přípravy surových dat z jednotlivých měřicích profilů. Pokud jde o záznamy nesouvisejících textů na projíždějících vozidlech – jedná se o předpokládanou chybovost algoritmu. Tyto texty budou identifikovány a případně spárovány jako další průjezd vozidla,
- *chybějící záznam* – algoritmus neřeší, chybějící záznamy jsou minimalizovány nastavením maximálního rozsahu detekce rozpoznávacího SW (za současného zvýšení počtu nepřesně rozpoznávaných RZ).

Algoritmus pracuje s následujícími vstupy:

- datová sada surových dat
- datová sada spárovaných dat
- matice relevantních návazností měřicích profilů (ve formátu matice sousednosti)
- matice T_{min} – stanovená dle spárovaných dat
- matice T_{max} – stanovená dle spárovaných dat s využitím limitních dob jízdy
- Mezní hodnoty dle SW parametrů spolehlivosti $LPMconfidence$, $OCRconfidence$
- Mezní hodnota pravděpodobnosti shody dvou záznamů
- Matice podmíněných pravděpodobností Matice $p(y|x)$

10.1 Kroky algoritmu a jejich popis:

1. Krok – Identifikace nespárovaných dat po proběhlém spárování
2. Krok – Sestavení množiny potenciálně spárovatelných záznamů
3. Krok – Určení pravděpodobností shody dvojic potenciálně spárovatelných záznamů
4. Krok – Posouzení vůči mezi pravděpodobností
5. Krok – Stanovení výsledků
6. Krok (nepovinný) - Manuální ověření nedopárovaných záznamů
7. Krok – Aktualizace surových dat
8. Krok – Opětovné spárování dat a vyhodnocení dat

Následuje bližší rozbor jednotlivých kroků algoritmu:

1. Krok – Identifikace nespárovaných dat

Spárování surových dat z jednotlivých měřicích profilů proběhlo dle postupu uvedeného v Metodice vyhodnocení směrových dopravních průzkumů (78). Spárované jsou záznamy registračních značek se zcela totožným řetězcem znaků.

Jsou identifikovány záznamy, které v předchozím kroku nebyly spárovány. S těmito záznamy bude dále blíže pracováno a algoritmus bude ověřovat, zda s využitím pravděpodobnosti shody nedojde k jejich spárování. Záznamy, které již byly v předchozím kroku spárovány, jsou automaticky považovány za správně spárované, protože se jedná o zcela shodné záznamy.

V tomto kroku je vhodné provést evidenci počtu spárovaných a nespárovaných záznamů, včetně rozlišení na jednotlivé měřicí profily. Po dokončení algoritmu bude možné získat informaci, jaká část záznamů byla na jednotlivých měřicích profilech algoritmem dopárována. Tím bude možné stanovit chybovost jednotlivých měřicích profilů a v případě kontinuálního měření bude možné na základě těchto znalostí provést lepší kalibraci měřicích technologií (upravit záběr, přiblížení a zaostření kamer, případně zvolit zcela jinou měřicí lokaci).

2. Krok – Sestavení množiny potenciálně spárovatelných záznamů

V tomto kroku je pro každý nespárovaný záznam vybrána množina potenciálně spárovatelných záznamů. Pokud existuje N nespárovaných záznamů (lze značit jako NZ), existuje i N množin potenciálně spárovatelných záznamů (lze značit jako M_PSZ , samotné potenciálně spárovatelné záznamy pak jako PSZ).

V nejširším pohledu by každá M_PSZ obsazovala stejný počet PSZ , který by se rovnal celkovému počtu záznamů v datové sadě surových dat, s výjimkou příslušného NZ , ke kterému M_PSZ patří. Toto by představovalo množství dvojic na prověření pravděpodobnosti shody, které odpovídá počtu NZ vynásobenému počtu záznamů (mínus 1) v datové sadě surových dat.

Z důvodu snížení výpočetní náročnosti jsou množiny potenciálně spárovatelných záznamů sestavovány s využitím následujících kritérií, která rovněž zajistí, aby v každé M_PSZ byly pouze relevantní PSZ. Tato kritéria jsou aplikována vždy všechna.

Územní kritérium – vymezuje relevantní měřicí profily dle Matice relevantních návazností měřících profilů, a to předcházející i následné profily vzhledem k měřicímu profilu řešeného nespárovaného záznamu; v těchto relevantních měřících profilech budou hledány potenciálně spárovatelné záznamy.

Časové kritérium – ke každému relevantnímu měřicímu profilu vymezuje relevantní časový úsek, ve kterém budou hledány potenciálně spárovatelné záznamy. Časové úseky jsou stanoveny s využitím matic T_{\min} a T_{\max} .

Kritérium meze SW spolehlivostních parametrů – jsou zhodnoceny parametry *LPMconfidence* a *OCRconfidence* řešeného nespárovaného záznamu vzhledem ke stanoveným mezním hodnotám. Pokud platí, že *LPMconfidence*, *OCRconfidence* řešeného nespárovaného záznamu nabývají hodnot větších nebo rovných mezním hodnotám, řešený záznam je prohlášen za správně rozpoznáný. V takovém případě jsou do množiny potenciálně spárovatelných záznamů zařazeny pouze ty záznamy, jejichž hodnoty SW spolehlivostních parametrů jsou nižší než mezní hodnoty. Jinými slovy, pokud je řešený nespárovaný záznam správně rozpoznáný, do množiny potenciálně spárovatelných záznamů nebudou zařazeny žádné správně rozpoznané záznamy (protože logicky by šlo o záznamy průjezdů různých vozidel).

Po sestavení dílčích matic potenciálně spárovatelných záznamů je tedy celkový počet potenciálně spárovatelných dvojic záznamů dán vztahem $\sum_i^N NZ_i \cdot PSZ_{ij}$, kde N je počet nespárovaných záznamů, NZ_i je označení jednoho konkrétního nespárovaného záznamu, PSZ_{ij} je označení potenciálně jednoho konkrétního potenciálně spárovatelného záznamu z množiny potenciálně spárovatelných záznamů patřící k příslušnému NZ_i .

3. Krok – Určení pravděpodobností shody dvojic potenciálně spárovatelných záznamů

Každé potenciálně spárovatelné dvojici záznamů je určena pravděpodobnost shody, dle metody představené v předchozí kapitole. Tento a následující krok je (za současného nastavení Matice $p(y|x)$) prováděn pouze pro ty potenciálně spárovatelné dvojice záznamů, kde mají obě RZ stejný počet znaků.

Výsledkem je normovaná pravděpodobnost, normovaná dle počtu znaků posuzovaných registračních značek.

4. Krok – Posouzení vůči mezi pravděpodobností

Pravděpodobnost shody každé potenciálně spárovatelné dvojice záznamů je vyhodnocena dle stanovené mezní hodnoty pravděpodobnosti shody. Jestliže je pravděpodobnost shody vyšší nebo rovna mezní hodnotě, je příslušná dvojice záznamů prohlášena za správně spárovanou. Jestliže je pravděpodobnost shody nižší než mezní hodnota, jsou záznamy z příslušné dvojice ponechány samostatně.

5. Krok – Stanovení výsledků

Ke každému původně nespárovanému záznamu jsou ponechány pouze ty potenciálně spárovatelné záznamy, které v předchozím kroku vyhověly posouzení dle mezní hodnoty. A to s tím, že pokud bylo ke stejnému nespárovanému záznamu nalezeno více potenciálně spárovatelných záznamů z jednoho měřicího profilu, bude z nich jako správný považován záznam s nejvyšší hodnotou pravděpodobnosti. Ostatní však budou ve výsledcích ponechány a budou rovněž spárovány, protože může jít o opakované záznamy stejného průjezdu vozidla na jednom měřicím stanovišti. Takto vytvořená množina M_{PSZ} se bude dále nazývat množina dopárování záznamů (M_{DZ}) k příslušnému původně nespárovanému záznamu (NZ).

V případě, že k příslušnému NZ nebyl nalezen žádný vyhovující potenciálně spárovatelný záznam a množina M_{DZ} je tedy prázdná, záznam bude potvrzen jako nespárovaný. Budou vyhodnoceny SW spolehlivostní parametry (*LPMconfidence* a *OCRconfidence*) daného záznamu. Pokud nabývají hodnot větších nebo rovných mezním hodnotám, řešený záznam je prohlášen za správně rozpoznáný a tvoří tedy zdrojovou nebo cílovou, případně vnitřní cestu ve zkoumané lokalitě. Pokud jsou SW spolehlivostní parametry nižší než mezní hodnoty, k záznamu je přidáno upozornění na možnost chybného SW rozpoznání.

V tomto kroku je vhodné doplnit a aktualizovat evidenci výsledků z kroku 2. Je vhodné provést evidenci počtu původně nespárovaných, ale nyní spárovaných záznamů, potvrzených nespárovaných záznamů, které vyhovují mezím SW spolehlivostních parametrů a lze je tedy prohlásit za správně rozpoznané, a nespárovaných záznamů, které mezím SW spolehlivostních parametrů nevyhovují. Tuto evidenci je vhodné provést souhrnně i s rozlišením na jednotlivé měřicí profily. Na základě zpracované evidence je možné stanovit chybovost jednotlivých měřicích profilů a v případě kontinuálního měření je možné na základě těchto znalostí provést lepší kalibraci měřicích technologií (upravit záběr, přiblížení a zaostření kamer, případně zvolit zcela jinou měřicí lokaci).

6. Krok (nepovinný) - Manuální ověření nedopárování záznamů

V tomto kroku je možné provést manuální ověření nespárovaných záznamů, které nevyhovují mezním hodnotám SW spolehlivostních parametrů. Cílem manuálního ověření je stanovit, zda se jedná o reálné záznamy průjezdů vozidel, s chybně rozpoznanou RZ, nebo zda se jedná o záznamy jiných textů, ať už na vozidlech nebo jinde v záběru. Pokud by šlo o záznamy jiných textů, pro dosažení co nejkvalitnějších výsledků je třeba je z datové sady zcela vyloučit. Nicméně, manuální ověřování je časově (a přeneseně i finančně) neefektivní a pokud nejsou k dispozici zdrojové videozáznamy, jeho provedení je nemožné. Proto se jedná o nepovinný krok.

Zároveň, v současné podobě algoritmus nebyl trénován na datech obsahující záznamy jiných textů. Proto neumožňuje automatizovanou identifikaci této kategorie záznamů. V příštím roce výzkumu je předpoklad rozšíření algoritmu o tuto kategorii.

7. Krok – Aktualizace surových dat

Na základě předchozích kroků je provedena aktualizace vstupní datové sady surových dat.

Záznamům, které byly v průběhu algoritmu spolu spárovány, bude přiřazen jednotný tvar textového řetězce registrační značky. Není nutné, aby se jednalo o reálně správný tvar registrační značky, podstatné je, aby byl jednotný. Bude zvolen ten tvar registrační značky, který má SW přiřazené nejvyšší hodnoty spolehlivostních parametrů a je tedy nejvíce pravděpodobné, že jde o reálně správný tvar.

Záznamy, které v průběhu algoritmu nebyly spárovány, ale byly označeny jako správně rozpoznané, budou v datové sadě surových dat ponechány beze změny.

Záznamy, které v průběhu algoritmu nebyly spárovány a zároveň byly označeny jako potenciálně chybně rozpoznané, budou v datové sadě surových dat ponechány s tímto upozorněním, pokud manuální kontrolou nedošlo k ověření, že se nejedná o průjezd vozidla. V takovém případě budou z datové sady surových dat zcela odstraněny.

8. Krok – Opětovné spárování dat a vyhodnocení dat

Aktualizovaná vstupní datová sada surových dat bude znovu spárována a dále vyhodnocena dle Metodiky vyhodnocení směrových dopravních průzkumů [zdroj – DP]. Při stanovování poměru tranzitní, vnitřní, a zdrojové/cílové dopravy bude zohledněna přítomnost nespárovaných záznamů s upozorněním na potenciálně chybně rozpoznané záznamy, které nemusí představovat reálný průjezd vozidla a při nezohlednění by mohlo docházet ke zkreslení výsledků.

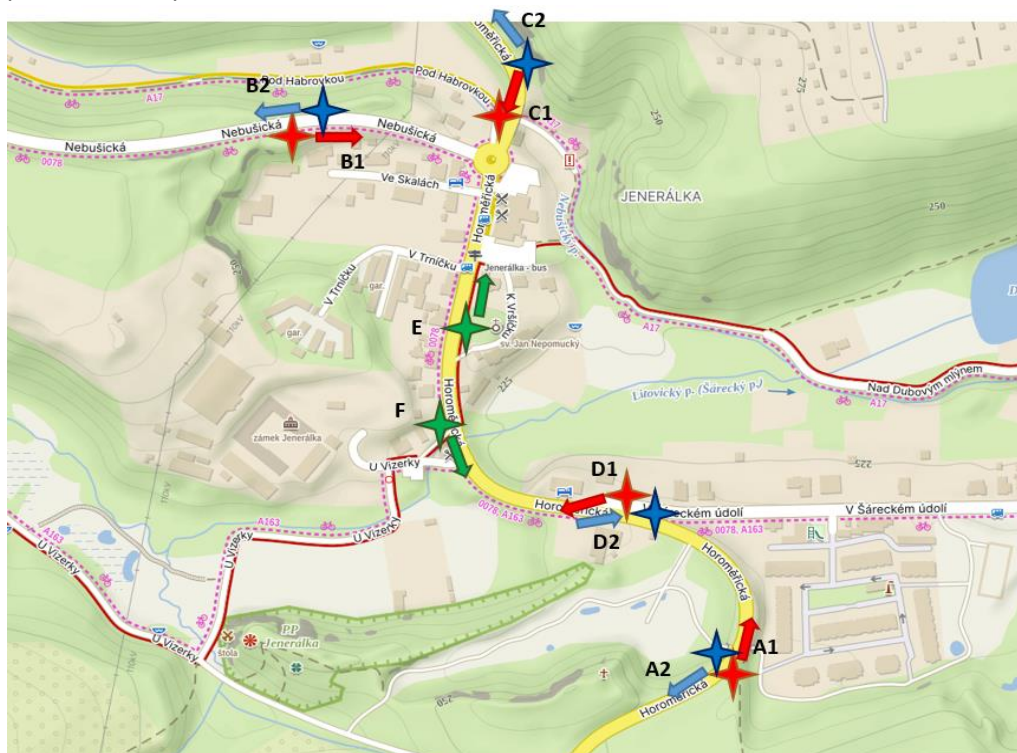
11 Ověření z dat z Jenerálky dle algoritmu

Ověření bylo zaměřeno na schopnost navrženého algoritmu zvýšit přesnost spárování záznamů a tím zvýšit kvalitu vyhodnocovaných dat. Ověřován byl tedy konkrétně 5. krok algoritmu – stanovení výsledků. Ověřování proběhlo s využitím prostředí R a s využitím manuálního ověření stanovených výsledků oproti referenčním videozáznamům.

11.1 Vstupy využité pro ověření

Vstupy, které byly pro ověření využity, vycházely z provedeného výzkumu a rovněž z provedeného směrového dopravního průzkumu v lokalitě pražské Jenerálky, který byl prováděn v čase 16:45-18:00 na 10 měřících profilech. Jedná se o následující:

- **datová sada surových dat**
 - o Jedná se o data ze směrového průzkumu, celkem 4306 záznamů, z 10 měřících profilů (vjezdové profily: A1, B1, C1, D1; výjezdové profily A2, B2, C2, D2; vnitřní profily E, F). Schéma rozmístění měřících profilů následuje na obrázku. Datová sada surových dat byla vytvořena rozpoznáním RZ pomocí SW Eyedea z těchto videozáznamů, bez následné manuální kontroly. Do datové sady byly použity záznamy ze všech měřících profilů v plném časovém rozsahu měření, s výjimkou měřícího profilu A2, kde vzhledem k méně vhodné zvolené lokaci měřícího profilu došlo k významnému snížení kvality posledních 15 minut záznamu. Tento časový interval byl nahrazen daty z měřícího profilu A1, z opačného směru.



Obrázek 5 Rozmístění měřících profilů, směrový průzkum Jenerálka

- **datová sada spárovaných dat**

- Jedná se o spárovaná data dle postupu v Metodice vyhodnocení směrových dopravních průzkumů (78).

- **matice relevantních návazností měřicích profilů**

- Jedná se o matici ve formátu matice sousednosti stanovující relevantní návaznosti mezi jednotlivými měřicími profily (v řádcích jsou uvedeny zdrojové a ve sloupcích cílové měřicí profily, číslice 1 pak označuje relevantní návaznost). Matice byla vytvořena na základě spárovaných dat a matice limitních jízdních dob, automaticky vytvořené dle postupu dle Metodiky vyhodnocení směrových dopravních průzkumů (78), s ověřením a úpravou dle reálné topologie silniční sítě ve zkoumané oblasti.

Tabulka 12 Matice relevantních návazností měřicích profilů, směrový průzkum Jenerálka

	A1	A2	B1	B2	C1	C2	D1	D2	E	F
A1	1	1	0	1	0	1	0	1	1	0
A2	1	1	0	0	0	0	0	0	0	0
B1	0	1	1	1	0	1	0	1	0	1
B2	0	0	1	1	0	0	0	0	0	0
C1	0	1	0	1	1	1	0	1	0	1
C2	0	0	0	0	1	1	0	0	0	0
D1	0	1	0	1	0	1	1	1	1	0
D2	0	0	0	0	0	0	1	1	0	0
E	0	0	0	1	0	1	0	0	1	0
F	0	1	0	0	0	0	0	1	0	1

- **matice T_{\min}**

- Jedná se o matici ve formátu matice sousednosti stanovující relevantní návaznosti mezi jednotlivými měřicími profily (v řádcích jsou uvedeny zdrojové a ve sloupcích cílové měřicí profily, číselná hodnota v polích pak odpovídá stanovenému minimálnímu časovému rozdílu mezi záznamy na návazných měřicích profilech pro zvážení jejich spárování). Matice byla vytvořena na základě spárovaných dat a matice limitních jízdních dob, automaticky vytvořené dle postupu dle Metodiky vyhodnocení směrových dopravních průzkumů (78), s ověřením a úpravou dle reálné topologie silniční sítě ve zkoumané oblasti.

Tabulka 13 Matice T_{\min} , směrový průzkum Jenerálka

	A1	A2	B1	B2	C1	C2	D1	D2	E	F
A1	0	0	0	40	0	33	0	15	25	0
A2	0	0	0	0	0	0	0	0	0	0
B1	0	30	0	0	0	10	0	25	0	15
B2	0	0	0	0	0	0	0	0	0	0
C1	0	30	0	8	0	0	0	25	0	15
C2	0	0	0	0	0	0	0	0	0	0
D1	0	25	0	35	0	28	0	0	20	0
D2	0	0	0	0	0	0	0	0	0	0
E	0	0	0	15	0	8	0	0	0	0

F	0	15	0	0	0	0	0	10	0	0
----------	---	----	---	---	---	---	---	----	---	---

- **matice T_{max}**

- Jedná se o matici ve formátu matice sousednosti stanovující relevantní návaznosti mezi jednotlivými měřicími profily (v řádcích jsou uvedeny zdrojové a ve sloupcích cílové měřicí profily, číselná hodnota v polích pak odpovídá stanovenému maximálnímu časovému rozdílu mezi záznamy na návazných měřicích profilech pro zvážení jejich spárování). Matice byla vytvořena na základě spárovaných dat a matice limitních jízdních dob, automaticky vytvořené dle postupu dle Metodiky vyhodnocení směrových dopravních průzkumů (78), s ověřením a úpravou dle reálné topologie silniční sítě ve zkoumané oblasti.

Tabulka 14 Matice T_{max} , směrový průzkum Jenerálka

	A1	A2	B1	B2	C1	C2	D1	D2	E	F
A1	180	180	0	225	0	183	0	99	120	0
A2	180	180	0	0	0	0	0	0	0	0
B1	0	210	180	180	0	72	0	189	0	114
B2	0	0	180	180	0	0	0	0	0	0
C1	0	222	0	57	180	180	0	177	0	111
C2	0	0	0	0	180	180	0	0	0	0
D1	0	138	0	228	0	209	180	180	129	0
D2	0	0	0	0	0	0	180	180	0	0
E	0	0	0	99	0	80	0	0	180	0
F	0	111	0	0	0	0	0	75	0	180

- **Mezní hodnoty dle SW parametrů spolehlivosti $LPMconfidence$, $OCRconfidence$**

- Na základě provedeného výzkumu (viz kapitola 8) byly mezní hodnoty stanoveny na $LPMconfidence = 100\%$ a zároveň $OCRconfidence \geq 97\%$.

- **Mezní hodnota pravděpodobnosti shody dvou záznamů**

- Na základě provedeného výzkumu (viz kapitola 9) byla mezní hodnota stanovena na hodnotu $M = 0,57$.

- **Matice podmíněných pravděpodobností Matice $p(y|x)$**

- Byla využita Matice $p(y|x)$ stanovená na základě předchozího výzkumu (viz kapitola 9.2).

11.2 Průběh ověření

V následujícím textu je popsán průběh ověření definovaného algoritmu. Nejdříve je popsán postup vyhodnocování dat dle algoritmu, konkrétně dle kroků 1 až 5. Následně je popsáno manuální ověření stanovených výsledků oproti zdrojovým videozáznamům.

Popis postupu dle algoritmu pro zvýšení kvality LPR dat

V prvním kroku byly identifikovány záznamy, které zůstaly po spárování shodných registračních značek nespárovány. Z celkového počtu 4306 záznamů se jednalo o 464 nespárovaných záznamů.

V druhém kroku byla ke každému nespárovanému záznamu sestavena množina potenciálně spárovatelných záznamů dle územního a časového kritéria (matice relevantních návazností měřicích profilů, matice T_{\min} a matice T_{\max} a se zpřesněním za využití mezních hodnot SW parametrů spolehlivosti.

Pro snadnější SW zpracování a manipulaci s daty byly všechny množiny potenciálně spárovatelných záznamů sloučeny do jedné tabulky. Tato tabulka obsahovala 50 291 řádků, kde každý řádek představoval jednu potenciálně spárovatelnou dvojici záznamů.

Ve třetím kroku byla potenciálně spárovatelným dvojicím záznamů s využitím Matice $p(y|x)$ přiřazována pravděpodobnost správného spárování, normovaná dle počtu znaků RZ. Vzhledem k tomu, že Matice $p(y|x)$ byla sestavena pouze se zohledněním dvojic registračních značek se stejným počtem znaků, byly nyní řešeny pouze potenciálně spárovatelné dvojice záznamů, u kterých se počet znaků obou registračních značek rovnal. Z celkového počtu 50 291 potenciálně spárovatelných dvojic záznamů bylo 33 305 dvojic se shodným počtem znaků obou RZ. 16 986 dvojic bylo s rozdílným počtem znaků RZ, tyto byly tedy z vyhodnocení dle Matice $p(y|x)$ vynechány.

Ve čtvrtém kroku proběhlo posouzení normované pravděpodobnosti správného spárování posuzovaných 33 305 dvojic potenciálně spárovatelných záznamů vůči mezi pravděpodobnosti shody, $M = 0,57$. Počet dvojic potenciálně spárovatelných záznamů, jejichž vypočítaná normovaná pravděpodobnost vyhovovala mezi, bylo nalezeno celkem 486. Tyto dvojice byly prohlášeny za správně spárované.

V pátém kroku byly stanoveny výsledky provedení dopárování záznamů, a to následovně:

Počet původně nespárovaných záznamů registračních značek, ke kterým byl provedeným postupem nalezen alespoň jeden záznam vyhovující mezi pravděpodobnosti shody, bylo 245. Těchto 245 původně nespárovaných záznamů bylo tedy nyní určeno jako spárované.

Počet původně nespárovaných záznamů registračních značek, které zůstaly nespárovány, byl 219. Tyto záznamy byly prověřeny vzhledem k definovaným mezím SW parametrů spolehlivosti. Celkem 73 záznamů vyhovovalo mezím SW parametrů spolehlivosti, těmto záznamům byl tedy přidělen příznak, že jsou správně rozpoznány. Celkem 146 nespárovaných záznamů nevyhovělo mezím SW parametrů spolehlivosti, byly tedy označeny jako potenciálně chybně rozpoznány registrační značky, které mohou představovat např. neexistující texty v obraze nebo jiné texty na vozidlech.

Další kroky algoritmu nebyly prováděny, protože nejsou nezbytné k ověření zvýšení přesnosti spárování záznamů algoritmem.

Popis ověření výsledků

Stanovené výsledky byly ověřeny dle referenčních videozáznamů a byly porovnány s původní datovou sadou spárovaných dat, aby mohla být stanovena míra zvýšení přesnosti spárování dat při využití navrženého algoritmu.

Na základě kontroly dle referenčních videozáznamů byl každému původně nespárovanému záznamu přiřazen referenční příznak, který definoval reálný stav záznamu vozidla, a to jednou z následujících možností:

- *nespárovaný záznam* (správné i chybné záznamy registrační značky vozidla),
- *neexistující vozidlo* (jiný text),
- *spárovaný záznam* – vozidlo mělo být spárováno se záznamem na jiném stanovišti (správné i chybné záznamy registrační značky vozidla),
- *spárovaný, opakovaný záznam* – vozidlo mělo být spárováno se záznamem na stejném stanovišti (opakovaný chybný záznam registrační značky vozidla).

Následně bylo ověřeno, zda po původním spárování dat byl záznam do výše uvedených kategorií zařazen správně nebo chybně – s využitím parametrů „ano“ (tzn. správně zařazen) a „ne“ (tzn. chybně zařazen). Stejně tak byl ověřen výsledek po provedení algoritmu.

11.3 Výsledky ověření

Výsledky ověření přehledně zobrazuje následující tabulka:

Tabulka 15 Výsledky ověření algoritmu

		Kategorie záznamu registrační značky			
		Nespárovaný	Neexistující	Spárovaný	Opakovaný
Reálný stav	Celkový počet	62	79	248	75
	Celkový počet	464	0	0	0
Bez algoritmu	Správně určeno	62	0	0	0
	Chybně určeno	402	0	0	0
	Procento správného určení vzhledem k reálnému stavu	100 %	0 %	0 %	0 %
	Celkový poměr správně určených k chybně určeným záznamům	62:402 (13 % : 87 %)			
	Celkový počet	219	0	210	35
S algoritmem	Správně určeno	62	0	205	35
	Chybně určeno	157	0	5	0
	Procento správného určení vzhledem k reálnému stavu	100 %	0 %	83 %	47 %
	Celkový poměr správně určených	302:162 (65 % : 35 %)			

	k chybně určeným záznamům				
Procentuální zpřesnění při využití algoritmu		0 %	0 %	+ 83 %	+ 47 %
Změna celkového poměru správně určených k chybně určeným záznamům	změna o 52 % ve prospěch správně určených záznamů s využitím algoritmu				

Z výsledků je vidět, že navržený algoritmus jednoznačně přispěl ke zvýšení přesnosti spárovaných dat, a to v celkovém pohledu o 52 %. Při rozboru podrobnějších kategorií záznamů lze konstatovat následující:

- *nespárované záznamy* – záznamy, které dle reálného stavu neměly být spárovány, byly z logiky věci správně určeny i bez využití algoritmu. Algoritmus pak nezpůsobil žádné falešné spárování záznamů z této kategorie; všechny záznamy zůstaly správně nespárovány. Naopak se s využitím algoritmu výrazně snížil podíl falešně nespárovaných záznamů.
- *neexistující vozidla* – tuto kategorii algoritmus neřešil, neboť nebyl trénován na datové sadě záznamů neodpovídajících registračním značkám. Nicméně bez využití algoritmu tato kategorie rovněž není řešena, nedochází zde tedy ke zhoršení přesnosti výsledků.
- *spárované záznamy* – záznamy registrační značky, které měly být spárovány se záznamem na jiném měřicím profilu, ale na základě prosté shody znaků RZ spárovány nebyly, byly hlavní cílovou kategorií navrženého algoritmu. V této kategorii algoritmus zaznamenal zvýšení přesnosti o 83 %.
- *opakované záznamy* – ve spárování opakovaných záznamů průjezdu vozidla na stejném měřicím stanovišti algoritmus rovněž přinesl zvýšení přesnosti, konkrétně o 47 %.

Přes prokazatelné a výše popsané přínosy navrženého algoritmu bylo na základě ověření zjištěno celkem 162 záznamů, který algoritmus zařadil do chybných kategorií. Konkrétně se jednalo o 157 záznamů, které algoritmus chybně určil jako nespárované, a 5 % záznamů naopak algoritmus chybně určil jako spárované.

Chybovost algoritmu byla způsobena zejména limitacemi trénovací datové sady, konkrétně:

- nezahrnutí záznamů neexistujících vozidel,
- nezahrnutí srovnávání záznamů o různé délce textu registrační značky,
- nízkým počtem trénovacích dat v množině reálně stejných, ale chybně rozpoznávaných záznamů (podrobněji viz kapitola 9.3), v důsledku čehož byla mezní hodnota pravděpodobnosti shody dvou záznamů nastavena výrazně ve prospěch robustní datové sady reálně různých záznamů, tedy tak, aby především nedocházelo k falešně správnému spárování,

Na chybovost algoritmu měly dále vliv rovněž chyby ve vstupní datové sadě, která po provedení dopravního průzkumu byla vytvořena pouze SW rozpoznáním naměřených videozáznamů, s minimální manuální kontrolou. Jednalo se o následující chyby:

- nerozpoznání RZ na některém z měřicích profilů z důvodu snížené kvality videozáznamu,
- nerozpoznání RZ motocyklů na měřicích profilech, které snímaly vozidla zepředu – C1, D2 (motocykly mají umístěnou RZ pouze vzadu),
- rozpoznání různých RZ u vozidel s přívěsem (RZ vozidla a RZ přívěsu),
- špinavé nebo poškozené RZ se zhoršenou čitelností.

V jednom případě došlo k selhání algoritmu z důvodu chyby vnesené do datech při předzpracování vstupní datové sady, konkrétně uložením textového řetězce RZ v číselném formátu, který způsobil ztrátu prvního znaku „0“ z řetězce znaků RZ.

Ověření zároveň prokázalo, že v případě snížené kvality videozáznamu na některém měřicím profilu může být vhodné využít data z měřicího profilu v opačném směru, pokud se tento profil nachází v těsné blízkosti a zaznamenával data v obou směrech jízdy.

12 Budoucí rozvoj

Navržený algoritmus pro zvýšení kvality LPR dat, konkrétně pro zvýšení přesnosti spárování záznamů, prokázal pozitivní přínosy. Nicméně do budoucího roku je předpoklad dalšího rozvoje a zrobustnění algoritmu.

Mezi možné oblasti dalšího rozvoje, které byly identifikovány v průběhu proběhlého výzkumu, patří:

- výzkum spolehlivosti parametru směru jízdy vozidla udávaného rozpoznávacím SW a identifikace a opravy chyb v určení směru jízdy vozidla,
- zpřesnění vyhodnocení maximálního a minimálního časového rozdílu záznamů vozidel v relevantních relacích měřicích profilů (T_{max} , T_{min}) pro zvážení spárování těchto záznamů,
- zaměření se na řešení opakovaných záznamů stejného vozidla na jednom měřicím profilu a zejména na záznamy jiných textů než RZ, např. zaměření se na možnost identifikace pomocí časových rozdílů záznamů na jednom měřicím profilu nebo na SW udávané spolehlivostní parametry
- rozšíření trénovací sady, zpřesnění matice $p(y|x)$ a podrobnější výzkum meze pravděpodobnosti shody dvou záznamů.

V druhém roce řešení projektu je předpoklad výzkumu výše uvedených bodů, hlavní práce budou však směřovány především k analýze vlivu různé míry chybovosti vstupních surových LPR dat na spolehlivost výstupů. Budou analyzovány praktické způsoby využití dopravních dat, především LPR dat v telematických systémech řízení dopravy v městském prostředí a na základě zjištěných požadavků na spolehlivost vstupních LPR dat bude rozvíjen a zpřesňován navržený algoritmus. Konkrétně jsou na následující rok řešení projektu předpokládány tyto činnosti:

- 1) Detailní analýza realizovaných způsobů využití předzpracovaných dopravně inženýrských dat a z nich získaných informací v telematických systémech pro účely řízení silniční dopravy, analýza požadavků jednotlivých aplikací na spolehlivost využívaných vstupních (předzpracovaných) dat a informací, se zaměřením na LPR data a na městské prostředí.
- 2) Prioritizace vybraných aplikací z oblasti řízení silniční dopravy vzhledem k požadavkům na spolehlivost výstupních dopravně inženýrských dat a informací z vybraných aplikací, se zaměřením na LPR data.
- 3) Stanovení závislosti mezi přesností surových naměřených LPR dat, spolehlivostí předzpracovaných dat (která tvoří vstup pro aplikace řízení silniční dopravy) a spolehlivostí výstupních dopravně inženýrských dat a informací.
- 4) Rozšíření algoritmu (viz bod 5) o specifikaci požadované spolehlivosti LPR dat za účelem jejich využití ve vybraných aplikacích řízení silniční dopravy (viz body 8 a 9).
- 5) Rozšíření modelové implementace a ověření algoritmu (viz bod 6) o specifikaci požadované přesnosti LPR dat za účelem jejich využití ve vybraných aplikacích řízení silniční dopravy v městském prostředí.
- 6) Vyvození závěrů z výsledků celého dvouletého výzkumu přenositelných do širšího kontextu využití různých datových zdrojů pro různé aplikace řízení silniční dopravy v městském prostředí, objektivní hodnocení přínosů pro zefektivnění řízení silniční dopravy v městském prostředí.

13 Závěr

Cílem studie bylo především představit návrh algoritmu pro zvýšení spolehlivosti předzpracování LPR pro jejich další využití v systémech řízení a managementu dopravy. Navržený algoritmus byl založen na provedené podrobné analýze současných metod sběru a zpracování LPR dat, na analýze zdrojů chyb a možností jejich eliminace. Analyzovány byly současné odborných výzkumných článků, kde byly zkoumány zavedené postupy i nové a inovativní výzkumné směry a trendy. Byla provedena rovněž detailní analýza současného stavu technologií dostupných na trhu pro sběr LPR dat – tedy kamer a SW aplikací pro rozpoznání registračních značek.

Na základě provedené analýzy byla vybrána hlavní metoda pro zvýšení přesnosti LPR, která využívá pravděpodobnostní model shody rozpoznávaných textů registračních značek na základě pravděpodobnosti záměny jednotlivých znaků dvou porovnávaných RZ, které vyplývá z technické funkce využitého rozpoznávacího SW. Tato metoda byla rozšířena na rozsáhlejší počet měřicích profilů. V algoritmu byla obohacena využitím parametrů pravděpodobnosti správnosti detekce RZ a rozpoznání textu RZ, který udávají některé rozpoznávací SW.

Bylo provedeno ověření algoritmu a stanovení procentuálního zpřesnění spárovaných dat. Ověření bylo provedeno na datové sadě získané provedeným směrovým dopravním průzkumem, tato datová sada obsahovala celkem 4306 záznamů z průběhu 1,25 h dlouhého měření, z 10 měřicích profilů (4 vjezdové profily na hranici zkoumané oblasti, 4 výjezdové profily a 2 vnitřní měřicí profily). Výsledky ověření prokázaly, že navržený algoritmus jednoznačně přispěl ke zvýšení přesnosti spárovaných dat. Ověření probíhalo na originálně nespárovaných datech (tzn. záznamech RZ, ke kterým prostou shodou textového řetězce nebyl nalezen v datové sadě žádný pár), protože algoritmus byl cílen zejména na tzv. dopárování chybně rozpoznávaných záznamů. Výsledky ověření ukázaly, že v kategorii nespárovaných záznamů navržený algoritmus přinesl zpřesnění, a to v celkovém pohledu o 52 %.

14 Seznam obrázků

Obrázek 1 Schéma rozložení měřicích profilů směrového průzkumu Strahov	24
Obrázek 2 Nastavení SW Eyedea.....	25
Obrázek 3 Vyhodnocení spolehlivostních parametrů – ilustrativní příklad	41
Obrázek 4 Matice $p(y x)$	47
Obrázek 5 Rozmístění měřicích profilů, směrový průzkum Jenerálka	57

Seznam tabulek

Tabulka 1 Struktura surových dat; zdroj: (DP), upraveno	7
Tabulka 2 Výhody a nevýhody obou popsanych řešení	13
Tabulka 3 Dostupné časové intervaly videozáznamů, tzn. referenční datové sady (doba měření byla 8:00-9:15)	24
Tabulka 4 Výstup ze SW Eyedea	26
Tabulka 5 Parametry doplňované na základě kontroly	27
Tabulka 6 Shrnutí výsledků kontroly datových sad, ve vztahu k parametru RZ.....	29
Tabulka 7 Kvalitativní přehled dat ve vstupní datové sadě pro vyhodnocení dle metodiky pro vyhodnocení směrových průzkumů	31
Tabulka 8 Chybovost původní zdrojové datové sady po opravě dle datové sady cest.....	33
Tabulka 9 Chybovost výsledné datové sady opravené dle datové sady cest	Chyba! Záložka není definována.
Tabulka 10 Popis počtu záznamů	35
Tabulka 11 Vyhodnocení spolehlivostních parametrů – ilustrativní příklad	42
Tabulka 12 Vyhodnocení spolehlivostních parametrů – ilustrativní příklad, závěr	43
Tabulka 13 Popis počtu záznamů	Chyba! Záložka není definována.
Tabulka 14 Matice relevantních návazností měřicích profilů, směrový průzkum Jenerálka	58
Tabulka 15 Matice T_{min} , směrový průzkum Jenerálka	58
Tabulka 16 Matice T_{max} , směrový průzkum Jenerálka	59
Tabulka 17 Výsledky ověření algoritmu	61

Seznam grafů

Graf 1 Histogram celkových četností záznamů dle parametru LPMconfidence	37
Graf 2 Histogram celkových četností záznamů dle parametru OCRconfidence.....	37
Graf 3 Histogram četností správných záznamů dle parametru LPMconfidence.....	38
Graf 4 Histogram četností chybných záznamů dle parametru LPMconfidence.....	38
Graf 5 Histogram četností správných záznamů dle parametru OCRconfidence	39
Graf 6 Histogram četností chybných záznamů dle parametru OCRconfidence	39
Graf 7 Grafické zobrazení správně/chybně rozpoznávaných záznamů ve vztahu k oběma parametrům LPMconfidence a OCRconfidence	40
Graf 8 Vývoj pravděpodobnosti chyby odhadu v závislosti na velikosti trénovací množiny.....	43
Graf 9 Hustota pravděpodobnosti zelené datové sady.....	49
Graf 10 Pravděpodobnost shody RZ, bez normování.....	49
Graf 11 Pravděpodobnost shody RZ, normovaný graf	49
Graf 12 Hustota pravděpodobnosti červené datové sady	50
Graf 13 Hustota pravděpodobnosti modré datové sady	51

Reference

1. Robinson, A., Niekerk, A. Uses of ANPR data in traffic management and transport modelling. *Proceedings of the 33rd Southern African Transport Conference (SATC 2014)*. <https://api.semanticscholar.org/CorpusID:106435641>. , 2014.
2. Zhan, X., Li, R., Ukkusuri, S. V. Lane-based real-time queue length estimation using license plate recognition data. *Transportation Research Part C: Emerging Technologies*, (57), 85-102. doi: 10.1016/j.trc.2015.06.001., 2015.
3. Li, R., Wang, S., Jiao, P., Lin, S. Traffic control optimization strategy based on license plate recognition data. *Journal of Traffic and Transportation Engineering*. 10(1), 45-57, 2023.
4. Bommès, M., Fazekas, A., Volkenhoff, T., Oeser, M. Video Based Intelligent Transportation Systems – State of the Art and Future Development. *Transportation Research Procedia*. (14), 4495-4504. doi: 10.1016/j.trpro.2016.05.372., 2016.
5. Navrátilová, K., Hajčiarová, E. Detailní analýza dostupných prostředků regulace parkování v uličním prostoru. *Výzkumná zpráva*., 2022.
6. Zmud, J., Wagner, J., Moran, M., George, J. License Plate Reader Technology: Transportation Uses and Privacy Risks. *American Association of State Highway and Transportation Officials, Texas A&M University School of Law Legal Studies Research Paper*. <https://ssrn.com/abstract=3443832>, 2016.
7. Langr, M., Růžička, J., Hajčiarová, E., Rek, R., Bečka, O. Studie strategického rozvoje dopravy ve městě Dobřichovice. . *Město Dobřichovice*. [Online] 2022. [Citace: 9. 20 2024.] <https://www.dobrichovice.cz/mesto/studie-strategickeho-rozvoje-dopravy-ve-meste-dobrichovice/>.
8. General dopravy ve městě Dačice: Dopravní průzkumy, analytická část. *Město Dačice*. [Online] 2017. [Citace: 20. 09 2024.] https://www.dacice.cz/e_download.php?file=data/editor/707cs_23.pdf&original=General%20dopravy_Da%C4%8Dice_Anal%C3%BDza_30_08_2017.pdf.
9. HBH Projekt, spol. s r.o. Územní studie obchvatu Kroměříže. *Město Kroměříž*. . [Online] [Citace: 20. 09 2024.] <https://www.mesto-kromeriz.cz/urad/dokumenty-a-informace/uzemni-studie/kromeriz/>..
10. Yao, W., Zhang, M., Jin, S., Ma, D. Understanding vehicles commuting pattern based on license plate recognition data. . *Transportation Research Part C: Emerging Technologies*. (128). doi: 10.1016/j.trc.2021.103142., 2021.
11. Yu, H., Yang, S., Wu, Z., Ma, X. Vehicle trajectory reconstruction from automatic license plate reader data. *International Journal of Distributed Sensor Networks*. 14(2). doi: 10.1177/1550147718755637., 2018.
12. Tang, J., Wan, L., Schooling, J., Zhao, P., Chen J., Wei S. Automatic number plate recognition (ANPR) in smart cities: A systematic review on technological advancements and application cases. . *Cities*. (129). doi: 10.1016/j.cities.2022.103833., 2022.
13. Purkrábková, Z., Hrubeš, P. Basic characteristics of floating car data from the perspective of traffic loss during the COVID-19 pandemic. *European Transport - Trasporti Europei*. (93). doi: 10.48295/ET.2023.93.5, 2023.
14. Hadachi, A., Pourmoradnasseri, M. A Survey on the Advancement of Travel Time Estimation Using Mobile Phone Network Data. *IEEE Transactions on Intelligent Transportation Systems*. 23(8), 11779-11788. doi: 10.1109/TITS.2021.3107212., 2022.
15. Ochôa, I., Calbusch, L., Viecelli, K., de Paz, J., Leithardt V., Zeferino, C. Privacy in the Internet of Things: A Study to Protect User's Data in LPR Systems Using Blockchain. *17th International Conference on Privacy, Security and Trust (PST)*. 1-5. doi: 10.1109/PST47121.2019.8949076., 2019.
16. Van de Vyvere, B., Colpaert, P. Using ANPR data to create an anonymized linked open dataset on urban bustle. . *European Transport Research Review*. 14(17). doi: 10.1186/s12544-022-00538-1., 2022.
17. REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - of 27 April 2016 (General Data Protection Regulation). *EUR-Lex: Acces to European Union law*. . [Online] 2016. [Citace: 18. 08 2024.] <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679..>

18. Chen, H., Yang, C., Xu, X. Clustering Vehicle Temporal and Spatial Travel Behavior Using License Plate Recognition Data. *Journal of Advanced Transportation*. . doi: 10.1155/2017/1738085, 2017.
19. Mo, B., Li, R., Zhan, X. Speed profile estimation using license plate recognition data. . *Transportation Research Part C: Emerging Technologies*. (82), 358-378. doi: 10.1016/j.trc.2017.07.006., 2017.
20. Mcleod, F., Cherrett, T. J., Box, S., Waterson, B. J. Pritchard, J. A. Using automatic number plate recognition data to investigate the regularity of vehicle arrivals. *European Journal of Transport and Infrastructure Research*. 17. 10.18757/ejtir.2017.17.1.3181., 2017.
21. Hrubeš, P., Langr, M., Purkrábková, Z. Review of Data Governance Approaches in the Field of Transportation Domain. *2024 Smart City Symposium Prague (SCSP)*. 1-7. doi: 10.1109/SCSP61506.2024.10552682, 2024.
22. Zhang, H., Chen, P., Zheng, J., Zhu, J., Yu, G., Wang, Y., Liu, H. X. Missing data detection and imputation for urban ANPR system using an iterative tensor decomposition approach. *Transportation Research Part C: Emerging Technologies*. (107), doi: <https://doi.org/10.1016/j.trc.2019.08.013>, 2019.
23. Bystrianský, A., Neuwirth, P., Striegler, R., Vyskočilová, L. *Metodika pro provádění a vyhodnocení směrových průzkumů automobilové dopravy*. místo neznámé : Centrum dopravního výzkumu, 2023. ISBN 978-80-88074-71-7.
24. Lubna, Mufti, N., Shah, S.A.A. Automatic Number Plate Recognition:A Detailed Survey of Relevant Algorithms. *Sensors*. (21). doi: 10.3390/s21093028, 2021.
25. S. Mala, V. H. R and K. Chanda. Yolo model-based license plate extraction and toll generation for smart parking systems. *2024 2nd International Conference on Networking, Embedded and Wireless Systems (ICNEWS)*. 1-7, doi: 10.1109/ICNEWS60873.2024.10730931, 2024.
26. Panahi, R. a Gholampour, I. Accurate Detection and Recognition of Dirty Vehicle Plate Numbers for High-Speed Applications. *IEEE Xplore*. 2017.
27. ALPR Results in the Real World. [Online] 2024. <https://platerecognizer.com/alpr-results/>.
28. ANPR Accuracy Unveiled: How Reliable Is Automatic Number Plate Recognition? [Online] 2024. <https://carmencloud.com/anpr-accuracy-unveiled-how-reliable-is-automatic-number-plate-recognition/>.
29. VIDAR ANPR/ALPR Camera. [Online] 2024. <https://adaptiverecognition.com/products/vidar-anpr-camera/>.
30. *Autogard*. [Online] 2024. <https://www.autogard.cz/produkty/parkovaci-systemy/>.
31. *ADIGlobal.com*. [Online] 2024. <https://adiglobal.cz/cz/produkty141/page/rozpoznavani-spz>.
32. *EFG*. [Online] 2024. <https://www.efg.cz/kamerove-systemy>.
33. *GreenCenter*. [Online] 2024. <https://green.cz/kategorie-produktu/doplňky-parkovacich-systemu/system-rozpoznavani-registracnich-znacek/>.
34. *ASParking*. [Online] 2024. <https://www.asparking.cz/cz/vyrobky/parkovaci-systemy/rozpoznavani-registracnich-znacek>.
35. *Smart4CitySpel*. [Online] 2024. <https://smart4city.spel.cz/kamera>.
36. *ZKTeco*. [Online] 2024. <https://zkteco.technology/en/catalog/smart-entrance-control/smart-terminal-smart-entrance-control/smart-vehicle-inspection/license-plate-recognition/>.
37. *Alimex*. [Online] 2024. <https://www.alimex.cz/index.php/technologie/identifikacni-technologie/identifikace-vozidel>.
38. *Designa*. [Online] 2024. <https://designa.com/en-us/license-plate-recognition-lpr>.
39. Co je edge computing. *APC, Schneider Electric*. [Online] 2024. <https://www.apc.com/cz/cs/solutions/business-solutions/edge-computing/what-is-edge-computing.jsp>.
40. Cloud a Edge computing. *Axians*. [Online] 2024. <https://www.axians.cz/nase-reseni/cloud-a-edge-computing/>.
41. Co je cloudový server. *Azure Microsoft*. [Online] 2024. <https://azure.microsoft.com/cs-cz/resources/cloud-computing-dictionary/what-is-a-cloud-server>.
42. Edge computing. *APC, Schneider*. [Online] 2024. <https://www.apc.com/cz/cs/solutions/business-solutions/edge-computing/>.

43. Mobilní aplikace pro ověření SPZ. *MATCH SYSTEMS*. [Online] 2024. https://www.machsystems.cz/produkty/rozpoznani-registracnich-znacek/mobilni-aplikace-overeni-spz?utm_source=chatgpt.com.
44. Serverové čtení SPZ. *MATCH SYSTEMS*. [Online] 2024. https://www.machsystems.cz/produkty/rozpoznani-registracnich-znacek/serverove-cteni-spz?utm_source=chatgpt.com.
45. AxxonNet. *AxxonSoft*. [Online] 2024. <https://www.axxonsoft.com/products/cloud-solutions/cloud-managed-vms>.
46. Archiv produktů. *EUROALARM*. [Online] 2024. <https://www.euroalarm.cz/eshop-zabezpecovaci-technika/dalsi-produkty/archiv-produktu/archiv-ids-2cd7a46g0-p-izhs8-32mmc>.
47. IP kamery Hikvision pro čtení, rozpoznání SPZ (ANPR). *EUROALARM*. [Online] 2024. <https://www.euroalarm.cz/bezpecnostni-kamerove-systemy-cctv/ip-kamery-hikvision-pro-specialni-urceni/cteni-rozpoznani-spz-hikvision-anpr-kamery/>.
48. SOFTPLUS LPR. *TSS GROUP*. [Online] 2024. <https://www.tssgroup.cz/produkty-1/zabezpecenie-objektov/tss-riesenia/softplus-lpr>.
49. Produkty. *NITTA Systems*. [Online] 2024. <https://www.nitta.cz/produkty/>.
50. Rozpoznávání SPZ. *IPSECURITY*. [Online] 2024. <https://www.rozpoznavanispz.cz/>.
51. Měření úsekové rychlosti. *SPEL*. [Online] 2024. <https://smart4city.spel.cz/mereni-usekove-rychlosti>.
52. Měření rychlosti. *CAMEA*. [Online] 2024. <https://www.camea.cz/cz/doprava/mereni-rychlosti/>.
53. Products. *GeoVision*. [Online] 2024. <https://www.geovision.com.tw/products.php?c1=7>.
54. Rozpoznávání SPZ. *ADI Global*. [Online] 2024. <https://adiglobal.cz/cz/produkty141/page/rozpoznavani-spz>.
55. Kamery pro čtení RZ. *Core - Protect*. [Online] 2024. <https://coreprotect.eu/kamery-pro-cteni-rz>.
56. License Plate Recognition. *MACH SYSTEMS*. [Online] 2024. <https://www.machsystems.cz/en/services/license-plate-recognition>.
57. Licence Plate Recognition. *Geutebrück*. [Online] 2024. <https://geutebrueck.cz/wp-content/uploads/sites/5/2020/01/g-tect-lpr.pdf>.
58. Hikvision kamery pro čtení a rozpoznání SPZ. *SMARTNESS*. [Online] 2024. <https://smartness.cz/kamerove-systemy/kamery-pro-cteni-a-rozpoznani-spz/>.
59. Doprava a cesty. *VISIONCRAFT*. [Online] 2024. <https://www.visioncraft.ai/doprava-a-cesty#co-a-jak-delame>.
60. Analýza obrazu z kamer s umělou inteligencí. *Certicon*. [Online] 2024. <https://www.certicon.cz/reseni/certiconvis/>.
61. Licence Plate Recognition (ALPR). *CCVis*. [Online] 2024. <https://ccvis.eu/funkce/license-plate-recognition/>.
62. Plate Recognizer Stream. *PLATE RECOGNIZER*. [Online] 2024. <https://platerecognizer.com/stream>.
63. *OpenALPR*. [Online] 2024. <https://www.openalpr.com/>.
64. *ATEAS*. [Online] 2024. https://www.absolon.cz/deploy/files/ateas_katalog_cz.pdf.
65. *VisionCraft*. [Online] 2024. <https://www.visioncraft.ai/ai-analyza-vidoa>.
66. *GoodVision*. [Online] 2024. <https://goodvisionlive.com/solutions/traffic-data-collection/>.
67. *Varnet*. [Online] 2024. <https://www.varnet.cz/zbozi/2206-078-axxon-one-enterprise-rozpoznani-spz-a-database>.
68. *HISYS*. [Online] 2024. <https://www.varnet.cz/zbozi/2206-078-axxon-one-enterprise-rozpoznani-spz-a-database>.
69. *ProfiComms*. [Online] 2024. <https://www.proficomms.cz/analyza-vidoa/c-981/>.
70. *GeoVision*. [Online] 2024. <https://www.geovision.com.tw/product/GV-LPR>.
71. Number Plate Reading. *Eyedeas*. [Online] 2024. <https://www.eyedeas.cz/cs/produkty/number-plate-reading-anpr>.
72. Stofan, D. The Development Of Traffic Data Collection Methods. *GoodVision*. [Online] 2023. [Citace: 20. 09 2024.] <https://blog.goodvisionlive.com/the-development-of-traffic-data-collection-methods>.

73. Červenka, P., Kruntorád, J. Optimizing the Operation of Portable Traffic Monitoring Devices Using Photovoltaic Panels for Sustainability and Efficiency. *2024 24th International Scientific Conference on Electric Power Engineering*. 1-5. doi: 10.1109/EPE61521.2024.10559514, 2024.
74. Město Český Krumlov. *Komplexní dopravní koncepce města Český Krumlov*. [Online] 2017. [Citace: 11. 05 2023.] https://www.ckrumlov.cz/docs/cz/20171222_Dopravni_koncepce.xml.
75. Rao, W., Wu, Y., Xia, J., Ou, J., Kluger, R. Origin-destination pattern estimation based on trajectory reconstruction using automatic license plate recognition data. *Transportation Research Part C: Emerging Technologies*. (95), 29-46, doi: 10.1016/j.trc.2018.07.002, 2018.
76. Oliveira-Neto, F. M, Han, L. D., Jeong, M. K. Online license plate matching procedures using license-plate recognition machines and new weighted edit distance. *Transportation Research Part C: Emerging Technologies*. 21(1), 306-320. doi: 10.1016/j.trc.2011.11.003, 2012.
77. Chen, Y., He, Z. Vehicle Identity Recovery for Automatic Number Plate Recognition Data via Heterogeneous Network Embedding. *Sustainability*. 12(8), 3074; <https://doi.org/10.3390/su12083074>, 2020.
78. Hajčiarová, E. Metodika vyhodnocení směrových dopravních průzkumů. *Diplomová práce*. ČVUT v Praze, Fakulta dopravní, 2023.