

RECOGNITION OF HANDWRITTEN DIGITS

Martin Dobrovolný

Bachelor Degree Programme (3), FIT BUT

E-mail: xdobro06@stud.fit.vutbr.cz

Supervised by: Adam Herout

E-mail: herout@fit.vutbr.cz

ABSTRACT

Recognition of handwritten digits is one of computer vision problematics that can not be solved with 100% success in these days. This document describes a method for handwritten digits recognizing based on shape features and randomized tree classifiers. These methods are known for their long time machine learning and quick characters recognizing. This method is due to use of relative angles and distances among key locations and is nearly invariant to substantial affine and nonlinear deformations.

1. ÚVOD

V dnešní době existuje mnoho metod pro rozpoznávání ručně psaných čísel. Mnou užitá metoda je založena na porovnávání tvarových charakteristik a hlasování více klasifikačních stromů. Tvarové charakteristiky užívají relativních úhlů a vzdáleností mezi význačnými body, což umožňuje jistou míru volnosti tvaru charakteristiky, a možnost porovnávat podobnost mezi testovaným vzorkem a natrénovanými tvarovými charakteristikami. Více klasifikačních stromů a jistá náhodnost při jejich vytváření v danou metodou nám umožňuje dosáhnout statisticky lepších výsledků.

Moje práce vychází z metody popsané v článku [1].

2. ROZPOZNÁVÁNÍ RUČNĚ PSANÝCH ČÍSEL

Jednotlivé fáze algoritmu popisovaného v článku [1] je možné rozdělit následovně:

1. Hledání význačných bodů v obrázku
2. Tvorba stromu tagů
3. Tvorba klasifikačního stromu
4. Rozpoznávání ručně psaných čísel

Fáze 1. až 3 nazýváme trénováním, algoritmus hledá charakteristické rysy jednotlivých číslic – učí se rozpoznávat. Fázi 4 nazýváme testováním. Fáze trénování trvá mnohem déle vzhledem k velikosti databáze trénovacích dat. Trénování i testování probíhá na dvoubarvných, černobílých obrázcích.

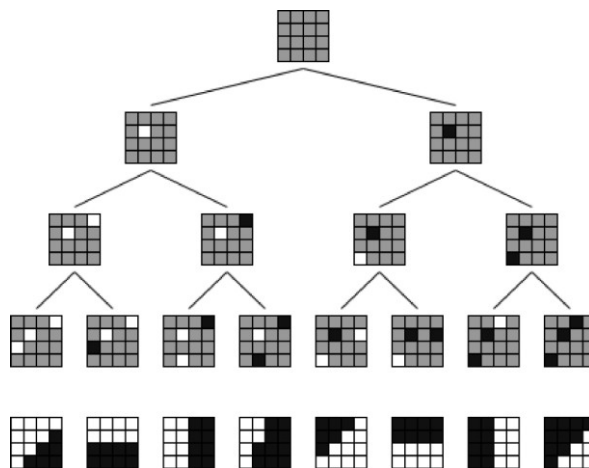
2.1. HLEDÁNÍ VÝZNAČNÝCH BODŮ

První fází algoritmu je hledání význačných bodů v trénovaných obrázcích. Význačný bod je malá část obrázku která nám poskytuje částečnou informaci o tvaru čísla. Tato malá část obrázku musí obsahovat alespoň jeden pixel jinak zbarvený než ostatní pixely. Tímto způsobem se získají význačné body položené na hranách napsaného čísla. Takto získané význačné body nazýváme tagy.

2.2. STROM TAGŮ

Poté co jsme získali tagy ze všech trénovaných obrázků je potřebujeme rozdělit do jednotlivých tříd podle toho, jak tag vypadá, respektive z jaké oblasti čísla tag pochází.

Toto nám umožňuje binární strom tagů. Každý uzel stromu může obsahovat libovolné množství tagů, ale množství tagů v jeho synech musí být rozděleno co nejrovnoměrněji. Kořenový uzel obsahuje všechny tagy a ty se dále dělí do synů podle barvy jednoho pixelu v tagu. Toto demonstruje obrázek 1 pro tagy o velikosti 4x4 px. Strom nám tagy rozděluje do požadovaného množství tříd. Toto množství ovlivňuje hloubka stromu. Jako třída tagů se užívají všechny skupiny tagů vzniklé v jednotlivých uzlech kromě kořenového uzlu. Mnou vytvářené stromy o hloubce 6 tedy rozdělují tagy do 62 tříd.



Obrázek 1: Strom Tagů, ukázka dělení tagů, převzato z [1]

Takový strom rozděluje tagy do tříd, v rámci kterých jsou si tagy navzájem vizuálně podobné. Například mohou všechny náležet horní hraně čáry, nebo levým spodním hranám čáry apod.

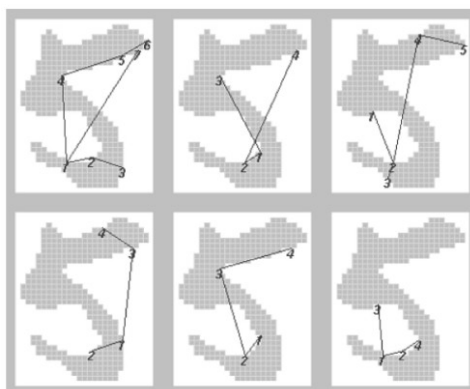
2.3. TVORBA KLASIFIKAČNÍHO STROMU

Použitý klasifikační strom je binární strom sloužící pro rozhodování o výsledku rozpoznávání. Díky jeho stromové struktuře je rozpoznávání velmi rychlé.

Kořenový uzel obsahuje všechny obrázky z trénovací databáze. V každém uzlu hledáme takový vzájemný vztah dvou tagů, který tyto obrázky rozdělí do synů na ty, které vztah obsahují a na ty které uspořádání neobsahují. Tento vztah lze tedy nazývat pravidlem.

Vztah obsahuje dvě různé třídy tagů a směr mezi nimi. Úhly mezi dvěma tagy jsou rozděleny do osmi směrů. Pro lepší představu je lze popsat světovými stranami (sever, severovýchod, východ, ... severozápad). Vzhledem k množství tříd tagů nám vzniká 62 x 62 x 8 možných kandidátů na vhodné pravidlo. Toto množství je z hlediska výpočetního času neúnosně velké, proto se vybírá pouze z malého množství náhodně vygenerovaných pravidel.

Kvalitu pravidla posuzujeme podle dvou kritérií. Prvním kritériem je vyváženost stromu kvůli budování stromu do šířky pro rychlejší vyhodnocování. Druhým kritériem je kvalita separace jednotlivých typů číslic. Snažíme se, aby listy obsahovaly jeden, nebo několik málo typů čísel, které mají početně co největší převahu nad ostatními typy. Všechna pravidla která v cestě mezi kořenovým uzlem a listem rozhodovacího stromu konkrétní vzorek obsahuje tvoří tzv. uspořádání.



Obrázek 2: Uspořádání vzorku nalezená v různých stromech, převzato z [1]

Vzhledem k náhodnosti při výběru tagů a při výběru rozhodovacího pravidla se pro zvýšení úspěšnosti rozpoznávání užívá vyhodnocování více stromů současně. Na obrázku 2 jsou vidět různá nalezená uspořádání v různých stromech, avšak u stejného trénovaného obrázku.

2.4. ROZPOZNÁVÁNÍ RUČNĚ PSANÝCH ČÍSEL

V předchozích odstavcích jsme si okrajově popsali, jakým způsobem se vytváří klasifikační strom. Nyní si ještě stručně popíšeme průběh vyhodnocování testovaného obrázku.

Postup při vyhodnocování obrázku číslice:

1. Nalezení tagů v testovaném obrázku
2. Nalezené tagy jsou rozřazeny do tříd pomocí dříve vytvořeného tagovacího stromu.
3. Průchod klasifikačním stromem až do listu. V listu nás zajímá informace o kvantitativním rozložení jednotlivých typů čísel při trénování.

3. ZÁVĚR

Cílem práce je vytvořit program, který bude dosahovat co nejvyšší úspěšnosti rozpoznávání. Nyní se program nachází ve fázi dokončování a prvních testů. Momentálně nejvyšší dosažená úspěšnost rozpoznávání ručně psaných číslic dosahuje 73% na databázi MNIST. Programy založené na výše popsaných principech běžně dosahují úspěšnosti nad 95% na stejné databázi.

LITERATURA

- [1] Amit, Y., German, D., WILDER, K.: Joint Induction of Shape Features and Tree Classifiers. IEEE Trans. on PAMI, 19/11, 1997