

Obrazová detekce klíčových změn v sekvenci prezentačních snímků zaznamenaných systémem Mediasite



Bc. Martin Paroubek
Ing. Jan Silovský, Ph.D.
Ústav Informačních technologií a elektroniky

Abstract

The goal of this thesis was to create a tool for identifying duplicate slides in a set of support materials for lectures. From the materials at hand, two sets were created: a test set and a training set. For the solution, 30 features were used, from which 4 described the whole image (e. g. the sum of absolute values of grayscale histogram differences). For all features vectors, a classification was performed by finding the limit which separated the duplicity images class from the unique images class. Application of LDA and PCA methods on the set of features followed and their success rate was measured separately. Finally, an evaluation was performed using Support Vector Machines, which was considered the best choice. Even though certain downsides of this approach were discovered, the method was recommended for implementation in practice, because it kept nearly all the unique images in the tested sets.

Úvod

Při nahrávání přednášek na Technické univerzitě v Liberci dochází k záznamu doprovodných textů přednášejícího (prezentace, psaní na tabuli, psaní rukou na papír). Výstupem ze systému je cca 1000 obrázků, ale pouze zhruba 3,5 % je jedinečných. Za jedinečné byly považovány nejobsáhlejší neduplicitní snímky.

Cíl

Úkolem bylo vytvořit nástroj, spustitelný z příkazové řádky, který redukuje počet duplicitních snímků. Vstupem programu je adresář se sadou obrázků. Výstupem programu je seznam jedinečných snímků.

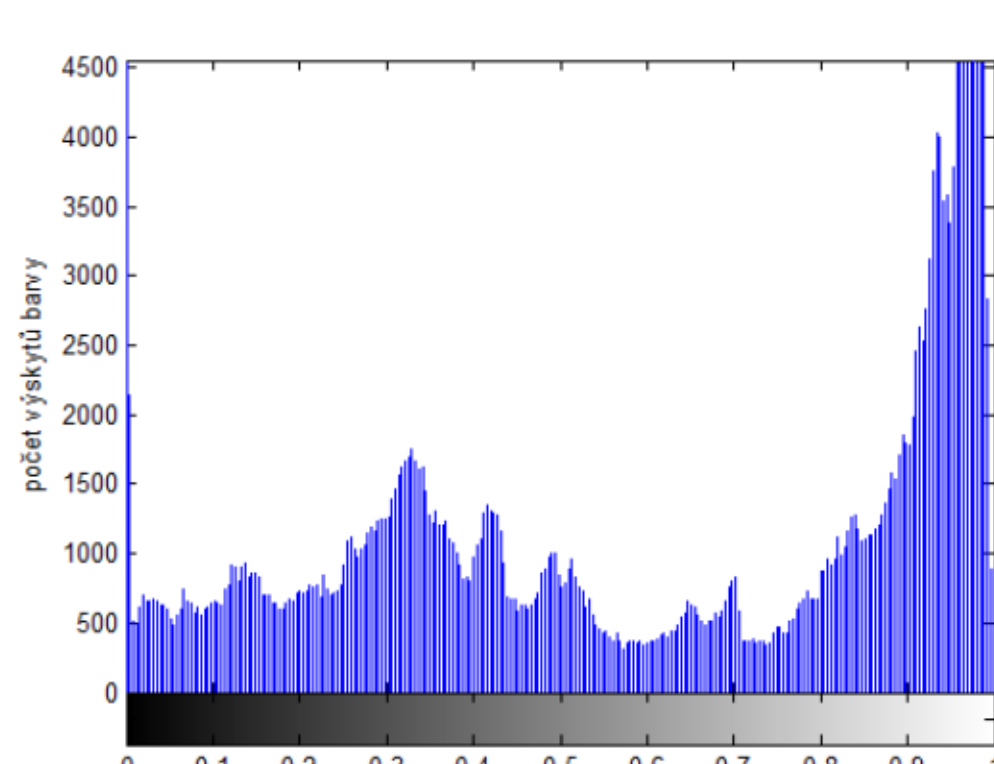
Metodika

Úspěšnost byla měřena vypočítáním F-score (1). Při trénování s tímto ukazatelem je výsledkem nižší počet nesprávně identifikovaných jedinečných snímků. Dále byla zjišťována hodnota True Negative Rate (TNR), která vyjadřuje úspěšnost identifikace duplicitních snímků.

$$F_{\beta} = (1 + \beta^2) \frac{\text{precision} \times \text{recall}}{(\beta^2 \times \text{precision}) + \text{recall}} \quad (1),$$

kde β je rovna 2 (ve prospěch jedinečných obrázků), precision je poměr správně klasifikovaných jedinečných snímků vůči všem klasifikovaným jedinečným snímkům a recall je poměr správně klasifikovaných jedinečných snímků vůči všem jedinečným snímkům.

Dále byla vytvořena testovací sada, která obsahovala 21 přednášek s 20 304 snímky a z toho bylo pouze 725 jedinečných. Všechny snímky prošly předpřípravou, která spočívala v převedení barev do stupňů šedi, v rozostření a zmenšení velikosti. Z takto připravených obrázků byly extrahovány příznaky s použitím jejich samotných, jejich hranové reprezentace, histogramů stupňů šedi a histogramů orientovaných gradientů (viz obr. 1 a obr. 2).

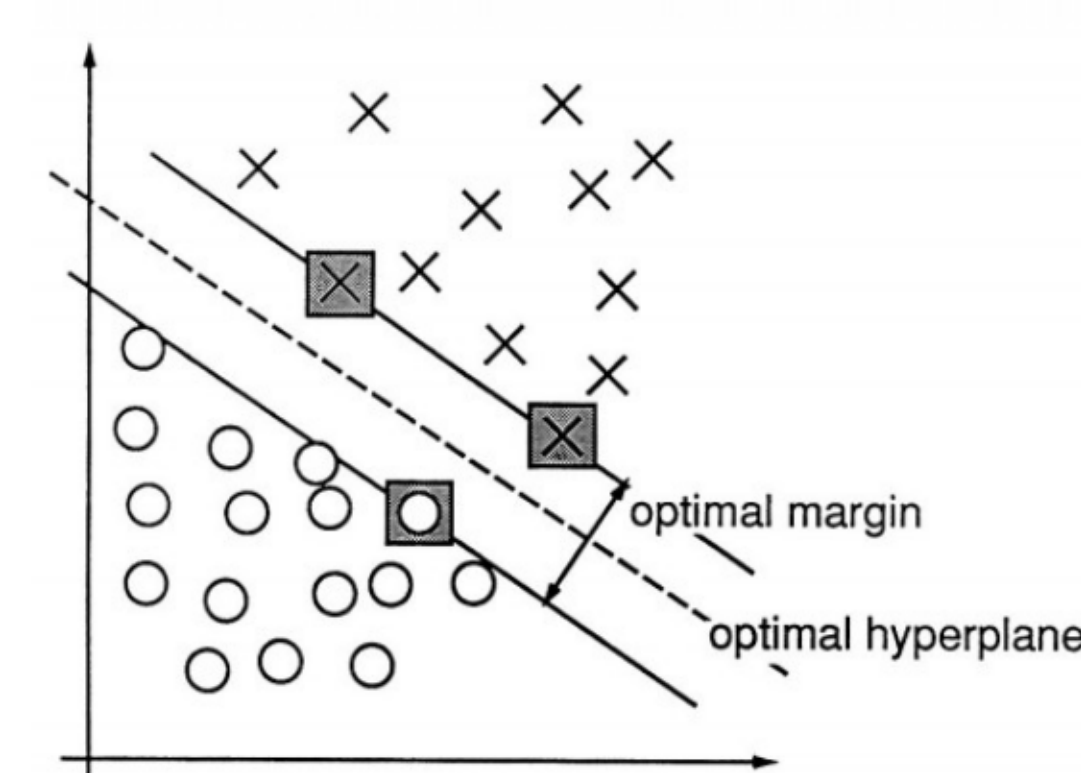


Obrázek 1: histogram stupňů šedi



Obrázek 2: hranová reprezentace snímku

Hlavním klasifikátorem byly zvoleny Support Vector Machines (dále jen SVM). Wang [1] konstatoval, že v nejjednodušších případech SVM využívají nadrovinu lineárně oddělující data k vytvoření klasifikátoru s maximálním odstupem (viz obr. 3). V jiných případech mohou být příznakové vektory transformovány do vícedimenzionálního prostoru (způsob transformace určuje kernel).



Obrázek 3: Příklad lineárně separovatelného problému ve 2dimenzionálním prostoru.

Výsledky:

Nejlepším řešením s celkovou úspěšností 98,21 % byl klasifikátor Support Vector Machines pro 30 příznaků s Radial Basis F. kernel a parametry $\log_2 C = -5$ a $\log_2 g = 6$. Byl přitom vytvořen modul pro klasifikaci s 817 Support Vectors.

Výsledný program byl také odzkoušen na sadě s 1841 snímky. Klasifikace snímků do tříd trvala 12 minut a 23 sekund. K měření byl použit notebook s 2 jádrovým procesorem Core 2 Duo 2,4 GHz a 2 GB RAM.

Tabulka 1. Výsledky pro natrénované cost a gamma vlastním algoritmem

Kernel type	F ₂ -score	Precision	Recall	TNR	log ₂ (C)	log ₂ (gamma)
Linear	90,21	100	88,06	100	-7	0
Polynomial	88,96	100	86,57	100	6	1
Radial basis f.	98,21	97,06	98,51	99,94	-5	6
Sigmoid	93,41	89,2	94,53	99,77	5	-5

Závěr:

Výsledný algoritmus dosáhl úspěšnosti při identifikaci klíčových snímků přes 98 %, ale v případě výskytu některých snímků (s rukou nebo stínem přednášejícího) docházelo k horší detekci snímků duplicitních.

V navazující práci by mohlo dojít k vytvoření dalších algoritmů řešících: detekci snímků bez informační hodnoty (jednobarevné přechodové snímky, nebo rozmazané snímky), invarianci osvětlení a zakrytí. V případě sériového zapojení těchto algoritmů by mohlo dojít i k časové úspoře.

Reference:

- [1] WANG, Lipo. *Support vector machines: theory and applications*. Berlin: Springer, 2005, s. 4. ISBN 3540243887.
- [2] HSU, Chih-Wei, Chih-Chung CHANG a Chih-Jen LIN. *LIBSVM: A library for support vector machines*. In: ACM Transactions on Intelligent Systems and Technology, 2011, volume 2, issue 3, s.1-27. Software dostupný z: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

Kontakt:

Bc. Martin Paroubek

martin.paroubek@gmail.com