

*Studentská tvůrčí a odborná činnost*  
**STOČ 2017**

**Detekce a analýza pohybu osob založená na analýze obrazu**

**Bc. Robin Antonič**

Mendelova univerzita v Brně, Zemědělská 1

**20. dubna 2017**  
FAI UTB ve Zlíně

**Klíčová slova:** Optický tok, Klasifikační modely, Známé detektory osob, Framework.NET, Analýza obrazu

**Anotace:** Práce se zabývá detekcí a analýzou trajektorií pohybu osob za účelem vyhodnocení pohybu osob ve vytyčeném prostoru z obrazu kamer. V první části práce je provedena komparace vybraných algoritmů pro výpočet optického toku, které slouží pro detekci dynamických objektů v obraze. Další část práce je věnována komparaci vybraných detektorů osob v obraze a klasifikačních modelů, které slouží pro verifikaci trajektorií.

Dále je navržen algoritmus, který využívá metodu optického toku TV-L1 pro detekci dynamických objektů. Tyto objekty jsou poté spárovány pomocí Maďarské metody, pokud detekce objektů selže a objekty nemohou být spárovány, tak je aplikován Kalmanův filtr pro predikci aktuálního stavu objektu. Pomocí neuronové sítě je provedena klasifikace, zda nalezené trajektorie patří lidem či ne. V dalším kroku je provedena analýza trajektorií.

## **Obsah**

1.	Cíl práce .....	4
2.	Princip algoritmu .....	4
3.	Vlastní práce .....	6
3.1	Pořízení testovacích dat .....	6
3.2	Komparace algoritmů optického toku.....	7
3.3	Komparace metod pro verifikaci osob .....	8
4.	Vyhodnocení přesnosti algoritmu .....	10
5.	Závěr .....	10
6.	Literatura.....	10

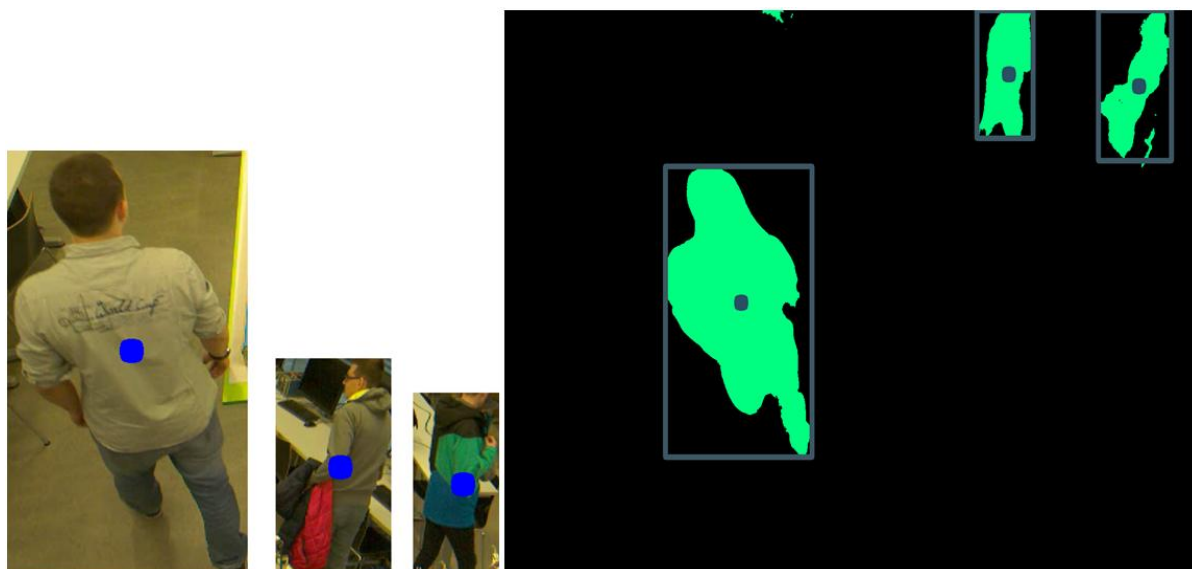
## 1. Cíl práce

Cílem práce je návrh a implementace algoritmu pro detekci a analýzu trajektorií pohybu osob z videosouboru či množiny snímků na základě analýzy obrazu. Dále byla vytvořena řídicí aplikace pomocí frameworku .NET, Accord.NET a knihovny EmguCV. Aplikace poskytuje funkcionalitu pro prezentaci a uložení získaných informací, které byly získány výše zmíněným algoritmem. V rámci vyhodnocení byla stanovena úspěšnost navrženého algoritmu.

## 2. Princip algoritmu

Hlavní myšlenka algoritmu je rozdělena do dvou částí. V první části jsou detekovány trajektorie pohybu objektů z celého videosouboru či jen jeho určité souvislé části. V druhém celku je provedena analýza detekovaných trajektorií pohybu. Algoritmus byl implementován pomocí frameworku .NET, knihovny EmguCV a frameworku Accord.NET.

V prvním kroku jsou v obraze detekovány dynamické objekty pomocí vybraných algoritmů pro výpočet optického toku, jedná se zejména o algoritmy TV-L1[3], Farneback, Brox[2] a Lucas-Kanade. Detekované dynamické objekty jsou následně převedeny do jednotné barvy pomocí segmentační metody prahování a poté je na objekty aplikována segmentační metoda rozrůstání oblastí viz. *Obrázek 1*. Následně jsou dynamické objekty filtrovány podle hustoty svého obsahu a také podle šířky a výšky takovým způsobem, aby poměrově odpovídali člověku. Poté jsou objekty extrahovány z původního obrazu a následně jsou vypočteny těžiště objektů, viz *Obrázek 1*.



Obrázek 1: V levé části jsou zobrazeny finální extrahované objekty s vypočtenými těžišti.  
V pravé části je zobrazen obraz, který obsahuje objekty před filtrováním.

Detekované a vyfiltrované objekty poté představují pohybující se osoby pro konkrétní snímek videosouboru. Trajektorie pohybu jsou tedy složeny z postupně detekovaných dynamických objektů z vybraného či celého časového intervalu videosouboru či množiny snímků.

Zde nastává situace, kdy se musí určit příslušnost objektů ke konkrétní trajektorii, tedy musí se vyřešit přiřazovací problém. Zde byla využita maďarská metoda pro spárování objektů mezi jednotlivými snímky. Jako klíčovou hodnotu pro porovnání jsme použili vzdálenost těžišť nalezených objektů napříč jednotlivými snímky. V rámci práce jsme museli také vyřešit

problém, kdy jsou nalezené objekty velmi blízko u sebe a splynou do jednoho detekovaného objektu nebo pokud dráhy trajektorií jsou překříženy a tím pádem mají společný průsečík.

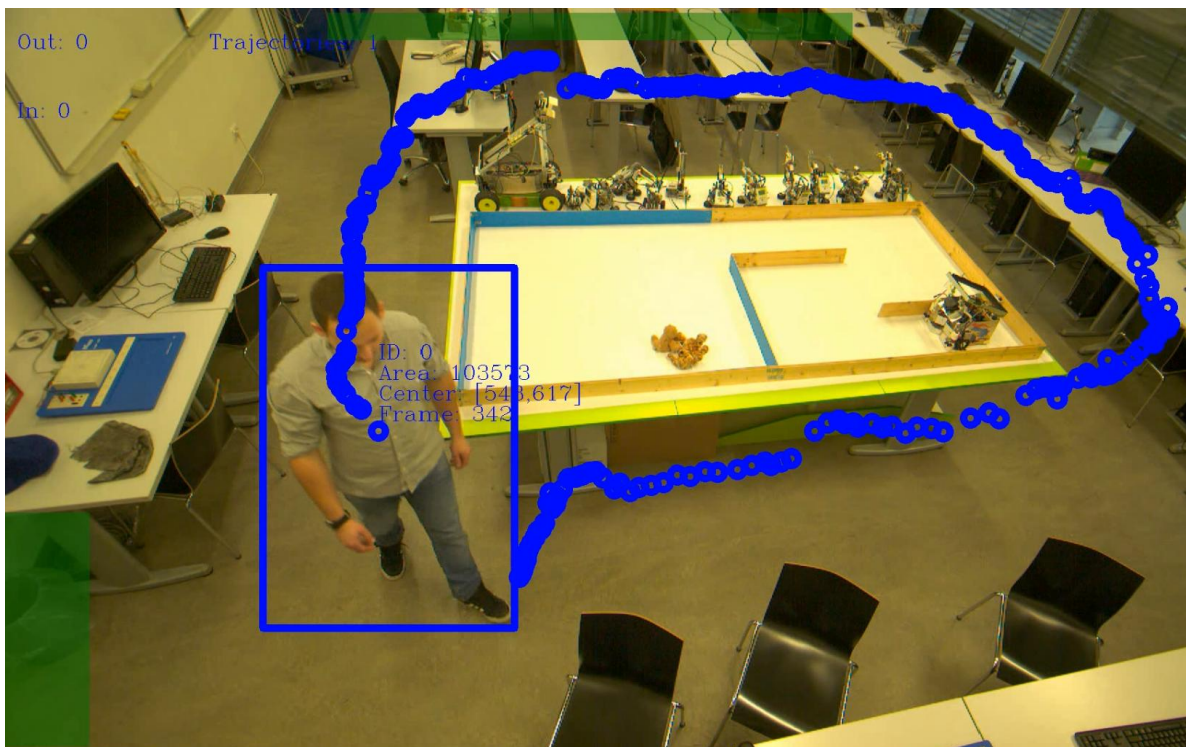
Dalším problémem nastává, pokud dynamické objekty zpomalí a mají velmi malou rychlost. Tyto problémy lze vyřešit pomocí implementace Kalmanova filtru. Kalmanův filtr je nástroj, který slouží pro odhad vnitřního stavu systému na základě vstupů a výstupů do systému. Tento nástroj nám tedy umožní predikovat aktuální pozici nedetekovatelného objektu, který se aktuálně nachází v obraze, a tím nám umožní predikovat trajektorie pohybu.



Obrázek 2: Vybrané objekty z trajektorie pohybu objektu ve zmenšené velikosti.

Vypočtené trajektorie pohybu dle předchozího postupu (viz. Obrázek 2.) musí být podrobeny analýze. Prvně musí být provedena verifikace jednotlivých trajektorií, aby v případě, že trajektorie nepatří člověku, byla trajektorie vyřazena z dalšího zpracování. Pro tento účel byly vybrány binární klasifikační modely strojového učení a vybrané známé detektory osob v obraze.

Princip verifikace je takový, že je vybráno náhodně deset snímků z každé trajektorie pohybu a poté je těchto deset snímků klasifikováno dle metod strojového učení nebo detekováno dle vybraných detektorů osob. Pokud jsou správně detekovány či klasifikovány alespoň tři snímky v trajektorii, tak je trajektorie považována za verifikovanou. Pokud tato podmínka není splněna, tak je trajektorie vyřazena z dalšího zpracování. Tento proces je proveden pro každou trajektorii.

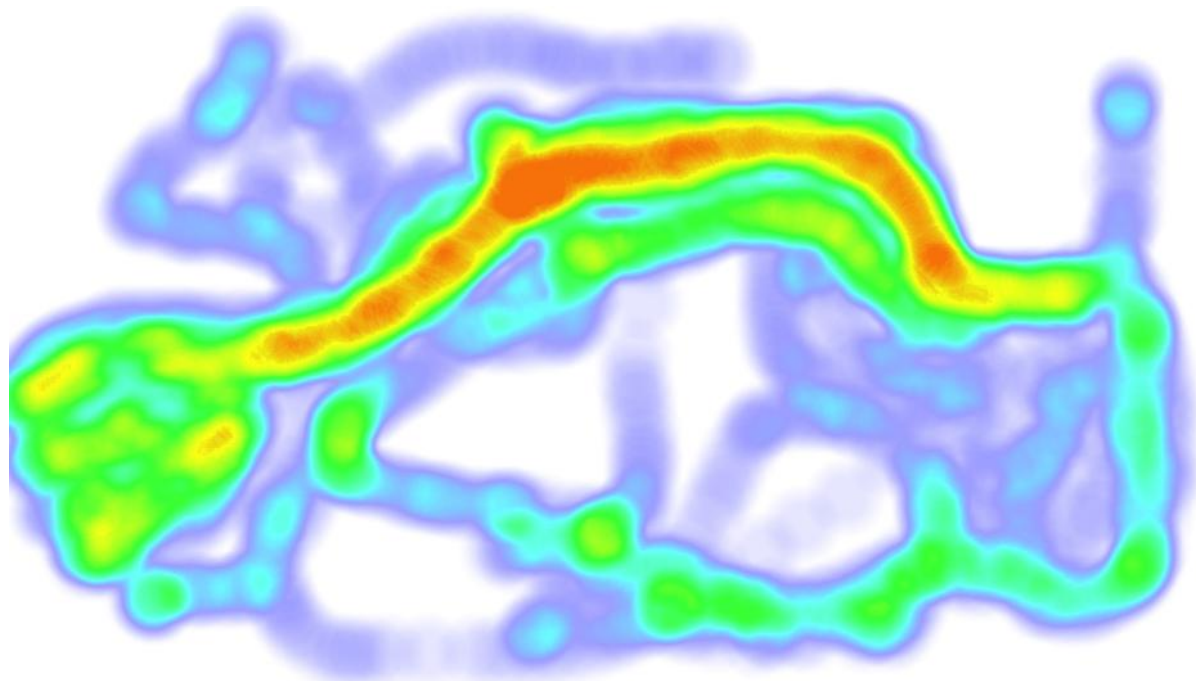


Obrázek 3: Zobrazení jedné trajektorie pohybu osoby, která prošla verifikací a je označena jako aktivní trajektorie.

Po procesu verifikace jsou jednotlivé trajektorie pohybu analyzovány podle svého směru a tvaru. Před začátkem detekce musí být stanoveny vstupně-výstupní oblasti, které určují oblasti, které umožňují odchod, či příchod do oblasti viz. *Obrázek 3*.

Pokud těžiště počátečního objektu trajektorie patří do vstupně-výstupní oblasti a těžiště posledního objektu v trajektorii patří do vstupně-výstupní oblasti, tak je trajektorie vyřazena z dalšího zpracování a je započten příchod či odchod do oblasti v závislosti na typu vstupně-výstupní oblasti. Pokud tato podmínka není splněna, tak je trajektorie označena jako aktivní.

Výslednou množinu trajektorií pohybu osob lze zobrazit pomocí zajímavých nástrojů, které umožňují poskytnout další pohled na získaná data. Jedná se zejména o zobrazení pomocí teplotních map viz. *Obrázek 4*, vybraných grafů či vyjádření pomocí statistických charakteristik. Výslednou množinu dat lze také omezit na určitý časový interval. Z těchto výsledků lze například vyčíst, v jaké části obrazu byl největší výskyt osob, kdy byl největší příchod lidí do oblasti a další.



Obrázek 4: Ukázka teplotní mapy, která byla vygenerována ze získaných trajektorií pohybu.

Dále je možné takto získaná data uložit pomocí nerelačního databázového systému MongoDB či do XML nebo JSON formátu. Zpracované data pak lze zkoumat v rámci širšího časového intervalu a tím získat věrohodnější údaje nebo je také možné napojit tyto data na vhodný analytický nástroj.

### 3. Vlastní práce

#### 3.1 Pořízení testovacích dat

V rámci práce byly vytvořeny dvě množiny testovacích dat. První množina obsahuje videosoubory, které byly pořízeny kamerami od výrobce Basler viz *Tabulka 1*. Pořízené videosoubory zachycují simulovaný a reálný pohyb osob. Videosoubory slouží zejména pro vytvoření datových sad, testování algoritmů a také pro zobrazení vhodných výstupů.

ID	Délka videa	Počet snímků	Rozlišení	FPS	Kamera
1	01:13:18	43980	2040x1086	10	Basler acA2000-165uc
2	00:43:48	52560	1624x1234	20	Basler acA1600-20uc

Tabulka 1: Vlastnosti videosouborů, které byly pořízeny kamerami od výrobce Basler.

Druhá množina dat je vytvořena z pořízených videosouborů a slouží k natrénování, optimalizaci a testování vybraných metod strojového učení a také pro testování vybraných detektorů osob viz *Obrázek 5*.



Obrázek 5: Vybrané pozitivní snímky z datové sady v rozlišení 80x60 pixelů zobrazené v odstínech šedi.

Tyto množiny dat jsou složeny ze snímků, které musí mít převedeny do odstínu šedi a musí mít stejnou velikost. Poté jednotlivé snímky ručně klasifikovány podle jejich obsahu. Pro každou datovou sadu byly vytvořeny snímky ve třech variantách viz. *Tabulka 2*.

Název	Velikost snímků [px]	Počet snímků	Negativní/pozitivní
Dataset č. 1	80x60	2288	851/1437
	55x40	2288	851/1437
	30x20	2288	851/1437
Dataset č. 2	80x60	2652	1058/1594
	55x40	2652	1058/1594
	30x20	2652	1058/1594
Dataset č. 3	80x60	3216	1219/1997
	55x40	3216	1219/1997
	30x20	3216	1219/1997

Tabulka 2: Popis vlastností datových sad, které budou využívány pro metodami strojového učení a detektory osob.

### 3.2 Komparace algoritmů optického toku

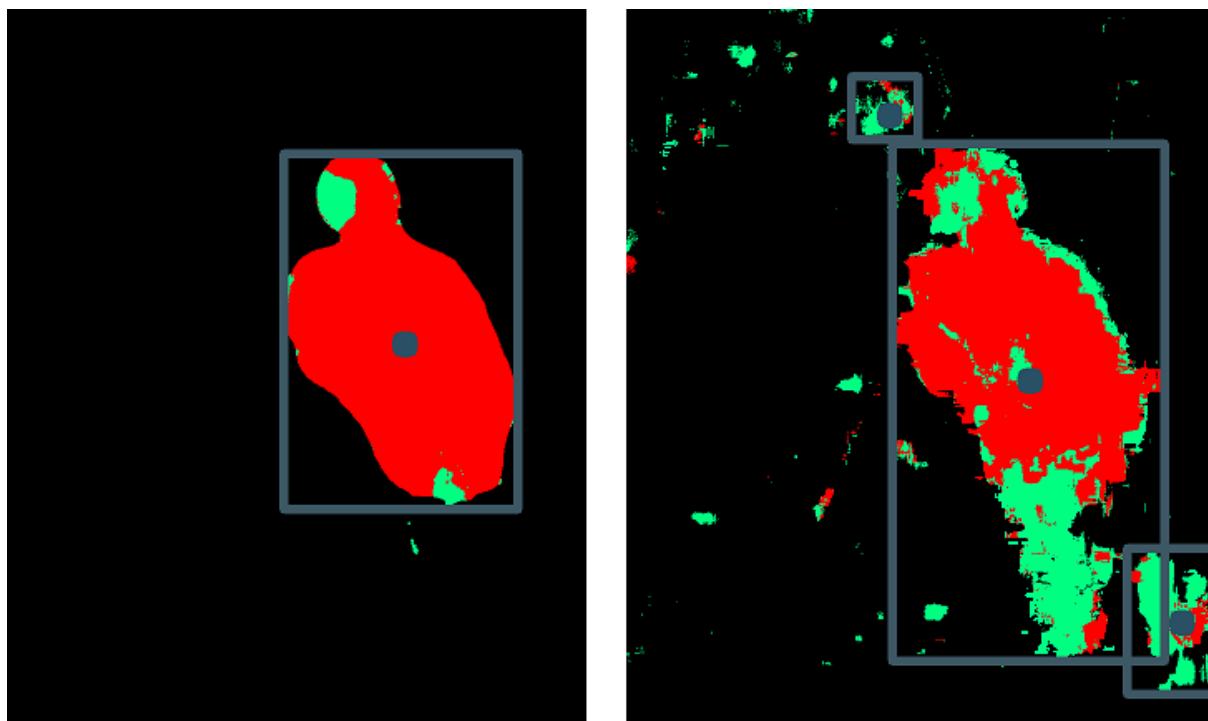
V rámci práce byla provedena komparace vybraných algoritmů pro výpočet optického toku na základě časové náročnosti a přesnosti algoritmu. Pro každý algoritmus byl proveden test, který určil střední hodnotu a hodnotu MCE na základě zpracování tisíce snímků daným algoritmem. Střední hodnota časové náročnosti a metrika MCE jsou uvedeny v tabulkách viz *Tabulka 3*, *Tabulka 4*.

Rozlišení	TV-L1	Lucas-Kanade	Gunnar-Farneback	Brox
1920×1080	18,14 ms	0,982 ms	1,972 ms	1,128 ms
1280×720	7,543 ms	0,542 ms	1,130 ms	0,789 ms
640x480	2,672 ms	0,282 ms	0,641 ms	0,378 ms
480x360	1,320 ms	0,189 ms	0,542 ms	0,292 ms

Tabulka 3: Porovnání časové náročnosti algoritmů optického toku na jeden snímek pomocí CUDA technologie na chipu GeForce GTX 1060.

Rozlišení	TV-L1	Lucas-Kanade	Gunnar-Farneback	Brox
1920×1080	3%	21%	8%	5%
1280×720	6%	22%	12%	10%
640x480	7%	24%	13%	14%
480x360	9%	28%	15%	17%

Tabulka 4: Porovnání přesnosti algoritmů optického toku na základě metriky MCE pomocí CUDA technologie na čipu GeForce GTX 1060.



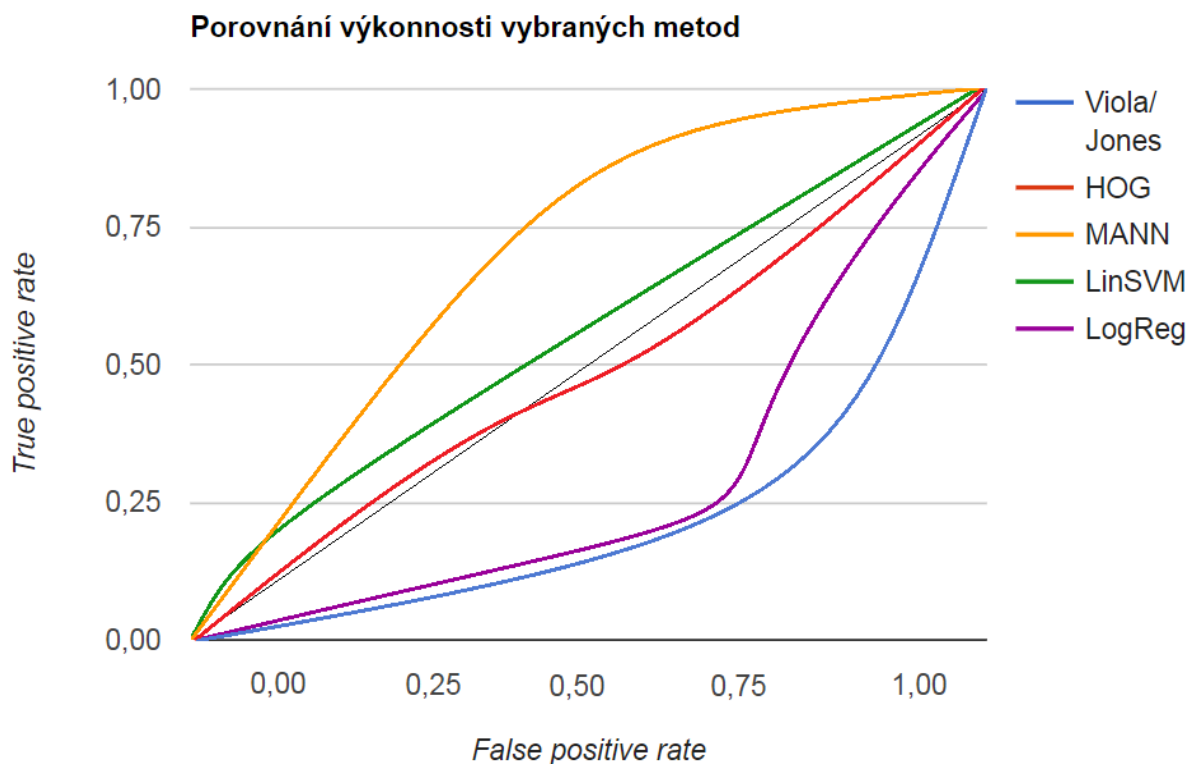
Obrázek 7: Levý obrázek ukazuje detekovaný pohyb osoby pomocí algoritmu TV-L1. V pravém snímku je zachycen pohyb pomocí algoritmu Lucas-Kanade.

Nejrychlejším algoritmus v testu je Lucas-Kanade, který je odhadem desetkrát rychlejší než nejpomalejší algoritmus TV-L1. Za to algoritmus Lucas-Kanade dosáhl největší nepřesnosti v detekci s hodnotou MCE 21% při rozlišení 1920x1080, zatímco algoritmus TV-L1 byl výrazně přesnější s hodnotou MCE 3% při stejném rozlišení viz. *Obrázek 7*. Algoritmus TV-L1 byl vybrán pro detekci dynamických objektů na základě toho, že přesnost algoritmu je v našem případě hlavní prioritou.

### 3.3 Komparace metod pro verifikaci osob

V rámci komparace byly testovány vybrané klasifikační modely lineární SVM, Logistická regrese a vícevrstvá neuronová síť se 3 třemi vrstvami, která využívá 20 skrytých neuronů. Jako učící algoritmus byla použita metoda zpětné propagace Levenberg-Marquardt. Ze známých detektorů osob byly testovány detektory Viola/Jones[1] a HOG detektorem [4].





Obrázek 6: ROC křivky zobrazující výkonnost jednotlivých metod, které byly naučeny, optimalizovány a testovány na datové sadě č. 3 s rozlišením snímků 80x60.

Výsledek testování stanovil matici záměn a metriky pokrytí a přesnost pro každou metodu viz *Tabulka 5*. Z těchto informací byla následně sestavena ROC křivka pro každou metodu v závislosti na vstupních parametrech algoritmu.

<b>Detektory</b>	<b>TP</b>	<b>FN</b>	<b>TN</b>	<b>FP</b>	<b>Pokrytí (Recall)</b>	<b>Přesnost (Precision)</b>
Detektor Viola/Jones	57	243	36	147	0,19	0,28
HOG+SVM-detektor	156	27	98	85	0,52	0,65
<b>Klasifikační modely</b>	<b>TP</b>	<b>FN</b>	<b>TN</b>	<b>FP</b>	<b>Pokrytí (Recall)</b>	<b>Přesnost (Precision)</b>
Lineární SVM	183	117	109	74	0,61	0,71
Vícevrstvá neuronová síť	249	51	152	31	0,83	0,89
Logistická regrese	63	237	42	141	0,21	0,31

Tabulka 5: Vybrané hodnoty metrik, které byly získány testováním výše uvedených metod na datové sadě č. 3 s rozlišením snímků 80x60 v testovací fázi.

Z výsledků testování je patrné, že nejlepší schopnost klasifikace poskytuje neuronová síť, která byla na základě tohoto testování implementována do řídicí aplikace. Detektor Viola/Jones dosáhl v testu nejhorších výsledků, jelikož tato metoda je již zastaralá.

Pokud je prováděna detekce a analýza trajektorií pohybu nad videosouborem, který obsahuje pouze lidi, tak není nutné provádět verifikaci trajektorií. Dále by bylo možné algoritmus rozšířit o další verifikaci např. aut v obraze, ale musely by být pořízeny vhodné datové sady snímků aut.

#### 4. Vyhodnocení přesnosti algoritmu

Přesnost algoritmu byla stanovena na základě porovnání reálných a detekovaných příchodů či odchodů z oblasti zachycené kamerou. Tyto informace byly získány detekcí a analýzou trajektorií pohybu osob z pořízených videosouborů viz *Tabulka 6*.

ID	Délka videa	Reálné příchody/Detekované příchody	Reálné odchody/Detekované odchody	Přesnost	Čas zpracování
1	01:13:18	52/43	47/39	82,69%/82,97%	4:35:12
2	00:43:48	39/33	38/31	84,61%/81,57%	3:21:46

Tabulka 6: Popis získaných informací, které byly získány detekcí a analýzováním trajektorií pohybu osob z pořízených videosouborů.

Celková úspěšnost algoritmu je vypočítána jako aritmetický průměr vypočtené přesnosti odchodů a příchodů. Výsledná hodnota byla stanovena na 82,96%.

#### 5. Závěr

V rámci práce byly pořízeny dva videosoubory, zachycují reálný a simulovaný pohyb osob viz *Tabulka 1*. Z těchto videosouborů byly vytvořeny tři ohodnocené datové sady, obsahující tři různé rozlišení snímků viz *Tabulka 2*.

Poté byla provedena komparace vybraných algoritmů pro výpočet optického toku na základě pořízených videosouborů. Následně byl vybrán algoritmus TV-L1 kvůli vysoké přesnosti 3% MCE v rozlišení 1920x1080 pro detekci dynamických objektů viz *Tabulka 3*, *Tabulka 4*.

Poté byla provedena komparace výkonnosti vybraných klasifikačních modelů a známých detektorů osob. Vícevrstvá neuronová síť dosahovala v komparaci nejpresnější klasifikace, a proto byla vybrána tato síť pro verifikaci nalezených trajektorií viz *Tabulka 5*.

Následně byla vytvořena řídicí aplikace pomocí frameworku.NET, knihovny EmguCV a frameworku Accord.NET. Aplikace implementuje funkcionalitu pro detekci a analýzu trajektorií pohybu osob v obraze a také obsahuje vhodné nástroje pro zobrazení a uložení výstupních dat. Celkový výsledek přesnosti algoritmu je stanoven na hodnotu 82,96% viz *Tabulka 6*. Použití algoritmu je vhodné zejména v oblastech, kde není vyžadována velká přesnost detekce, ale je spíše důležitý alespoň částečný odhad. Jedná se například o nákupní centra, kde algoritmus může poskytnout manažerům zajímavé informace o pohybu zákazníků kolem vybraného zboží apod.

#### 6. Literatura

- [1] Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, díl 1, str. I511-I518, 8.12 2001, Kauai, Spojené Státy americké.
- [2] Brox, T., Bruhn, A., Papenber, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping, díl 3024, str. 25-36, 2004, Heidelberg, Německo.
- [3] Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime TV-L1 optical flow. 29th Annual Symposium of the Deutsche Arbeitsgemeinschaft für Mustererkennung, díl 4713, str. 214-223, 14.11 2007, Heidelberg, Německo.
- [4] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition, díl 1, str. 86-893, 25.6 2005, San Diego, Spojené Státy americké.