Projekt č. VI20172020068

Nástroje a metody zpracování videa a obrazu pro zvýšení efektivity operací bezpečnostních a záchranných složek (VRASSEO)

Detekce anomálií na obrazovém záznamu z dronu

Technická zpráva

Bc. David Bažout, Ing. Vítězslav Beran Ph.D.

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

Prosinec 2019

Obsah

1 Úvod **1**

2 Uživatelské úlohy **2**

3 Průzkum vhodných datových sad **4**
3.1 Termální datové sady 4
3.2 RGB datové sady 4
3.3 Návrh vlastní datové sady 5
3.4 Generování syntetických dat 6

4 Analýza metod pro detekci anomálie v davu **8**
4.1 Analýza davu na základě detekce objektů 8
4.2 Analýza příznakových vektorů 10
4.3 Využití architektury GAN 11

5 Návrh metody pro detekci anomálií v davu **12**
5.1 Části detekčního systému 12
5.2 Stabilizace obrazu 13
5.3 Extrakce příznaků a detekce anomálie 13
5.4 Vizualizace do mapového podkladu 14

6 Návrh uživatelské aplikace **15**

7 Závěr **16**

Abstrakt

Cílem této práce je analýza možností řešení uživatelských úloh v oblasti detekce anomálií z obrazového záznamu z dronu na základě posledních vědeckých výsledků. Z uživatelských úloh je zásadní zejména analýza davu z pohledu zrychlení, zastavení, rozdělení, srocení, zamaskování, detekce ležící osoby a detekce pohybu osoby mimo hlavní pohyb. Práce obsahuje průzkum vhodných datových sad, průzkum existujících přístupů pro detekci anomálie a návrh metody pro řešení těchto uživatelských úloh.

1 Úvod

V posledních letech dochází k rychlému rozvoji využití dronů v mnoha odvětvích. Své uplatnění nachází i v oblasti národních bezpečnostních složek. Cílem této práce je analýza možností řešení uživatelských úloh v oblasti detekce anomálií z obrazového záznamu z dronu na základě posledních vědeckých výsledků.

V první fázi byla práce zaměřena výhradně na vývoj metody umožňující detekci ležící osoby na videozáznamu z termokamery. Z tohoto důvodu byl proveden podrobný průzkum dostupných termálních datových sad a vhodných metod pro detekci objektů na těchto datech. Zde byl ovšem zásadní problém s nedostatkem vhodných dat, a proto byly prozkoumány způsoby generování syntetických datových sad pomocí neuronových sítí typu GAN a návrh podoby a scénáře pro nahrávání vlastní datové sady.

Později došlo k upřesnění požadavků na základě rozhovorů s uživateli, jehož výsledkem bylo zaměření na analýzu chování davu na obrazových datech. Z uživatelských úloh je zásadní zejména analýza davu z pohledu zrychlení, zastavení, rozdělení, srocení, zamaskování, detekce ležící osoby a detekce pohybu osoby mimo hlavní pohyb. Uživateli budou zpřístupněny výsledky analýzy skrze intuitivní grafické rozhraní.

Práce se dále zabývá detailním rozborem uživatelských úloh v kapitole 2, průzkumem vhodných datových sad v kapitole 3, analýzou metod pro detekci anomálie v davu v kapitole 4, návrhem metody vhodné pro řešení uživatelských úloh v kapitole 5 a nakonec návrhem podoby uživatelské aplikace v kapitole 6.

2 Uživatelské úlohy

Pro hlubší pochopení problému je důležité provést nejprve přesnou specifikaci typických uživatelských úloh společně s případy jejich užití. Vstupní data pro veškeré úlohy představuje videozáznam z kamery umístěné na dronu. Lze předpokládat, že dron bude ve stacionární poloze se záběrem prostoru o rozloze zhruba 500 x 500 metrů. V současné době lze uvažovat obrazová data z oblasti viditelného spektra (RGB), ale v budoucnu je možné rozšíření o data z termální kamery. Vlivem povětrnostní situace, trajektorie letu, dostupné vzdálenosti o cílového místa a dalších okolností mohou být data snímána se záchvěvy někdy i výraznějšího charakteru. Na datech se dále může vyskytnout nízké osvětlení v ranních a večerních hodinách, zhoršená viditelnost vlivem dýmu nebo mlhy a další faktory snižující kvalitu záznamu. Příklad obrazového zdroje je na obrázku 1.

Obrázek 1: Ukázka snímků davu v reálných podmínkách.

Potřebou uživatele je sledování mnohdy i rozsáhlého prostoru a rychlá reakce v případě výskytu neobvyklé a potenciálně nebezpečné situace v davu. Dav v této úloze představuje shromáždění osob s vysokou hustotou, ale je nutné uvažovat i analýzu shromáždění osob s řídkým výskytem.

Uživatel potřebuje mít přehled o situaci celkového proudění davu osob a potřebuje být upozorněn zejména na jeho zrychlení, zastavení, rozdělení nebo srocení. Tyto situace nemusí vždy představovat bezpečnostní riziko - změna proudění může být ovlivněna červenou signalizací na semaforu. Z druhé strany pohledu může být výraznější změna proudění davu způsobena rvačkou, nepokoji nebo jinou situací, která vyžaduje zásah hlídky. Vyhodno-

cení potenciálního nebezpečí je v odpovědnosti uživatele a i z tohoto důvodu je zde tolerován výskyt falešných upozornění.

K vyhodnocení rizika ve změně proudění davu může výrazným způsobem napomoci analýza změny vzhledu davu, která může být způsobena například maskováním osob před hrozícím nebezpečím. Navrhovaná metoda by této vizuální informace měla využívat a zpřístupňovat ji pro uživatele.

Další uživatelskou úlohou je upozornění na ležící osobu z důvodu včasného zásahu lékařů v případě jejího zranění. Ležící osobu je nutné detekovat i ve větším shromáždění osob, kde hrozí její částečné nebo úplné zakrytí na záznamu z dronu. Pro tuto úlohu by mohlo být významné využití termálních dat pro použití v noci a je vhodné tuto alternativu prozkoumat.

Poslední zkoumanou úlohu představuje detekce osoby provádějící netypický směr pohybu. Může se jednat o opačný směr pohybu oproti ostatním osobám v davu s hustým zastoupení. Tato situace může značit například snahu osoby uniknout policejní hlídce a její detekce může napomoci jejímu rychlejšímu zadržení. Druhým příkladem je netypický směr nebo rychlost pohybu vzhledem k danému místu. Pro příklad mohu uvést situaci, kdy osoba přistupuje k nějakému objektu cestou určenou pro východ, což může a nemusí být podezřelé. V tomto případě zde ani nemusí docházet k výskytu ostatních osob.

3 Průzkum vhodných datových sad

Po přesné specifikaci uživatelských úloh a případů jejich užití lze provést průzkum a výběr vhodných datových sad. Vzhledem k poměrně úzké specifikaci úloh nelze očekávat výskyt datové sady jednoznačně určené pro potřebu této práce. Při hledání je nutné zaměřit se na alespoň částečně odpovídající datové sady, které by měly odpovídat následujícím požadavkům.

Při hledání je vhodné zaměřit se i na průzkum datových sad s termálními snímky vzhledem k možnosti jejich budoucího využití. Je vhodné, aby datová sada obsahovala záznamy snímané z dronu, případně alespoň ze střechy budovy. Záznamy z malé výšky jsou oproti záznamům dronu příliš odlišné a tudíž pro tuto úlohu nevhodné. Vhodná datová sada by měla obsahovat scény s dostatečnou hustotou výskytu osob. Z anotací je pro tuto úlohu klíčová specifikace výskytu anomálie na daném snímku, informace o směru proudění davu a případně oblast detekovaných osob a dalších tříd objektů k odvození informací o jejich pohybu a jejich kontextu ke scéně.

3.1 Termální datové sady

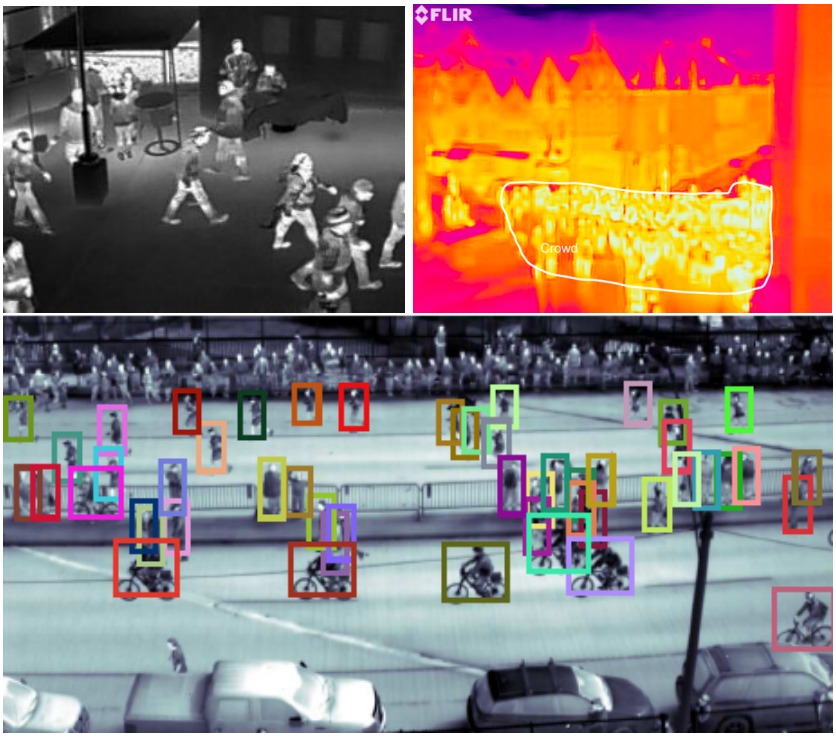
Z volně dostupných datových sad v infračerveném spektru předcházejícímu popisu alespoň částečně odpovídají datové sady CAMEL Dataset [1], BU-TIV (Thermal Infrared Video) Benchmark [2] a The Last Post Dataset [3]. CAMEL Dataset obsahuje několik vhodných sekvencí s vyšší hustotou výskytu osob. Celkem 5000 snímků o rozlišení 336x256 px je dostupných v RGB a IR spektru, což může být využitelné pro generování syntetických dat. Příklad dat je na obrázku 2.

Datová sada BU-TIV dataset obsahuje pouze infračervené snímky o rozlišení 1024x512 nebo 512x512px o celkovém počtu 9000. Ke snímkům jsou přiloženy anotace pro veškeré chodce a další třídy objektů. Ukázka je na obrázku 2.

Datová sada Last Post dataset je pravděpodobně nejblíže očekávané podobě dat. Skládá se ze záznamů ulice s dostatečnou hustotou výskytu osob. Celá scéna je snímána pravděpodobně ze střechy budovy. Jejím problémem je absence anotací o pohybu objektů a velmi nízké rozlišení 160x120px. Naproti tomu ji tvoří zhruba 230 minut záznamu. Ukázka je na obrázku 2.

3.2 RGB datové sady

Z průzkumu je tedy zřejmé, že datových sad v infračerveném spektru je pouze omezené množství a relevantnost vůči řešenému problému je rovněž spekulativní. Datové sady se záznamy lidských davů snímaných pomocí dronů nebo z vyvýšených míst jsou mnohem dostupnější v RGB spektru. Datová sada Train Station Dataset [4] obsahuje 30 minutový záznam rozsáhlého prostoru nádražní haly (ukázka na obrázku 3). Záznam je o rozlišení 720x480



Obrázek 2: Ukázka snímků z termálních datových sad. Ve směru od levého okraje se jedná o Camel, Last Post Dataset a dole BU-TIV.

px a jsou k němu dostupné anotace pro směrnice pohybu osob. Multidrone Public Dataset je rozsáhlá datová sada sestávající z několika desítek hodin různorodých záznamů s výskytem scén lidských davů snímaných dronem. Datová sada je v HD rozlišení a pro akademické účely je volně dostupná na vyžádání. Tuto datovou sadu získal řešitelský tým k dispozici těsně před koncem roku a nyní probíhá její analýza.

3.3 Návrh vlastní datové sady

Z existujících datových sad infračerveného spektra zde není žádná datová sada zabývající se analýzou davu (odhad velikosti, detekce anomálie, tracking jednotlivce v davu) [5]. Existující datové sady jsou obvykle zaměřeny na detekci chodců na ulici, která je snímána pomocí stacionární kamery. Na těchto datech je ale nízká hustota výskytu snímaných osob a pro účely analýzy davu jsou tyto záznamy obtížně využitelné.

Pro úlohu detekce anomálie v davu na infračervených snímcích není řešitelskému týmu znám dostupný dataset. Práce zabývající se detekcí anomálie referují výhradně datovou sadu z tohoto článku [6]. Datová sada ale obsahuje šedotónové záznamy chodníku v parku, kde je za anomálii považován výskyt objektu různého od chodce. Ze získaných informací lze usuzovat, že pořízení

Obrázek 3: Ukázka snímku z datové sady Train Station Dataset.

datové sady zabývající se analýzou davu a zejména detekce anomálie by mohlo být pro odbornou komunitu přínosem.

Podle průzkumu [5] je v odborné komunitě nedostatek datových sad zabývajících se detekcí anomálie v davu na snímcích v RGB spektru a pro IR spektrum zde taková datová sada není žádná. Z tohoto důvodu by bylo vhodné pořídit multispektrální datovou sadu (RGB a T) a pomoci tím vývoji metod v obou spektrech a zároveň otevřít cestu k trénování modelu pro generování syntetických dat na základě RGB spektra.

Z důvodu uplatnění v projektu VRASSEO i vysoké využitelnosti pro aplikace reálného světa by bylo vhodné datovou sadu zaměřit na záznamy ze sportovních akcí nebo koncertů s vysokou návštěvností snímaných z dronu nebo pomocí statické kamery umístěné na střeše budovy. Za žádoucí lze považovat pořízení denních i nočních záznamů a záznamů za různého počasí a teplot. V záznamu by bylo vhodné provést a později anotovat scény popisované v kapitole 2. Ležící osobu nebo osobu pohybující se netypickým směrem či rychlostí vzhledem k davu lze anotovat pomocí rámečku s přiřazením do příslušné kategorie. Lokální anomálie odpovídající netypickému chování davu lze specifikovat pomocí masky celočíselných hodnot, které opět odpovídají zařazení do vhodné kategorie.

3.4 Generování syntetických dat

V současné době se objevují první výsledky nového přístupu tzv. "zero-shot learning" [7], který je velmi slibný při trénování modelů v doméně, pro které nejsou trénovací data. Přesto zatím pro využití konvolučních neuronových sítí obecně platí, že je potřeba značné množství trénovacích dat. Z důvodu

nedostatku datových sad z infračerveného spektra množství řešení pro detekci objektu nebo anomálie v infračervených datech stále využívá klasických metod. Tím vzniká motivace pro generování syntetických datových sad pro rozšíření stávajících datasetů.

Odborných prací zabývajících se touto problematikou není v současnosti mnoho. V článku [8] je popisována metoda generování termálních snímků na základě dvojic obrázků z RGB a IR spektra pomocí architektury neuronové sítě pix2pix. Je zde popisována i metoda umožňující překlad obrázku do termálního spektra bez trénovacích dvojic s pomocí architektury CycleGAN. Tato metoda ovšem dosahuje výrazně horších výsledků.

Celkový dopad synteticky generovaných dat davových scén na proces trénování metod detekce objektu nebo anomálie a jejich následnou úspěšnost na reálných datech vyžaduje důkladnější prozkoumání.

4 Analýza metod pro detekci anomálie v davu

Detekce anomálie v davu je stále výrazně otevřeným problémem a to nejen v oblasti vývoje metod, ale i v pouhé interpretaci problému. Obecně lze za jádro úlohy považovat modelování obvyklého chování davu [5]. Aktuální chování davu je pomocí algoritmu porovnáváno s modelem, na jehož základě je vyhodnoceno běžné a netypické chování. Za netypické chování lze považovat změnu rychlosti nebo směru pohybu, změnu vzhledu nebo výskyt netypického objektu.

Úlohu detekce anomálie lze rozdělit na globální a lokální [5]. U globální detekce je výstupem algoritmu pouhá informace o výskytu anomálie. Lokální detekce anomálie připojuje k globální detekci ještě informace o její poloze. Modelování obvyklého chování davu lze provádět za přítomnosti učitele i bez učitele.

Většina metod pro detekci abnormalit závisí na složitých a ručně získávaných příznacích pro reprezentaci pohybu a vzhledu davu. Ukázalo se, že konvoluční neuronové sítě mohou být mocným nástrojem pro extrakci příznaků, které mohou být využity pro detekci anomalit [9]. V případě malých datových sad jsou ale náchylné na přílišnou adaptaci na trénovací data a neschopnost dostatečného zobecnění. Nedostatek kvalitních datových sad lze považovat za jednu z největších překážek v řešení této úlohy [9, 5].

Vzhledem k nedostatečné formalizaci celého problému a v souvislosti s uživatelskými potřebami, které jsou popisovány v kapitole 2, lze na úlohu detekce anomálie rovněž nahlížet jako na detekci nežádoucího objektu. V tomto případě lze anomálii modelovat jako šablonu popisující nežádoucí objekt a využít k tomu obecné metody detekce.

4.1 Analýza davu na základě detekce objektů

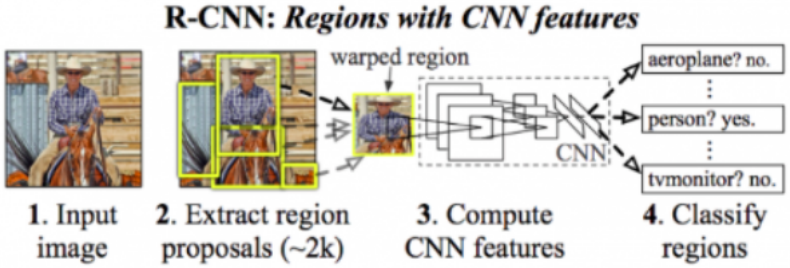
Řešením úlohy detekce objektu ve snímku je nalezení jeho ohraničujícího rámečku. Existující metody detekce lze rozdělit na metody klasické a metody využívající neuronových sítí. Princip klasické metody spočívá v extrakci příznakového vektoru aplikací určité matematické funkce na výřez obrázku. O výskytu objektu v daném výřezu rozhoduje na základě extrahovaného příznakového vektoru klasifikátor. Velice často jsou v odborné literatuře zmiňovány příznaky typu HAAR [10], LBP [11] nebo HOG [12] v kombinaci s kaskádovým nebo SVM klasifikátorem. Výhodou klasických metod zůstává nutnost menšího počtu trénovacích dat oproti neuronovým sítím. Klasická metoda lze využít v kombinaci s R-CNN architekturou pro generování region proposal [13].

Metody založené na konvolučních neuronových sítích jsou obvykle složené ze základní a klasifikační sítě [3]. Základní síť se využívá pro extrakci příznakového vektoru pro danou oblast ve snímku. Příznakový vektor slouží jako vstup pro klasifikátor, který rozhoduje o výskytu objektu na daném

místě. Architektura základní sítě se liší podle konkrétní úlohy, ale obvykle se skládá z konvolučních a pooling vrstev. Klasifikační síť obvykle využívá plně propojených vrstev neuronů.

První skupina detektorů připojuje přímo na výstup základní sítě klasifikátor rozhodující o výsledku detekce. Do této kategorie spadají například v literatuře často zmiňované detektory založené na architekturách R-CNN a YOLO [3]. Pro obě architektury existují volně dostupné implementace i nástroje pro trénování vlastního detektoru pro specifický objekt. Existují i předtrénované modely, kterých lze využít u přístupu přeneseného učení. Společnou vlastností těchto metod je omezená změna měřítka, se kterou jsou schopny objekt detekovat.

Druhá skupina detektorů (SSD, MS-CNN) připojuje klasifikátor do různých úrovní v architektuře neuronové sítě a částečně tím řeší problém s měřítkem objektu. Obě metody využívají sliding window klasifikátorů pohybujících se po extrahovaných příznakových mapách. Důležitá je volba parametrů pohyblivého okna.



Obrázek 4: Struktura detekce objektu pomocí R-CNN. Převzato z [3].

Problém nastává při detekci objektu na termálních snímcích, protože jsou tvořena pouze jediným kanálem, zatímco nejběžnější architektury pracují se 3 barevnými kanály. V odborné literatuře bylo uvedeno několik postupů, jakým způsobem lze neuronovou síť na termální data adaptovat.

Podle experimentu z článku Multispectral Pedestrian Detection using Deep Fusion Convolutional Neural Networks [14] lze provést redukci počtu konvolučních jader v architektuře základní sítě sloužící pro extrakci příznakového vektoru a tím ji lépe adaptovat na jednokanálová data. S úpravou architektury již nelze využít dříve předtrénovaných vah a vznikají tedy vyšší nároky na rozsah datové sady. Tento problém lze částečně řešit předtrénováním základní sítě skrze architekturu autoencoder na termálních datech bez anotací a na anotovaných datech lze poté dotrénovat klasifikační část. Články [3] a [14] popisují dosažení lepších výsledků se zavedením batch normalizace.

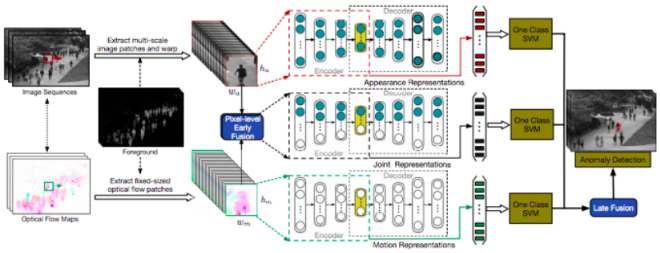
Kromě využití existující architektury určené primárně pro detekci objektu v RGB snímcích existují práce [15, 14], kde autoři experimentují s

návrhem speciální architektury neuronové sítě pro tuto konkrétní úlohu. V článku [14] je popisována architektura, která je určena pro 4 kanály vstupních dat (RGB + termální kanál). Z důvodu nedostatku trénovacích dat pro infračervené snímky byla neuronové síť předtrénována na RGB datech tím způsobem, že byl IR kanál nahrazen kopií kanálu R ze spektra RGB. Proces předtrénování byl složený ze dvou kroků - trénování úlohy klasifikace na datové sadě ImageNet a jemné doladění vah na IR obrázcích z datové sady CALTECH benchmark. S použitím předtrénovaných vah konvoluční části byly na datové sadě KAIST natrénovány váhy plně propojených vrstev určených pro klasifikaci.

4.2 Analýza příznakových vektorů

Analýza příznakových vektorů je jednou z dalších metod detekce anomálie. Myšlenkou tohoto přístupu je reprezentace obvyklého chování pomocí vhodného příznakového vektoru a samotné řešení úlohy detekce anomálie spočívá v detekci outlierů porovnáním s dříve extrahovanými příznaky [6, 16]. V některých článcích je referován ještě mezikrok v podobě clusterizace příznakových vektorů za účelem získaní jejich diskrétního identifikátoru [17]. Příznaky lze extrahovat pomocí klasických přístupů, ale v posledních letech věnuje odborná literatura větší pozornosti extrakci příznaků s využitím neuronové sítě z důvodu jejich větší robustnosti vůči nepříznivým vlivům [5].

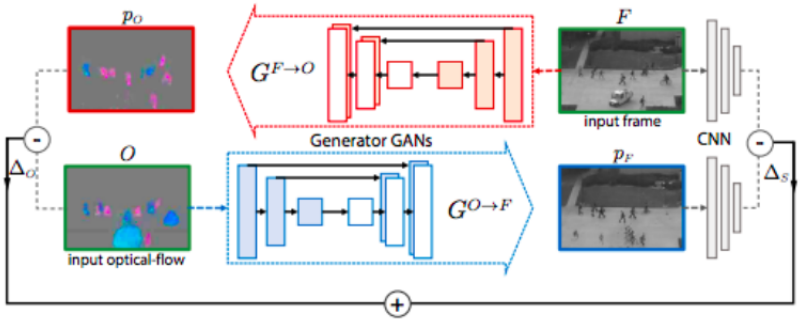
Jeden z nejnovějších přístupů detekce anomálie je publikován v článku Learning Deep Representations of Appearance and Motion for Anomalous Event Detection [16]. Hlavní myšlenkou metody je extrakce pohybových, vzhledových a fůzních příznaků pomocí hlubokých neuronových sítí a jejich následná binární klasifikace. Fůzní příznakový vektor je extrahovaný ze snímku vzniklého fůzí optického toku a obrazových dat. Vstupní snímek je rozdělen do mřížky tvořené buňkami. Pro každou buňku jsou extrahovány příznakové vektory popisující její stav. Výsledné skóre o výskytu anomálie je vyhodnocováno nezávisle na ostatních buňkách pomocí hierarchie binárních SVM klasifikátorů. O výskytu anomálie poté rozhoduje kombinace hodnot skóre pro jednotlivé složky. Ilustrace metody je na obrázku 5.



Obrázek 5: Ilustrace metody Learning Deep Representations of Appearance and Motion for Anomalous Event Detection. Převzato z [16].

4.3 Využití architektury GAN

Do samostatné kategorie spadá metoda publikovaná v článku Abnormal Event Detection in Videos Using Generative Adversarial Nets [9]. Je zde popisována metoda využívají neuronových sítí typu GAN. Neuronová síť typu GAN je natrénována na snímcích bez výskytu anomálie v kombinaci s jejich optical flow variantou. Vzhledem k tomu, že se neuronová síť při procesu trénování nesetká s výskytem anomálie, není schopna ji ani vygenerovat. Během procesu detekce jsou porovnána reálná a uměle vygenerovaná data a každá odlišnost je považována za anomálii. Schéma algoritmu této metody je na obrázku 6.



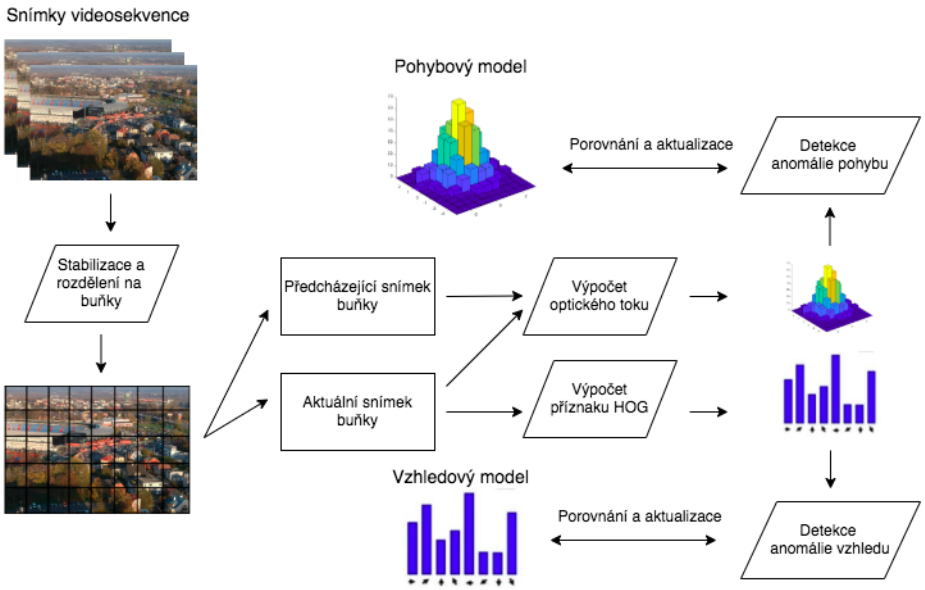
Obrázek 6: Schéma algoritmu metody využívající architektury GAN pro detekci anomálie. Převzato z [9].

5 Návrh metody pro detekci anomálií v davu

Na základě analýzy aktuálních uživatelských potřeb (viz kapitola 2), jejich priorit a analýzy dostupných metod a datových sad, bude navržen nejdříve koncept systému a základní metoda pro detekci změny pohybu v davu s využitím klasických přístupů. Toto základní experimentální řešení bude sloužit pro další vývoj a využití pokročilých metod.

5.1 Části detekčního systému

Navrhovaný systém se skládá ze základních funkčních bloků zobrazených ve schématu na obrázku 7. Snímky videozáznamu jsou nejdříve stabilizovány a poté rozděleny do mřížky tvořené buňkami. Pro každou buňku je proveden výpočet histogramu optického toku, který je následně porovnán s pohybovým modelem. Na základě jejich odchylky je detekována míra anomálie vzhledem k chování na předchozích snímcích. Analogický postup je realizován pro extrakci vzhledového příznaku. Jednotlivé funkční bloky jsou podrobněji specifikovány dále v textu.



Obrázek 7: Blokové schéma detekčního systému.

5.2 Stabilizace obrazu

Ačkoliv je stabilizace obrazu řešena v snímacím zařízením (jak závěsné zařízení pro kameru, tak vlastní kamera), vzniká řada situací, kdy toto nestačí - např. krátký výpadek v přenosu dat, nečekaná změna nastavení expozice, prudší závan větru, výkyv pozice dronu operátorem apod.

Drobné korekce nestability záznamu lze vyřešit detekcí klíčových bodů a jejich sledováním do následujícího snímku. Z trajektorie jejich posunu lze vypočítat transformační matici sloužící ke stabilizaci. Pro tento účel jsou vhodné např. metody FAST+BRIEF [18] z důvodu jejich nízké výpočetní náročnosti. Tato metoda je náchylná k chybě v případě, že detekované klíčové body vykovávají ve scéně pohyb. Do hledání vhodných klíčových bodů může být zapojen i uživatel s tím, že systému specifikuje statické oblasti v obrazu.

Při stabilizaci obrazu předchozí metodou může dojít po zpracování většího množství snímků k hromadění chyby. Další odchylka může vzniknout při výskytu výraznějšího záchvěvu obrazu. Z tohoto důvodu je vhodné jednou za určitý počet snímků provést detekci spolehlivějších klíčových bodů (např. SIFT [18]). Jejich detekci nelze provádět pro každý snímek z důvodu vysoké výpočetní náročnosti. Mezi klíčovými body aktuálního a dřívějšího snímku jsou porovnávacím algoritmem (např. Flann-based matcher [18]) nalezeny odpovídající dvojice. Hledání prostorové transformace (např. metodou RANSAC [18]) mezi dvojicemi klíčových bodů určuje transformační matici pro stabilizaci obrazu.

5.3 Extrakce příznaků a detekce anomálie

Po stabilizaci obrazu následuje jeho rozdělení do mřížky složené z menších buněk. Vhodný rozměr mřížky závisí na různorodosti snímané scény, ale zároveň má výrazný vliv i na výpočetní náročnost.

Informaci o pohybu jednotlivých pixelů mezi dvojicí snímků lze získat výpočtem hustého optického toku např. pomocí Gunner-Farnebackova algoritmu [18], který je implementován v knihovně OpenCV. Později lze využít přesnější analýzy pomocí konvoluční neuronové sítě.

Pro každou buňku mřížky je uchováván model obvyklého pohybu davu v podobě normalizovaného 2D histogramu. Pixely buňky jsou podle rychlosti a směru pohybu rozděleny do vhodného počtu skupin a hodnota histogramu pro danou kombinaci rychlosti a směru představuje četnost jejího výskytu. Porovnáním histogramu optického toku v aktuálním snímku s histogramem optického toku pro určitý počet předcházejících snímků lze určit hodnotu odchylky od běžného chování pro danou buňku.

Podobně jako pro pohyb je pro každou buňku vypočten model její vzhledové stránky. Pro prvotní implementaci lze extrahovat vzhledový příznak pomocí některé z klasických metod (např. HOG [12]). Aplikací vhodné

statistické metody na sekvenci historických hodnot příznaku lze získat jeho obvyklou hodnotu. Porovnáním aktuální hodnoty s hodnotou obvyklou lze určit míru odchylky od běžného stavu pro danou buňku.

Výsledná implementace by měla váhovat příspěvek jednotlivých pixelů podle jejich vzdálenosti od středu buňky např. pomocí Gaussovy křivky. Eliminuje se tím negativní vliv posunu výraznějších pixelů na okrajích buněk. Ze stejného důvodu je vhodné, aby se okraje buněk do určité míry překrývaly.

5.4 Vizualizace do mapového podkladu

Výstupem detekce anomálie je tedy hodnota odchylky od obvyklého stavu pro každou z buněk mřížky. S ohledem na prostorový charakter dat představuje jejich vhodnou grafickou reprezentaci teplotní mapa. Teplotní mapa je typ grafického zobrazení, ve kterém je každá hodnota reprezentována barvou spojitého barevného spektra. Zobrazení teplotní mapy jako další vrstvy mapového podkladu představuje velice intuitivní způsob vizualizace výsledků analýzy.

Pro vizualizaci teplotní mapy do mapového podkladu je nezbytné nalezení transformace mřížky do reálné scény. Jedním ze způsobů řešení tohoto problému je nalezení odpovídajících klíčových bodů v satelitních snímcích a ve videu z dronu. Zjednodušení může představovat uživatelský vstup s přibližnou nebo v počáteční fázi implementace úplnou specifikací polohy videozáznamu vzhledem k mapovému podkladu.

6 Návrh uživatelské aplikace

Cílem uživatelské aplikace je jednoduché a intuitivní zpřístupnění výsledků analýzy chování davu skrze grafické uživatelské rozhraní. Jednou z možností je vizualizovat stav pohybu davu přímo do mapových podkladů (např. mapy.cz). Koncept vyvíjené uživatelské aplikace je následující.

Klíčové pro efektivní vizualizaci výsledků je zobrazení výsledků detekce např. přímo do topografické mapy je znalost transformace mezi videozáznamem a mapového podkladu. Toto by bylo možné řešit zcela automatickou metodou, která by byla schopna při znalosti orientační polohy kamery (GPS) vypočítat s využitím mapových podkladů její přesnou polohu a směr pohledu. Vývoj takovéto zcela automatické pokročilé netriviální metody je ovšem značně drahý a pro potřeby projektu není aktuálně klíčový. Počáteční transformaci lze zadat uživatelem ručně a lze tak vyvinout pouze metodu, která bude tuto transformaci automaticky udržovat. Prvním krokem při používání aplikace tedy bude získat od uživatele několik vzájemně odpovídajících bodů, podle kterých bude nalezena transformace mezi souřadnicemi videozáznamu a topologické mapy.

Hlavní vizualizační komponentou tedy bude mapový podklad, do kterého budou vizualizovány výsledky detekce. Tato mapová vrstva má podobu teplotní mapy, jejíž barva odpovídá hodnotě odchylky od běžné situace. Vzhledem k výstupů analýzy je vhodné umožnit uživateli specifikovat atributy metody (prostorové rozlišení a citlivost, délku historických záznamů a případnou možnost smazání historie). Součástí bude panel pro nastavení chování aplikace (parametry alarmu, citlivé oblasti apod.), popř. chování detekční metody. Finální vlastnosti aplikace a GUI budou upřesněny na základě testování a dialogu s potencionálními uživateli.

7 Závěr

Proběhla analýza možností řešení uživatelských úloh v oblasti detekce anomálií na obrazovém záznamu z dronu. Na základě zpřesnění uživatelských požadavků vyplynula důležitost zejména analýzy davu z pohledu detekce nečekaného chování: zrychlení, zastavení, rozdělení, srocení, zamaskování, detekce ležící osoby a detekce pohybu osoby mimo typický pohyb davu.

Analýza vědeckých výsledků v této oblasti ukázala možnosti a omezení při využití jak obrazových dat, tak termo-snímků. Výsledek analýzy vedl k výběru úloh, které lze řešit na základě pouze obrazových dat s využitím sledování pohybu/změny v obraze (optický tok, obrazové příznaky a jejich změna) a následné analýzy historie těchto změn a detekce netypických vzorů v těchto změnách. Provedené dílčí experimenty s různými konfiguracemi hlubokých neuronových sítí zatím nevedly v praxi k použitelnému řešení a na vývoji těchto metod se bude dále pokračovat. Současně započal vývoj základní verze řešení s využitím klasických metod zpracování obrazu.

Pro potencionální uživatele byl navržen koncept a rozpracována základní verze GUI aplikace, která bude integrovat vyvinutou metodu pro detekci anomálií v pohybu davu osob, umožní úpravu klíčových parametrů metody a vizualizaci výsledků detekce efektivním způsobem - vykreslení do topografické mapy, možnosti alarmu apod. Výsledná podoba využití výsledků metody bude ještě konzultována s potencionálními uživateli.

Literatura

[1] E. Gebhardt and M. Wolf, "Camel dataset for visual and thermal infrared multiple object detection and tracking," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, IEEE, 2018.

[2] Z. Wu, N. Fuller, D. Theriault, and M. Betke, "A thermal infrared video benchmark for visual analysis," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2014.

[3] E. Valldor, "Person detection in thermal images using deep learning," 2018.

[4] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2871–2878, IEEE, 2012.

[5] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan, "Crowded scene analysis: A survey," *IEEE transactions on circuits and systems for video technology*, vol. 25, no. 3, pp. 367–386, 2014.

[6] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1975–1981, IEEE, 2010.

[7] W. Wang, V. W. Zheng, H. Yu, and C. Miao, "A survey of zero-shot learning: Settings, methods, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, p. 13, 2019.

[8] V. V. Kniaz, V. A. Knyaz, J. Hladuvka, W. G. Kropatsch, and V. Mizginov, "Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 0–0, 2018.

[9] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, "Abnormal event detection in videos using generative adversarial nets," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 1577–1581, IEEE, 2017.

[10] P. Viola, M. Jones, *et al.*, "Robust real-time object detection," *International journal of computer vision*, vol. 4, no. 34-47, p. 4, 2001.

[11] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European conference on computer vision*, pp. 469–481, Springer, 2004.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp. 91–99, 2015.

[14] J. Wagner, V. Fischer, M. Herman, and S. Behnke, "Multispectral pedestrian detection using deep fusion convolutional neural networks.," in *ESANN*, 2016.

[15] C. D. Rodin, L. N. de Lima, F. A. de Alcantara Andrade, D. B. Haddad, T. A. Johansen, and R. Storvold, "Object classification in thermal images using convolutional neural networks for search and rescue missions with unmanned aerial systems," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2018.

[16] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe, "Learning deep representations of appearance and motion for anomalous event detection," *arXiv preprint arXiv:1510.01553*, 2015.

[17] M. Ravanbakhsh, M. Nabi, H. Mousavi, E. Sangineto, and N. Sebe, "Plug-and-play cnn for crowd motion analysis: An application in abnormal event detection," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1689–1698, IEEE, 2018.

[18] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.