

# Detektor obličejů ve videu pro dohledové systémy

Technical Report - FIT -  
VG20102015006 - 2012 - 05

Ing. Aleš Láník



Faculty of Information Technology, Brno University of Technology

20. prosince 2012

# Obsah

Obsah.....	2
1.Úvod.....	1
2.Detekce obličejů.....	2
2.1.Předzpracování obrazu.....	3
2.2.Segmentace.....	5
2.3.Extrakce příznaků.....	5
2.4.Klasifikace.....	7
2.5.Trasování obličejů v čase.....	9
3.Návrh začlenění do vyvíjeného systému.....	11
4.Závěr.....	12
5.Literatura.....	13

# Detektor obličejů ve videu pro dohledové systémy

Ing. Aleš Láník

20. prosince 2012

## Abstrakt

Technický report se zabývá problematikou detekce lidských obličejů v obraze a jejich následné trasování v časové doméně. V reportu jsou rozebírány jednotlivé části detekcí a jejich typické algoritmy. V poslední části reportu je rozebírám způsob jakým bude popisován systém začleněn do funkčního prototypu.

## 1. Úvod

Lidský obličej je jedena z nejvíce výrazných částí lidského těla, obsahuje z hlediska vizuálních vlastností, mnoho kontrastních podčástí, které jsou však vůči sobě vždy stejně umístěny a orientovány. Příkladem takových primitiv jsou např. oči, nos a rty, které má téměř každý člověk umístěné vždy stejně. Kombinace těchto primitiv je natolik jedinečná, že dostatečně vizuálně odděluje lidský obličej od pozadí.

Z těchto důvodů je právě detekce obličeje velmi vhodná, při dodržení vhodných podmínek snímání, pro detekci osob např. pro dohledové systémy.

Existují různé přístupy, jak lidský obličej v obraze detekovat, v této práci se však budeme zabývat pouze jedním v současnosti nejvíce používaným způsobem a to statistickým klasifikátorem ADABOOST.

Statistické klasifikátory jsou hojně užívány pro detekci objektů či detekci vzorů v rastrových obrazech. Jejich základním principem je, že odhadují pravděpodobnost výskytu objektů dané třídy na základě předem známých dat, které se použijí k natrénování klasifikátoru a na základě stávajících dat. Z těchto dvou skupin informací se dá odhadnout, zda se v aktuální chvíli daný objekt v obraze nachází či ne.

Vstupními daty do klasifikátorů typicky nebývají přímo pixely daného obrazu ale vektory příznaku z těchto pixelů vypočítané. Je to především z důvodu snížení dimensionalitu prostoru příznaků a tím snížení výpočetní složitosti, dále k odstranění vysoké korelace

pixelů, která by vedla k neseparovatelnosti dané třídy v prostoru příznaků. Dalším důvodem je zvýšení invariantnosti detektoru pro různé typy geometrických deformací (rotace, změna měřítka, ...) a v neposlední řadě také ke snížení vlivu okolního osvětlení.

Nedá se obecně říci, že existuje nejlepší algoritmus pro extrakci příznaků, který by byl využitelný pro libovolný třídu objektů a proto většinou vznikají specifické extraktory příznaků pro specializované případy. V našem případě se však budeme zabývat extraktory příznaků vhodné pro detekci obličejů v obraze.

Obecně lze říci, že výhoda použití statistických klasifikátorů spočívá v automatickém naučení z předem známých (trénovacích) dat. Jednou z nejdůležitějších podmínek, která však musí být splněna je však dostatečná odlišitelnost hledaných objektů od zbytku obrazu.

## 2. Detekce obličejů

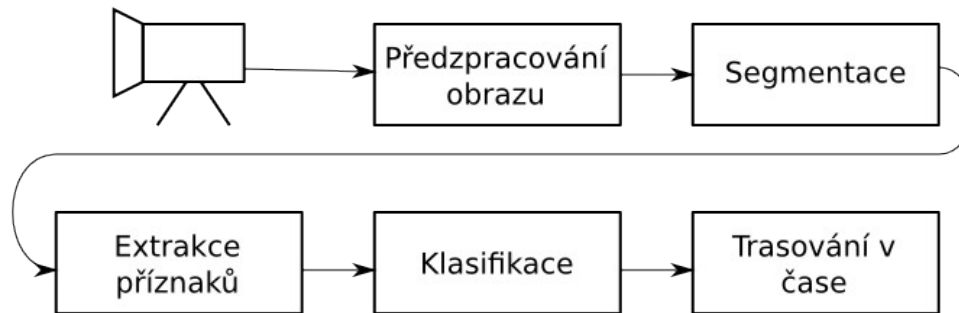
Pro potřeby dohledových systémů, a systémů pro bezpečnosti složky, vznikl požadavek pro automatickou detekci části videí z dohledových kamer, na kterých je přítomna lidská osoba. Tato funkcionality by výrazně zjednodušila práci, která se v současnosti provádí ručně a je tak velmi časově náročná.

Jako vhodným kandidátem pro detekci stavu, kdy je v obraze přítomna lidská osoba, se jeví právě detektor obličejů, u kterého lze říci, že funguje s velkou mírou spolehlivosti.

Tato skutečnost vychází, jak již bylo zmíněno, z velmi jedinečných vlastností rozmístění vizuálních primitiv (oči, nos, rty, ...) lidského obličeje, jejich stálosti u lidské populace a jedinečnosti takového vzoru na reálně pořízených datech.

Obecně lze detekci obličejů a tedy i osob rozdělit do základních bloků, které jsou popsány v obrázku 1. Vstupem celého systému je snímač dat, v našem případě kamera, která dodává rastrové data (převážně v RGB barevném modelu).

Tato rastrová data projdou blokem, který se stará o předzpracování dat, který ideálně připraví data, tak aby další bloky měly usnadněnou práci.



Obrázek 1: Schéma detektoru osob ve videu

V následujícím kroku se data segmentují do menších obrazových oblastí (většinou čtvercového tvaru), které pokrývají celý vstupní obraz, tak aby mezi sousedními bloky byl určitý překryv. Nad jednotlivými oblastmi se spočítají příznakové vektory popisující obsah oblasti. Tyto vektory jsou pak vstupem samotného klasifikátoru, který určí pravděpodobnost zda daná oblast obsahuje hledaný objekt (v našem případě obličeje).

Všechny tyto kroky jsou prováděny snímek po snímku a vzniká tak sekvence výsledků klasifikátoru. O sjednocení a přiřazení jednotlivých detekcí klasifikátoru skutečným objektům se stará poslední blok.

## 2.1. Předzpracování obrazu

Vstupní obraz je vhodné převést do grayscale barevného modelu. Informace o barvě se na první pohled zdá užitečná, protože lidská kůže má pevně daná specifika (především v odstínu kůže)[1], ovšem pro tento druh použití je zcela nevhodná z hlediska nemožnosti nastavení vhodných podmínek snímání takových, aby informace o odstínu kůže byla užitečná. Tímto krokem se také eliminuje datová náročnost dalšího zpracování.

Vstupní snímky jsou velmi často zatíženy různým typem šumu, jenž může negativně ovlivnit výsledky dalšího zpracování. Tento šum má velmi často předem známou charakteristiku a není tedy velkým problémem jeho nepříznivé dopady eliminovat.

Dalším typickým krokem v předzpracování obrazu je redukce geometrických zkreslení snímací soustavy. Tento krok je velmi důležitý vzhledem k prostorově pevné závislosti jednotlivých obličejových elementů, které by díky zkreslení mohly mít pozměněný charakter.

Většinou bývá tento krok prováděn poloautomaticky, kdy prvním krokem je snímání kalibračního obrazce známých vlastností např. šachovnice a následném výpočtu kalibrační matice. Dá se

předpokládat, že míra a způsob zkreslení bude po celý čas životnosti snímacího senzoru nezměněna proto stačí tuto kalibraci provést jen jednou za životní cyklus snímače.

Vzhledem k tomu, že extraktory příznaků používané pro tuto doménu jsou často založené na sčítání hodnot nad obdélníkovými oblastmi obrazu [3], se často používá optimalizační technika zvaná integrální obraz. Pomocí této techniky jsme schopni sečíst libovolně velkou oblast vždy v konstantním čase.

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

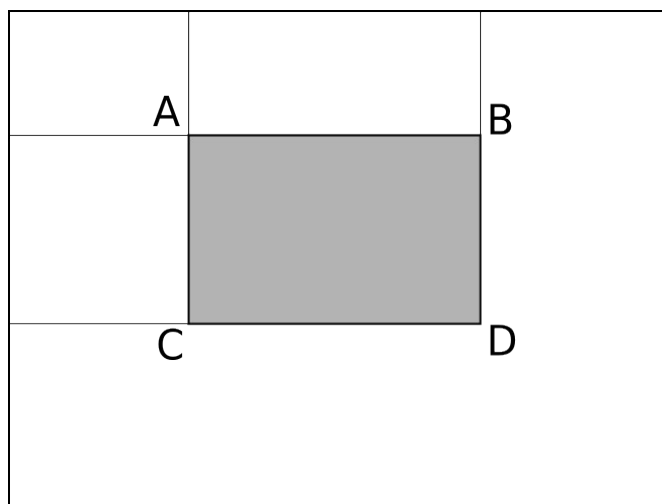
*Vzorec 1: Výpočet integrálního obrazu*

Integrální obraz je transformace původního obrazu  $i$  do stejně velkého obrazu  $ii$ , ve kterém souřadnice  $x, y$  obsahuje součet oblasti  $v$  v  $i$  od toho bodu do středu souřadného systému.

$$suma(A, B, C, D) = ii(x_D, y_D) + 2 * ii(x_A, y_A) - ii(x_B, y_B) - ii(x_C, y_C)$$

*Vzorec 2: Výpočet obdélníku s rohovými body A, B, C, D*

Po provedení tohoto výpočtu, právě v bloku předzpracování obrazu, lze poté spočítat součet hodnot nad jakoukoliv obdélníkovou oblastí jako součet pěti hodnot z integrálního obrazu.



*Obrázek 2: Výpočet součtu hodnot pixelů nad obdélníkem A, B, C, D*

## 2.2.Segmentace

Vzhledem k tomu, že vstupní data jsou velmi často velmi rozsáhlého charakteru, je vhodné tato rastrová data dělit do menších bloků, které se pak zpracovávají samostatně. Toto dělení má mimo jiné za následek to, že jsme schopni lokalizovat hledaný obličej na místě v obraze, což je pro detekci obličejů důležitá skutečnost především pro následující blok ve zpracování a to trasování obličejů v čase viz kapitola 2.5.

Dalším pozitivním důsledkem tohoto dělení je fakt, že takto rozdělený obraz lze zpracovávat paralelně např. na více jádrových procesorech, GPU [2,4] či specializovaných FPGA chipsetech [4] čímž se dosáhne akcelerace výpočtu. Toto je výhodné obzvláště pro specializovaný grafický hardware, který obsahuje velké množství výpočetních jader.

Pro účely dále popsaného klasifikátoru se segmentuje obraz po blocích velikosti 24x24 pixelů, které se jím postupně vyhodnocují.

## 2.3.Extrakce příznaků

V této kapitole budou popsány možné algoritmy extrakce příznaků vhodné pro detekci obličejů v obraze. Pro popsané algoritmy předpokládáme jako vstup obraz v odstínech šedi.

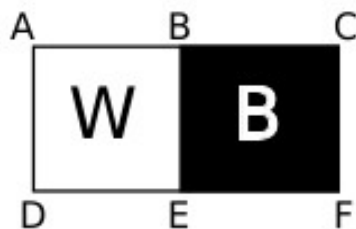
### Haarovy příznaky

Jedná se o jednoduchý typ příznaku, který je založen na jednoduchých aritmetických operacích. Jedná se o konvoluci jádra příznaku s danou oblastí. Vzhledem k tomu, že jádro obsahuje jen dvě možné hodnoty (1,-1) lze výpočet optimalizovat pomocí integrálního obrazu.

$$haar(ker) = \sum_{w \in ker_w} w - \sum_{b \in ker_b} b$$

*Vzorec 3: Výpočet  
Haarova příznaku pro  
jádro ker*

Pro velikost jádra 24x24 pixelů existuje 180 000 kombinací jader, které lze využít pro klasifikátor.



Obrázek 3: Princip výpočtu Haarova příznaku

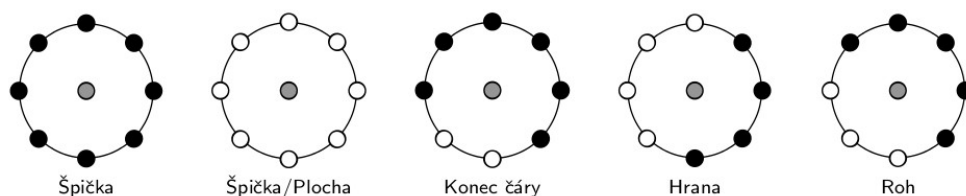
## Local binary patterns

Local binary patterns (dále jen LBP) vychází z principu porovnání centrálního pixelu se svým okolím. Díky tomuto principu je příznak invariantní vůči lokální změně osvětlení.

$$LBP_p = \sum_{p=0}^{P-1} sig(i(p) - i(c)) 2^p \quad sig(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Vzorec 4: Výpočet LBP příznaku

Jak je uvedeno ve vzorci 4 je příznak vypočten jako součet mocnin čísla 2. Přičemž o tom zda bude daná mocnina přičtena závisí na skutečnosti, zda hodnota pixelu, ke kterému přináleží je větší jako hodnota středového pixelu. Pro lepší ilustraci je na obrázku 4 ukázáno několik případů jak může LBP příznak vypadat.



Obrázek 4: Ukázka LBP vzorů a jejich významu v obraze

V případě, že se provede minimalizace LBP pomocí bitové rotace, dosáhne se kromě již zmiňované invariance vůči změnám jasu, také rotační invariance.



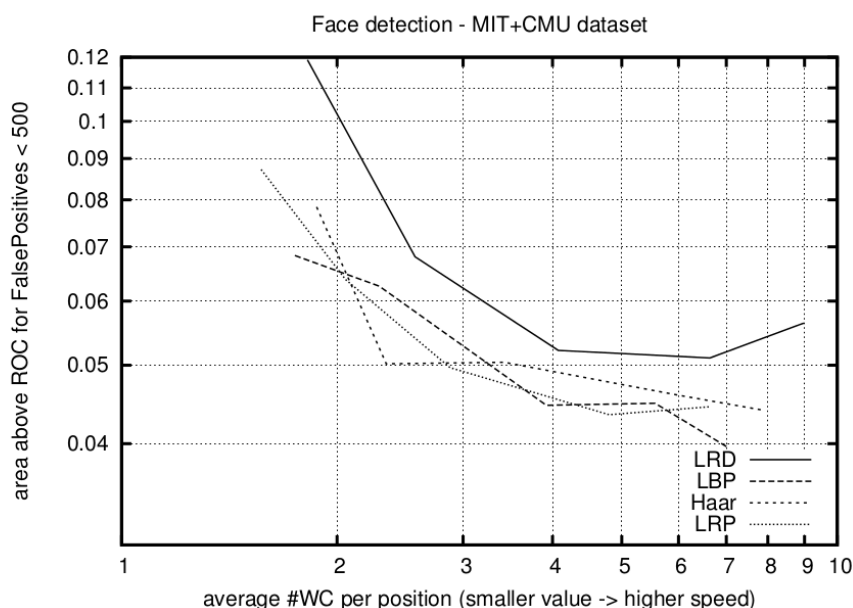
## 2.4.Klasifikace

Statistických klasifikátorů existuje široká škála typu, v poslední době se však pro detekci obličejů hojně využívá klasifikátor ADABOOST, který je založen na kaskádě slabých klasifikátorů, které se postupně vyhodnocují. Přičemž na jakékoli úrovni kaskády lze posuzovaný snímek (většinou podoblast obrazu 24x24 pixelů) odmítnout jako oblast neobsahující hledaný objekt.

Slabé klasifikátory jsou většinou implementovány jako váhované odezvy příznaků na aktuální okno. Tyto váhy jsou klasifikátoru nastaveny automaticky během procesu trénování.

Tímto chováním lze velmi rychle zavrhnout oblasti obrazu, na kterých se dané objekty zcela jistě nevyskytují, což je typicky převažující část obrazu.

Klasifikační algoritmus ADABOOST může na vstupu přijímat různé typy příznaků. Velmi často právě Haarovy příznaky[3] či LBP[4].



Obrázek 5: Srovnání rychlosti a kvality detekce pro různé příznaky na klasifikátoru ADABOOST [4]

Na obrázku 5 jsou zobrazeny výsledky porovnání kvality a rychlosti vyhodnocení jednotlivých příznaků. Na kolmé ose je plocha nad ROC<sup>1</sup> křivkou, přičemž platí, že v ideálním případě je plocha nad křivkou 0.

Z grafu je patrné, že pro všechny použité příznaky stačí v průměru vyhodnotit kolem čtyř slabých klasifikátorů na okno.

1 ROC (Receiver Operating Characteristic) křivka, je poměr skutečně pozitivních případů oproti falešně pozitivním případům

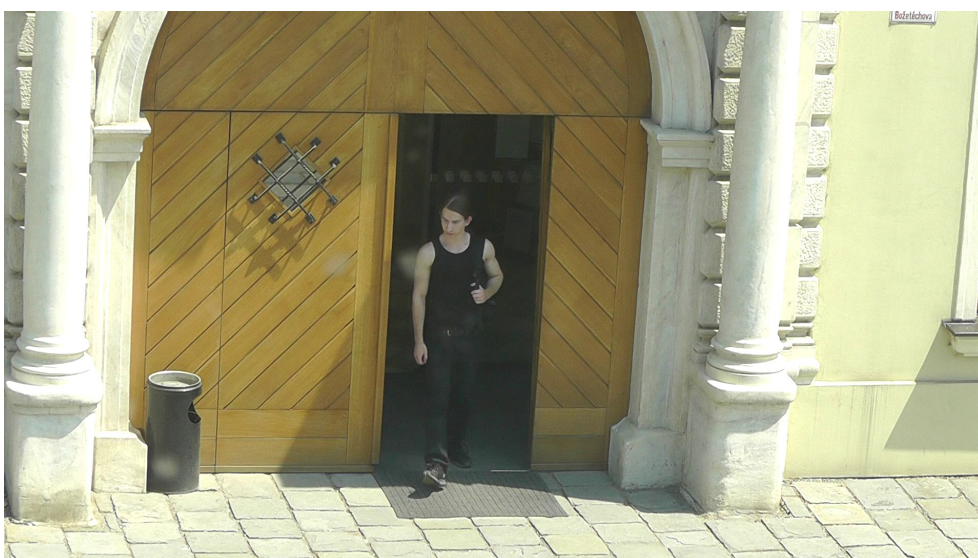
Vyjímkou jsou příznaky LRD<sup>2</sup>, které jsou ovšem na druhou stranu optimalizovány pro rychlý výpočet.



*Obrázek 6: Ukázka výstupu klasifikátoru obličejů*

Na obrázku 6 je výstup klasifikátoru ADABOOST na testovacích datech k projektu VideoTerror. Datová sada byla vytvořena pro potřeby tohoto projektu a simuluje snímání dat v podmínkách dohledových systémů.

Na obrázku 7 je zobrazena celá snímaná scéna z testovacího datasetu. Jedná se o vstupní bránu do areálu FIT na ulici Božetěchova.



*Obrázek 7: Ukázka snímané scény z testovacího datasetu.*

---

<sup>2</sup> LRD (Local Rand Diferences) více o příznamech viz [4]

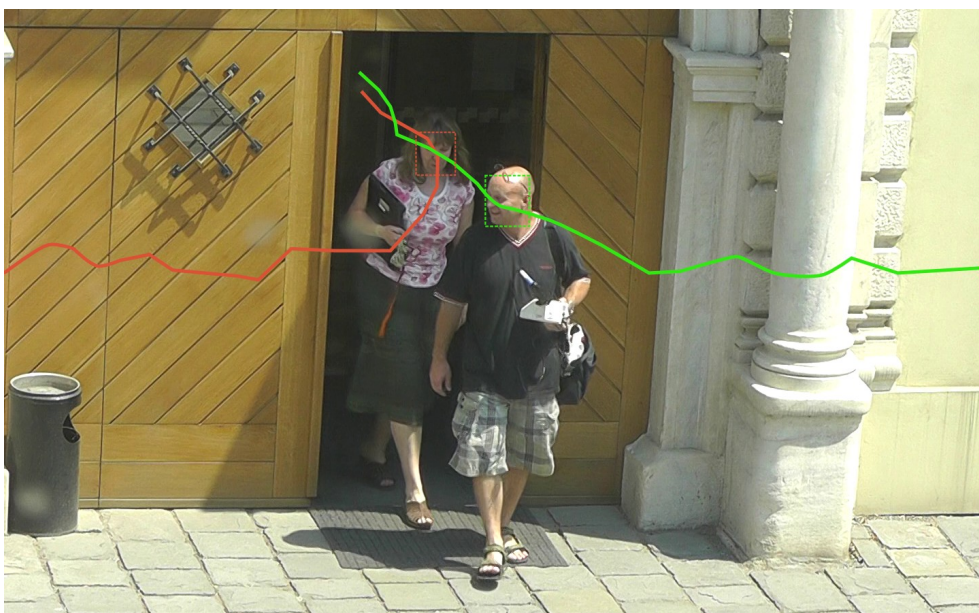
## 2.5. Trasování obličejů v čase

Aby bylo možné detekovat celý výskyt osob ve videu, je zapotřebí spojit výsledky klasifikace z jednotlivých časových okamžiků do jedné trajektorie.

V jednoduchém případě stačí daný detekovaný objekt, v našem případě obličej, popsat jednoznačným deskriptorem (např. HOG<sup>3</sup>) a detekce z jednotlivých snímků na základě těchto deskriptorů párovat.

Toto spojování se však musí umět vypořádat s faktem, že během celé doby výskytu objektu (obličeje) ve scéně nedochází vždy k pozitivní detekci obličeje. Toto může být například způsobeno otočením se osoby od snímacího zařízení, zákrytem obličeje jiným objektem (např. projíždějícím autem), sklopením hlavy o 90° směrem k zemi, atd.

Je proto záhodné použít mechanismu odhadu pohybu objektu a v době, kdy není objekt ve scéně viditelný odhadovat jeho teoretickou pozici a případně upravovat jeho deskriptor.



*Obrázek 8: Ukázka trasování obličejů v čase, barevné linky označují trajektorii jednotlivých osob.*

K vyřešení tohoto problému lze použít odhad parametru pozice pohybujícího se objektu pomocí Kalmanova filtru. Kalmanův filtr je vhodný také proto, že dokáže odhadnout pozici nejen

<sup>3</sup> Histogram of oriented gradients – je deskriptor popisující statistické rozložení množství gradientů v jednotlivých směrech.

z předchozích snímků ale také na základě budoucích. Tato skutečnost je výhodná obzvláště v případech offline zpracování videa, ve kterém můžeme předpokládat, že se nejprve zpracuje celé video pomocí klasifikátoru obličejů a teprve až následně se provede trasování pozitivních detekcí.

Na obrázku 8 je znázorněna trajektorie trasovaných postav ve videu.

### 3. Návrh začlenění do vyvíjeného systému

V současnosti není popisovaný mechanismus začleněn do funkčního prototypu. Toto začlenění je náplní práce v první polovině roku 2013.

Detektor osob lidí ve videu bude začleněn do vyvíjeného prototypu, následujícím způsobem.

O interakci s uživatelem se bude starat uživatelské rozhraní (dále jen UI) skrze webový prohlížeč. Webový server obstarávající tvorbu tohoto UI bude spuštěn na některém z projektových serverů (prozatím `medusa.fit.vutbr.cz`) a bude využívat technologii Django. UI bude mít na starosti nahrávání dat od uživatele a demonstraci výsledků detekce.

Veškeré zpracování obrazu tzn. Předzpracování, segmentace, klasifikace a trasování bude probíhat na serveru `vidte.fit.vutbr.cz` s tím, že vypočítané příznaky se budou ukládat na tomto serveru skrze datové API VTAPI vyvinuté v předchozím roce v tomto projektu. Toto API zjednodušuje přístup k datům z různých programů a dovoluje tak komfortní sdílení výsledků různých algoritmů unifikovaným způsobem.

Server Vidte je specializovaný pro náročné výpočty a obsahuje také širší škálu akceleračních jednotek (GPU a FPGA kity). Hodí se tak pro aplikaci detekce lidského obličeje ve videu. Díky tomuto faktu je předpoklad, že zpracování bude probíhat v reálném čase případně rychleji.



Obrázek 9: Vztah projektových serverů mezi sebou

Samotná obrazová data budou uložena na serveru `matylda1.fit.vutbr.cz`, který je specializován pro ukládání velkých objemů dat.

V budoucnu se předpokládá sjednocení všech částí výpočtu, datových úložišť a části generující uživatelské rozhraní na jeden server.

## 4. Závěr

Podle dosažených výsledků se jeví popisovaný systém jako vhodný pro použití např. v kriminalistice, ve které by mohl pomoci během dlouhodobých sledovacích akcí zkrátit čas strávený zpracováváním časově náročnějších video záznamů sledovaného objektu a zjišťování průchodu osob tímto objektem.

V minulém roce byla v tomto směru navázána spolupráce s Policií České Republiky, která zajistila vzorové data, na kterých by bylo možné ověřit použitelnost výše uvedeného systému pro kriminalistické účely. Toto ověření bude náplní činnosti v roce 2013.

Do budoucna počítáme s akcelerací daných algoritmů díky přesunu výpočtu na akcelerační jednotky (GPU a FPGA). Což by dozajista zvýšilo výkonnost a interaktivitu celého systému.

## 5. Literatura

- [1] Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. *International Journal of Computer Vision (IJCV)* 46(1) (2002) 81-96
- [2] Polok, L., Herout, A., Zemčık, P., Hradiš, M., Juránek, R., Jošth, R.: "Local Rank Differences" Image Feature Implemented on GPU, In: *Proceedings of the 10th International Conference on Advanced Concepts for Intelligent Vision Systems*, Berlin, Heidelberg, DE, Springer, 2008, s. 170-181, ISBN 978-3-540-88457-6
- [3] P. Viola, M. Electric, and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," *Learning*, 2001.
- [4] Herout, A., Zemčık, P., Hradiš, M., Juránek, R., Havel, J., Jošth, R., Žádník, M.: *Low-Level Image Features for Real-Time Object Detection*, Pattern Recognition, Recent Advances, Vienna, AT, IN-TECH, 2010, s. 111-136, ISBN 978-953-7619-90-9