Projekt č. VI20172020068

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

Technická zpráva 2018

Detekce a identifikace hledaných osob ve videu

David Bažout, Vítězslav Beran

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

Prosinec 2018

Obsah

1 Úvod **1**

2 Výběr aktuálních metod **2**

3 Datové sady, experimenty a výsledky **4**

4 Uživatelská aplikace **8**

Abstrakt

Technická zpráva prezentuje výpočetní modul a základní uživatelskou aplikaci pro efektivní vyhledávání osob ve video sekvenci. Hledané osoby jsou systému zadány pomocí několika fotografií jejich obličeje. Modul umožňuje analyzovat video v online (zdroj např. IP kamera) nebo offline (ze souboru) režimu. Při detekci a identifikaci osoby ve snímcích videa systém hlásí výskyt hledané osoby. Obsahem zprávy je představení běžně využívaných metod pro detekci a popis obličeje, výběr vhodné datové sady pro vývoj a vyhodnocení modulu. Součástí je i popis samostatné uživatelské aplikace a způsob jejího využití.

1 Úvod

Jednou z potřebných úloh pro sledování budov a objektů technologie umožňující efektivní vyhledávání osob podle obličeje ve videu. Hledané osoby jsou systému zadávány pomocí několika fotografií jejich obličeje. Video může pocházet z IP kamery v reálném čase, nebo může být načítáno ze souboru. Systém je schopen určit, na kterých snímcích videa a ve které části se hledaná osoba nachází. Tento problém lze rozdělit na detekci obličeje a jeho následné rozpoznávání.

Úloha detekce má za úkol určit, kde se ve fotografii obličej nachází. Detekce obličeje je zapouzdřena do logického celku, který je označován jako detektor. Existuje množství detektorů, které s různou úspěšností a výpočetními nároky dokáží vyřešit tuto úlohu. Mohou být založeny na klasických přístupech provádějící převod snímku do odstínů šedi a následné aplikaci různých algoritmů pracujících s maticí hodnot intenzit šedotónového obrázku. V současné době jsou ale nejlepší výsledky v oblasti detekce obličeje dosahovány detektory, které jsou založeny na konvolučních neuronových sítích.

Cílem úlohy rozpoznávání je v prvním kroku provést extrakci příznakového vektoru obličeje ze snímku videa a porovnat ho s příznaky obličejů uložených v databázi. Porovnání příznakových vektorů spočívá ve výpočtu euklidovské vzdálenosti. Experimenty na vhodné datové sadě lze určit optimální práh euklidovské vzdálenosti pro shodu mezi dvěma příznakovými vektory. Výsledkem úlohy je určení identifikátoru osoby na obrázku.

2 Výběr aktuálních metod

Mezi nejúspěšnější technologie pro detekci obličeje ve fotografii patří klasické metody Viola-Jones, HOG a LBP a metody založené na konvolučních neuronových sítích.

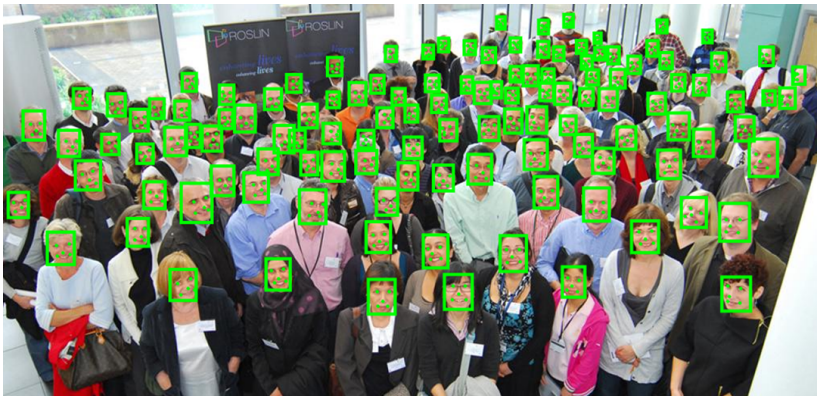
Metoda Viola-Jones [1] spadá do kategorie metod detekce na základě extrahovaných příznaků. V tomto případě je využíváno příznaků typu Haar. Metoda rovněž využívá pro optimalizaci rychlosti výpočtů strukturu integral image a upravený algoritmus strojového učení AdaBoost. Interně pracuje pouze s daty v odstínech šedi. Tím je dosaženo velice rychlého výpočtu a možnosti zpracování obrazových dat v reálném čase. V současné době je pravděpodobně nejrozšířenější metodou při řešení běžných úloh jako například detekce obličeje implementovaná ve fotoaparátu.

Metoda HOG (Histogram of oriented gradients) [2] patří rovněž mezi metody, které při detekci obličeje využívají extrakce příznaků. Metoda získává hodnoty a směry gradientu pro jednotlivé pixely aplikací konvolučního filtru. Výsledný příznak odráží rozložení jednotlivých vektorů gradientu v obrázku. Tím je dosaženo lepších výsledků za nepříznivých světelných podmínek. Na extrahovaný příznakový vektor je aplikován klasifikátor určující výskyt obličeje. Ke klasifikaci objektů se v praxi využívá Support Vector Machine [3].

LBP detektor je dalším z detektorů využívající extrakce příznakového vektoru [3]. Metoda rozděluje pixely na menší buňky ve tvaru čtverce s rozměry 3x3 pixely. Porovnáváním šedotónové hodnoty středového pixelu buňky s pixely okolními získává 8-bitové čísla reprezentující jednotlivé buňky. Příznak představuje vektor s 256 dimenzemi reprezentující histogram rozložení hodnot získaných předchozí operací. Výskyt obličeje je opět určen pomocí natrénovaného klasifikátoru.

V současné době jsou dosahovány nejlepší výsledky v oblasti detekce obličeje aplikací neuronových sítí a existuje množství detektorů obličejů na nich založených. Jedním z rozšířených metod je detektor publikovaný v práci Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks [4]. Tento detektor nejdříve určí ve fotografii oblast s vysokou pravděpodobností výskytu obličeje pomocí neuronové sítě s jednodušší strukturou. Postupnou aplikací složitějších neuronových sítí pouze na některé oblasti fotografie je dosažena významná úspora výpočetního výkonu a tím i zpracování v reálném čase za vysoké kvality detekce. Ukázka detekce pomocí této metody je na obrázku 1. Výsledné řešení bude založeno na využití tohoto detektoru.

V úloze rozpoznávání obličeje je k extrakci příznakového vektoru využíváno speciálně navržených konvolučních neuronových sítí. Trénovací algoritmus těchto sítí se snaží minimalizovat vzdálenost příznakových vektorů mezi fotografiemi stejné osoby a maximalizovat vzdálenost mezi fotografiemi, na kterých se nachází stejná osoba. V anglické literatuře se tento mechanizmus



Obrázek 1: Ukázka detekce obličeje [4].

označuje jako *Triplet loss*. Vyvíjený modul využívá upravené neuronové sítě typu RESNET-34 [5]. Tato síť byla trénována na kombinaci datových sad *The face scrub dataset*[1] a *VGG dataset*[2] s celkovým počtem okolo 3 milionů fotografií tváří s celkovým počtem individuálních osob přesahujícím 7 tisíc.

[1] `http://vintage.winklerbros.net/facescrub.html`
[2] `http://www.robots.ox.ac.uk/~vgg/data/vgg_face/`

3 Datové sady, experimenty a výsledky

Datových sad určených pro úlohu detekce nebo identifikace tváří byla vytvořena celá řada. Mezi nejznámější patří datová sada *Labeled Faces in the Wild* [6]. Tato datová sada sestává z 13000 anotovaných fotografií z internetových zdrojů. Celkově se na ní vyskytuje 1680 různých osob. Fotografie této datové sady jsou v příliš dobrém rozlišení a kvalitě oproti datům, pro které je aplikace určena.

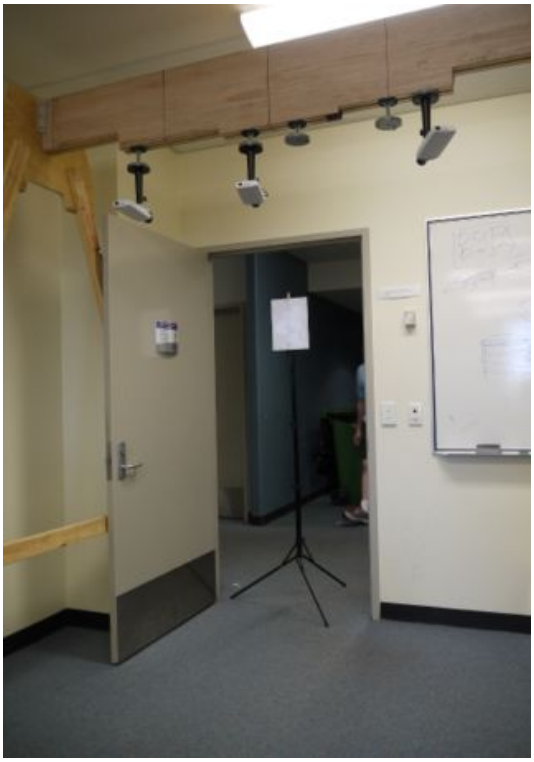
Face Detection Data Set and Benchmark Home (FDDB) [7] je další ze známých datových sad určených pro úlohu detekce tváře. Zajímavostí této datové sady je odlišnost v anotaci plochy, na které se nachází obličej. Tato plocha je určena pomocí elipsy namísto čtvercového ohraničení. To umožňuje provádět přesnější analýzu a porovnávání detektorů. Datová sada sestává z 2845 fotografií s 5171 anotovanými tvářemi.

Pro testování vyvíjeného modulu bylo použito datové sady *ChokePoint* [8]. Tato datová sada je tvořena 48-mi sekvencemi videozáznamů nahrávaných 3-mi kamerami umístěnými nad dveřmi v odlišných úhlech (obrázek 2). Celkový počet snímků sekvence je 100 000 a na 64 000 snímcích se vyskytuje obličej. V datové sadě se vyskytuje 29 různých osob. Osoby přicházejí ke dveřím z různých úhlů a procházejí směrem dovnitř. Data se podobají záznamům z interního kamerového systému. Datová sada je určena pro úlohu identifikace. Její anotace lze využít s menší přesností i k testování úlohy detekce. Anotace nese ke každé osobě souřadnice levého a pravého oka a její identifikátor. Příklady z videozáznamů datové sady ChokePoint jsou na obrázku 3.

Testování detektorů obličeje spočívalo v porovnání různých detektorů, jejich různých nastavení a výpočetní náročnosti. Výstup testování představují precision-recall křivky (obrázek 4) umožňující přehledné grafické porovnání kvality detekce a sloupcový graf (obrázek 5) porovnávající výpočetní náročnost na referenčním stroji s procesorem Intel-i5@1.8GHz. Na základě těchto informací bylo možné určit nejvhodnější konfiguraci pro danou úlohu.

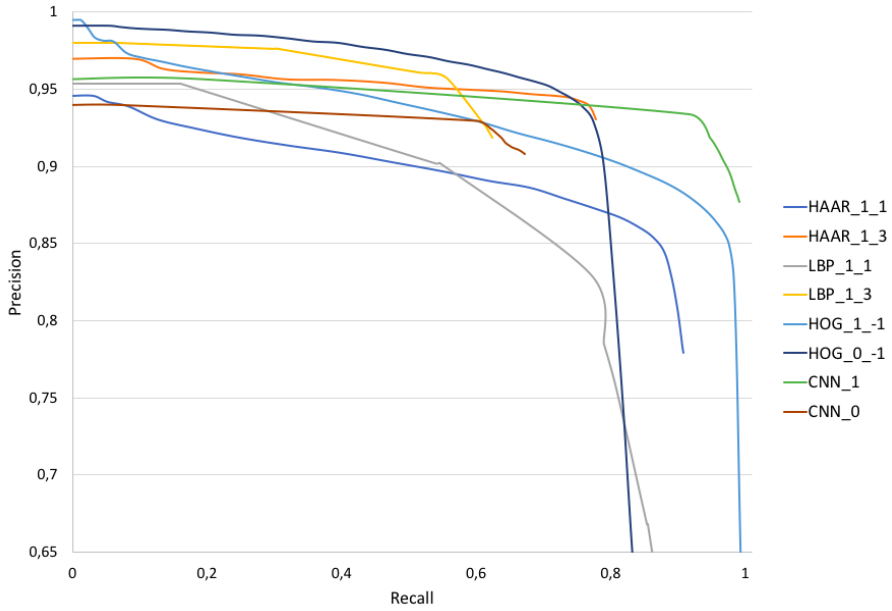
Legenda na obrázku 6 přiřazuje křivkám na obrázcích 4 a 5 použitou metodu a její nastavení. Parametr `minNeighbors` určuje minimální počet detekcí objektu na stejné pozici ve scale pyramid definované parametrem `scaleFactor` takový, aby byla detekce považována za úspěšnou. Parametr `minSize` určuje minimální rozměry prohledávacího okna. U metod detekce HOG a CNN implementovaných v knihovně dlib je možnost specifikovat `upsample factor` pro vstupní obrázek.

Dalším cílem bylo otestovat úspěšnost rozpoznávání tváře pomocí upraveného modelu neuronové sítě RESNET-34. Jako metrika byla použita hodnota *mAP* (*mean average precision*). Hodnota *mAP* pro rozpoznávání obličeje na datové sadě ChokePoint s použitím modelu RESNET-34 je 97,42% [5].

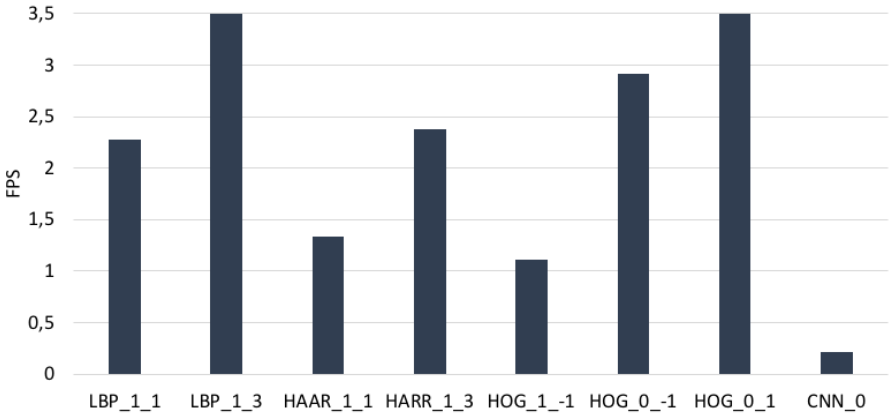


Obrázek 2: Umístění kamer při snímání datové sady ChokePoint.

Obrázek 3: Příklad obrázků z datové sady ChokePoint.



Obrázek 4: Precision - recall křivky pro různé detektory.



Obrázek 5: Výkonnostní porovnání různých metod detekce.

HAAR_1_1	minNeighbors=5 minSize=(10, 10) scaleFactor=1.1 nebo 1.3
HAAR_1_3	
LBP_1_1	minNeighbors=5 minSize=(10, 10) scaleFactor=1.1 nebo 1.3
LBP_1_3	
HOG_1_-1	upsampleFactor = 1 nebo 0
HOG_0_-1	
CNN_1	upsampleFactor = 1 nebo 0
CNN_0	

Obrázek 6: Legenda ke grafům porovnávajícím metody detekce.

4 Uživatelská aplikace

Uživatelská aplikace je implementována v interpretovaném jazyce Python verze 3. Jedná se o terminálovou aplikaci, které je pomocí parametrů příkazové řádky předávána adresa IP kamery, složka s databází fotografií hledaných osob a složka pro uložení výstupních dat. Výstupem je video soubor a textový soubor ve formátu *csv*, kdy každý řádek nese informace o jedné detekované osobě:

- číslo snímku,
- identifikátor osoby,
- souřadnice detekce (souřadnice x, souřadnice y, šířka, výška) a
- příznakový vektor dané osoby.

Video soubor pak nese obrazová data získaná z kamery s vizualizací detekcí osob pomocí ohraničujícího obdélníku s popisujícím identifikátorem. Časová značka počátku nahrávání dat a IP adresa kamery je zakódována do názvu souboru.

K načítání a ukládání obrazových dat ze zdroje je využíván objekt videoCapture z knihovny OpenCV. Dále je využito implementace detektoru HOG pro detekci obličeje a modelu neuronové sítě RESNET-34 pro extrakci příznakových vektorů za účelem identifikace z knihovny dlib [5]. K porovnávání příznakových vektorů je využita funkce knihovny Scipy. Jako abstrakci pro práci s detekovanými osobami je využita vlastní třída *Face* s atributy pro uložení identifikačního řetězce, příznakového vektoru a objektu *trackeru*. Implementace demo aplikace využívá korelačního *trackeru* z knihovny Dlib pro optimalizaci výpočtu.

Jádro aplikace bude integrováno jako výpočetní modul do výsledného systému. Aktuálně probíhají experimenty ještě s jedním, nově vyvinutým detektorem, který by měl mít ještě menší výpočetní nároky a lepší detekční vlastnosti.

Literatura

[1] M. Šonka, *Image processing, analysis, and machine vision.* Toronto: Thomson, 3rd ed. ed., 2008.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *In CVPR*, pp. 886–893, 2005.

[3] G. R. Bradski, *Learning OpenCV.* Computer Programming. Robotics, Sebastopol: O'Reilly, 2008.

[4] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, pp. 1499–1503, Oct 2016.

[5] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[6] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

[7] V. Jain and E. Learned-Miller, "Fddb: A benchmark for face detection in unconstrained settings," Tech. Rep. UM-CS-2010-009, University of Massachusetts, Amherst, 2010.

[8] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 81–88, IEEE, June 2011.