

Електротехнички факултет

Бањалука

Дипломски рад на тему:
Оптичко препознавање знакова

Ментор:

Др. Зденка Бабић

Кандидат:

Винчић Тихомир

Садржај:

1	Увод	1
2	Основни појмови	1
2.1	Основне карактеристике писма	1
2.2	Уређаји за унос	2
2.3	Процес препознавања	3
2.3.1	Улога препознавања знакова	3
2.3.2	Језички модели	4
2.3.3	Улога анализе слике документа	5
2.4	Будућност аутоматског препознавања писма	5
3	Анализа слике документа	6
3.1	Текстуални документи	7
3.2	Обрасци	7
3.3	Поштанске адресе и читање чекова	8
3.4	Цртежи	8
3.5	Напредак у анализи докумената	9
4	Поступци при оптичком препознавању штампаних знакова	9
4.1	Приступ препознавању анализом слике документа	10
4.1.1	Одређивање размјештаја	10
4.1.2	Изградња знакова	10
4.1.3	Утицај фонтова	11
4.2	Помоћни поступци при препознавању знакова	11
4.2.1	Одређивање особина знакова	11
4.2.2	Учење симбола	12
4.2.3	Контекстна обрада	12
4.2.4	Очекивања у области оптичког препознавања знакова	13
4.3	Технике сегментације представе свјетлине	13
4.3.1	Основе сегментације свјетлине	13
4.3.2	Сегментација кориштењем ивичних тачака	15
4.3.3	Поступак итеративног бирања	15
4.3.4	Поступак анализе варијансе хистограма свјетлина	16
4.3.5	Сегментација употребом ентропије	17
4.3.6	Сегментација помоћу фази скупова	19
4.3.7	Сегментација најмањим одступањем	21
4.3.8	Поређење поступака сегментације једним прагом	22

4.3.9	Употреба обласних прагова	25
4.3.10	Поступак Чоуа и Канека	26
4.3.11	Узорковање освијетљености помоћу ивица	28
4.3.12	Релаксациони поступци	29
4.3.13	Помични просједи	30
4.4	Одређивање нагнутости	30
5	Препознавање знакова	33
5.1	Поступак придруживања образаца	35
5.2	Поступци вјештачке интелигенције	39
6	Закључак	40
7	Литература	41

Увод

Јављају се многе примјене у којима је потребан унос штампаних или писаних докумената у рачунар. Постоји огромна количина информација у штампаним документима и књигама, међутим руковање тим информацијама је отежано јер се оне не могу искористити у модерним рачунарским системима за индексирање и добављање докумената. Због тога постоји тежња да се што већа количина информација преведе у симболичку представу у рачунару, погодну за обраду и руковање рачунаром. Симболичка представа омогућава лако претраживање, добављање, пренос, обраду и чување информација. Такође, обрада података из разних ручно попуњених образаца захтијева њихов унос у рачунарске базе података што представља велики трошак у случају ручног уноса. Постоји потреба и за разврставањем докумената на основу информација садржаних на њима (нпр. писама) што је далеко јефтиније извршити рачунаром.

Писани језик се јавља у облику рукописа и штампе. С обзиром на важност писаног језика (писама) у међуљудској комуникацији његово машинско препознавање у данашње вријеме добија све већи значај.

Основни појмови

Основне карактеристике писма

Писмо се састоји од симболичких представа (симбола) на површини. Сврха му је комуникација, што се остварује помоћу договореног значења симбола у оквиру језика. Иако је говор далеко природнији за човјека сматра се да је писмо омогућило већину културних и цивилизацијских достигнућа.

Различита писма представљају лингвистичке јединице (ријечи, слоге и гласове) на различитим структурним нивоима. У словним писмима, нпр. Ћирилице, Латинице и Алфabetу, симболи представљају слова, чијим слагањем се формирају ријечи. У писму Деванагари симболи представљају слова и слоге. Код кинеског писма симболима су представљени идеограми. Јапанско писмо се састоји од кинеских идеограма (Канџи) и симбола који представљају слоге (Кана). Поред тога, Јапански се може писати и помоћу два словна писма (Катакана и Хирагана).

Поједина писма се употребљавају у оквиру више различитих језика. У данашње вријеме се употребљава тридесетак различитих писама, не узимајући у

обзир минорне разлике у правопису између различитих језика који користе исто писмо, као нпр. између Српског и Бугарског.

Свако писмо има свој скуп симбола који имају одређене основне облике. Поред тога у сваком писму постоји скуп правила за слагање основних симбола да би се добили облици који представљају више лингвистичке јединице, нпр. правила за слагање писаних слова у Ћирилицы која се користе при писању ријечи у Српском језику.

Поред лингвистичких симбола свако писмо има систем за представљање бројева, било посебним скупом симбола (нпр. Индо-Арапски бројеви у Ћирилицы и Латиници) било слагањем слова (нпр. Глагољица, Срарогрчки, Старојеврејски, Латиница у вријеме Римског царства). У данашње вријеме се такође јављају и симболи за специјалне знакове на рачунарским тастатурама (табулатор, размакница, Delete, Backspace, Insert).

Уређаји за унос

Од открића штампарске пресе писмо се већином јавља у облику штампе. Код таквих докумената слова су представљена симболима из ограниченог скупа фонтова, а контраст симбола у односу на позадину је велик. Овакви документи се уносе у рачунар оптичким скенером који аналогне нивое освијетљености на документу претвара у низ бројева одређене резолуције по простору и нивоу освијетљености. Просторна резолуција се изражава бројем тачака по инчу - dots per inch (dpi) у земљама које користе империјални систем јединица, док би у земљама које користе интернационални систем јединица (si) требало да се користи број тачака по милиметру (mm^{-1}). У данашње вријеме је стандардна просторна резолуција рачунарских скенера 300 dpi ($11,8 \text{ mm}^{-1}$) по оба правца и 8 бита (256 нивоа) по нивоу освијетљености. Код факс машина стандард је 200 dpi ($7,87 \text{ mm}^{-1}$) по правцу оптичког читавања, а 100 dpi ($3,94 \text{ mm}^{-1}$) по правцу кретања папира и 1 бит (2 нивоа) по нивоу освијетљености. За поуздано препознавање мањих слова потребна је већа просторна резолуција. Писмо се такође јавља у облику рукописа исписаног на папиру или регистрованог на осјетљивој електронској површини. Рукопис се уноси у рачунар било скенирањем исписаног папира или писањем специјалном оловком по осјетљивој електронској површини (нпр. екран са течним кристалима или дигитализаторска табла). Ова два приступа уносу се називају

вансистемско (off-line) и унутарсистемско (on-line) писање. Код унутарсистемског писања се редоследом по времену читавају дводимензионалне координате врха оловке у току писања, тако да се добија низ потеза које је писац направио у току писања, док се код вансистемског писања добија само слика написаног писма без могућности сазнања како се кретала оловка при писању. Унутарсистемско писање рукује са једнодимензионалном представом улаза (положај у зависности од времена), док вансистемско писање захтијева анализу дводимензионалне слике. Количина података које треба обрадити је далеко већа код вансистемског писања (за просјечну писану ријеч при унутарсистемском писању 230 бајтова при учестаности одмјеравања од 100 одмјерака у секунди, а при вансистемском писању 80 килобајта при просторној резолуцији 300 dpi). Такође, успјешност препознавања је далеко већа код унутарсистемског писања.

Процес препознавања

Задатак препознавања писма је претварање писма из просторног облика састављеног од графичких знакова у симболичку представу у рачунару. За Енглески правопис уобичајен начин ове симболичке представе је у облику ASCII кода, док је већину осталих језика могуће представити помоћу Уникода (Unicode).

Главни задаци при препознавању писма су препознавање слова и препознавање ријечи. Неопходан корак у предобради је одређивање мјеста на којима се налази текст код докумената са сложеним распоредом текста и графике. Овај поступак се назива анализа слике документа.

Улога препознавања знакова

Основни задатак је да се одреди којој симболичкој класи припада дигитализовани знак. Препознавање штампаних симбола се назива оптичко препознавање знакова, док се препознавање ручно писаних симбола за штампу назива интелигентно препознавање знакова.

Типичне класе су велика и мала слова, цифре, и специјални симболи (интерпункцијски, новчани и други). Да би се одредила класа знака употребљава се неки од алгоритама за препознавање узорака. Најлакши метод за реализацију класификатора је помоћу вјештачке неуронске мреже. Алгоритми засновани на

методима највеће сличности су прецизнији, али уколико је број могућих различитих класа велик далеко су спорији од неуронских мрежа.

Знакове из једног фонта са квалитетно одштампаног документа је могуће веома прецизно препознати. Међутим, уколико се користи велик број фонтова или је квалитет штампе лош препознавање је отежано. У том случају потребно је употријебити моделе који ће ограничити могуће изборе на нивоу знака или ријечи. Овакви модели су основа при препознавању рукописа због велике промјењивости ручно штампаних и писаних слова.

Алгоритми за препознавање ријечи препознају слику читаве ријечи додјељујући вјероватноће појединим ријечима из ријечника дозвољених ријечи. Ово се ради или аналитичким приступом, гдје се препознају поједини карактери и додјељују им се вјероватноће на основу којих се бира ријеч са највећом вјероватноћом, или холистичким приступом, гдје се препознавање читаве ријечи врши у једном кораку. Холистички приступ се користи у случају међусобног додиривања сусједних знакова и препознавања рукописа.

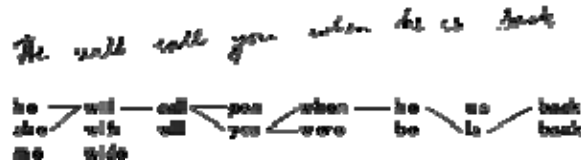
Језички модели

Језички модели се користе за опоравак низа ријечи након проласка кроз канал са шумом какав је рукопис или штампа слабог квалитета. Најзначајнији модел при препознавању писма је ријечник. Ријечник садржи лингвистичка ограничења, тако да је у току препознавања текста скуп могућих ријечи за сваку слику ријечи ограничен синтаксом, семантиком и прагматиком реченице.

Укључивање статистичких информација на нивоу низа ријечи повећава перформансе система за препознавање. До повећања перформанси долази због могућности избора ријечи која је мање вјероватна на основу слике, уколико због таквог избора реченица постаје вјероватнија. Сусједне ријечи одређене системом за препознавање ријечи се модификују техником колокационе анализе и њој сличним техникама. Модификације укључују: мијењање вјероватноћа појединих ријечи, брисање и предлагање нових ријечи. Колокати су ријечи које се често у неком језику јављају у близини, нпр. ако је ријеч А присутна онда је велика вјероватноћа да се у њеној близини налази ријеч Б.

Међу методима за примјену лингвистичких знања су: н-грамски модели низова ријечи, н-грамски модели низова класа (нпр. дијелова говора), безконтекстне

граматике и стохастичке безконтекстне граматике. Примјер рукописне реченице Енглеског језика са избором ријечи система за препознавање и граматички одређеном исправном путањом је приказан на слици 1. Употребом језичких модела могуће је код система за препознавање ријечи повећање тачност првоизабране ријечи за 80% до 95%.



Слика 1:

Препознавање реченице из рукописа.

Путања кроз првоизабране ријечи је одређена кориштењем ознака дијелова говора.

Улога анализа слике документа

Анализа слике документа је међузависна са препознавањем писма. Њен задатак је одређивање просторне и логичке структуре садржаја документа. Постоји велика промјенивост међу структурама докумената као што су новине, магазини, књиге, обрасци, писма и руком писане биљешке. У случају новинске стране сврха анализе слике документа је: одређивање просторног распореда дијелова документа и додјеливање одговарајућих ознака појединим дијеловима (ознаке су нпр. слика, текст, графика, линија за раздвајање), груписање дијелова слике у смислене јединице (нпр. слика и њен назив, наслов и поднаслови), одређивање редоследа дијелова текста.

Анализа слике документа користи на штампаном тексту традиционалне операције обраде слике попут разних метода побољшања слике (нпр. филтрирање шума, појачање контраста), анализе текстуре, просторна и сегментација по нивоу и других. Битно је напоменути да су при препознавању рукописа отежани следећи поступци: раздвајање редова текста, раздвајање ријечи у реду и раздвајање знакова који се додирују.

Будућност аутоматског препознавања писма

Истраживање аутоматског препознавања писма се врши већ неколико деценија. У данашње вријеме се текст у добро одштампаним документима са простим распоредом може поуздано препознати комерцијалним програмима. Постоје неки помаци и у препознавању рукописа, нарочито раздвојених штампаних

слова и ријечи при унутарсистемском писању, што је омогућило развој личних дигиталних помоћника (personal digital assistants). У ограниченим доменима као што су поштанске адресе, чекови и пописни обрасци постигнут је напредак и при вансистемском писању. Анализа докумената са сложеним распоредом, препознавање оштећеног штампаног текста и препознавање брзо писаног рукописа и даље остају у истраживачкој области. Неки од главних истраживачких изазова у препознавању рукописа су: раздвајање ријечи и редова, подјела ријечи у слова, препознавање ријечи из обимних ријечника и употреба језичких модела ради побољшања предобrade и препознавања.

Анализа слике документа

Анализа слике документа је ток обраде који даје потпуно тумачење слике документа. Њен циљ је комбиновање знања о језику, уређењу документа, обради слике и препознавању знакова да би се обрадили документи одређене намјене. Према томе анализа документа се бави општим питањима препознавања писма са слике. Она надграђује оптичко препознавање знакова јер утврђује уређење документа и примјењује претходна знања при његовом тумачењу.

Одређивање уређења документа се врши према израженом или подразумијеваном обрасцу за врсту докумената која нас занима. Образац описује изглед и међусобне односе основних градивих јединица документа. Оптичко препознавање знакова је често задњи поступак при обради, тј. оно даје коначно кодовање симбола садржаних у логичкој јединици попут параграфа или табеле након што је она одређена у претходним етапама. Међутим, битно је напоменути да се при одређивању уређења документа користи и оптичко препознавање знакова. На примјер може се десити да при поступку вађења новинског чланка систем мора да препозна на дну стране низ знакова „наставља се на страни 5“ да би нашао читав текст.

Практично систем за анализу документа врши основне задатке сегментације слике, разумијевања уређења, препознавања симбола и примјену контекстуалних правила на обједињен начин. Тренутно се рад у овој области може сажети под четири главне врсте примјена.

Текстуални документи

Коначни циљ текстуалних система се може назвати обрнуто уређивање или затварање Гутенбергове петље, што значи да је скенирани документ поново преведен у језик за опис документа из ког се може по потреби поново одштампати. На истраживачком нивоу овом се тежи у областима као што су технички чланци, пословна писма и хемијски структурни дијаграми. Неки комерцијални системи за оптичко препознавање знакова већ имају ограничене могућности обрнутог уређења, пошто дају кодове за основне јединице попут параграфа, колона и табела. Данашњи системи за оптичко препознавање знакова препознају али не кодује слике и линијску графику.

У појединим примјенама није неопходно потпуно тумачење документа. Систем за индексирање и добављање текстуалних докумената може препознавати само поједине дијелове документа. На примјер комерцијално расположив систем за добављање техничких чланака садржи обрасце разних стилова журнала који му омогућавају налажење и препознавање наслова, аутора и сижеа сваког чланка и вађење кључних ријечи. Корисници претражују кодовани материјал али за читање добијају скениране слике жељених чланака.

Обрасци

Обрасци представљају односе у бази података. Образац се обично састоји из више података чланова који се могу представити као уређени парови (име члана, вриједност члана). Оптичко препознавање знакова се користи при препознавању вриједности члана. Да би се препознало и име члана најчешће је потребно извршити општију анализу документа.

Постоје многе примјене у области пословања и државне управе у којима је потребно пронаћи податке на обрасцу, установити имена чланова и сачувати пратеће вриједности у базу података као податке чија врста има то име. Уређење обрасца одређене намјене је најчешће непромјењиво у оквиру једне установе. Због тога се образац уређења докумената једне врсте може састојати већином од физичких параметара чије се вриједности процјењују са узорка таквог документа. Такви системи за скупљање података са образаца постоје у продаји.

Поштанске адресе и читање чекова

Сличност ових примјена је у добро утврђеном логичком поретку, јако промјењивом уређењу и великим контекстним ограничењима симболичких података. Контекстна ограничења омогућавају постизање велике тачности. Контекстна правила могу промијенити резултате оптичког препознавања знакова да би обезбиједила слагање имена градова и поштанских бројева или да ускладе бројно писане валутне износе на чековима са онима писаним словима. Контекстна ограничења могу помоћи и при откривању погрешно препознатих докумената како би се они могли обрадити ручно или неким рачунски захтјевнијим поступком. Иако су поштанске пошиљке и чекови у ствари подврста образаца, они су посебно набројани због огромних напори уложених у њихову обраду.

Опрема која се тренутно користи у ту сврху само дјелимично користи контекстне информације и може читавати само поштанске бројеве из руком писаних адреса или бројне износе са чекова. Новији поштански уређаји читавају читаву адресу и постижу већу тачност због примјене контекстних ограничења. Истовремено омогућају већу уситњеност у распореду пошиљки. На примјер у Сједињеним Државама новији уређаји врше распоред пошиљки по путањама појединачних поштара.

Цртежи

Највећи напори у овој области су усмјерени на уношење нацрта у системе за израду нацрта помоћу рачунара и производњу помоћу рачунара (CAD/CAM). На пројекту за унос нацрта интегрисаних кола је постигнут јефтинији унос нацрта од уобичајеног ручног уноса. Овај пројект доказује да је нацрте нових кола најјефтиније израдити на папиру а затим унијети системом за препознавање. Тврди се да је то боље од непосредног уноса на терминалу због мале величине приказа на данашњој опреми. Такви системи постоје у продаји. Врши се истраживање у циљу добијање тродимензионалних модела који омогућају вишеструке погледе на нацрте производних дијелова. Постигнут је напредак и у претварању катастарских мапа у информациони облик.

Напредак у анализи докумената

Један од мотива за рад у анализи докумената је велико повећање потребе за системима за рад са сликама у пословању и управи. Ови системи омогућују брз смјештај, позивање и расподјелу докумената у обради пословања и другим примјенама. Анализа докумената помаже при индексирању ради смјештаја и добављања и омогућује подјелу слике у по областима занимања ради лакшег приступа корисника.

У блиској будућности овакве могућности ће се раширити и до стварања електронских библиотека које ће се исто тако користити услугама аутоматског индексирања и уређивања. У даљој будућности ће се увећати напори да се протумачи више информација представљених на сачуваним сликама да би се омогућила већа прилагодљивост при добављању и руковању.

Јавља се велика потреба да се у нацртима система за анализу докумената предвиди већа употреба контекста, нарочито језичког. Знање о контексту се мора употријебити и ради свођења броја грешака на најмању мјеру и ради одбијања докумената који се не могу аутоматски протумачити. Значајан предмет истраживања је уопштен начин дефинисања таквих правила, тако да се могу једноставно прерадити за различите примјене. Битно питање је и начин преношења правила у софтвер који се интегрише са процесом препознавања који ће омогућити највећи учинак.

Језичка анализа ће у будућим системима за анализу докумената бити укључена у поступак анализе умјесто да буде одвојен корак, након обраде као што је данас. Савремени поступци препознавања изводе пробне сегментације слика знакова и на основу поузданости препознавања бирају најбољу сегментацију из скупа могућих. Такав поступак се изводи најпоузданије као редни низ поступака гдје контекстна правила управљају избором низа знакова.

Поступци при оптичком препознавању штампаних знакова

У данашње вријеме постоји велика мотивација да се обезбиједи рачунарски системи за анализу докумената. Допринос оптичког препознавања знакова се састоји у омогућавању аутоматског претварања великих количина података у информациони облик. Многи комерцијални програми рекламирају тачност

препознавања изнад 99.9 %, што даје утисак да је проблем оптичког препознавања знакова ријешен. Међутим, неуспјех у неким стварним примјенама показује да је учинак проблематичан на мјешовитим и оштећеним документима (јак шум, мјешовити фонтови, нагиб, итд.) што оставља простор за даљи напредак. Предложени су различити поступци за повећање тачности препознавања. Изазов је у развијању отпорних поступака који ће уклонити што више штампарских и шумних ограничења а задржати тачност приближну комерцијалним системима са ограниченим бројем фонтова.

Приступ препознавању анализом слике документа

Знакови су поредани у редове и задовољавају правила штампе која се могу употребити да се нађу знакови и одреди њихов начин штампе. Правила штампе омогућују разликовање сличних знакова различите величине или висине положаја у реду. Она такође ограничавају област претраге на основу међусобних разлика у положају и висини знакова у односу на основну линију реда.

Одређивање размјештаја

Одређивање положаја знакова у документу се врши након анализе размјештаја слике документа. Анализа размјештаја се врши помоћу неколико операција као што су одређивање нагиба, раздвајање слика од текста и подјела текста у колоне, редове, ријечи и повезане саставне дијелове. Текстуални дио се даље сегментира. Касније ће бити дат преглед техника сегментације.

Изградња знакова

При изградњи слика знакова често се суочавамо са међусобно повезаним или прекинутим знаковима на које наилазимо у оштећеним документима. Развој техника за правилно расчлањење ријечи у слова је још увијек изазов. Међу познатим техникама описаћемо следеће:

- Израчуна се вертикална пројекција тачака у реду. Затим се претражују различите подјеле и спајања симбола алгоритмом „branch-and-bound“ ограниченим оцјенама поузданости ријечи изведених из поузданости симбола.

- Израчуна се вертикална пројекција тачака у реду. Након тога се користи структура података „дрво одлуке“ ради разрјешавања нејасноћа.
- Изврши се рекурзивно расчлањење. Затим се употребљавају контекстне информације и врши провјера правописа ради исправљања грешака насталих усљед нетачног расчлањења.
- Направе се хипотезе за спајање и раздвајање знакова. Затим неколико експертских система провјерава хипотезе да би се видјело да ли представљају важећи симбол. Претрагу надгледа „А*“ алгоритам, разрјешавајући повратну обраду. Експертски системи се састоје из дијела за препознавање симбола и скупа алгоритама за обраду контекста.

Утицај фонтова

Читалац мора изаћи на крај с многим узроцима промјењивости у документу, од којих су најбитнији различити фонтови и величине знакова. У почетку развоја технологије оптичког препознавања знакова су за потребе појединих институција кориштени посебни фонтови (нпр. OCR-A и OCR-B) утврђене величине, што је и поред неразвијене технологије омогућавало високу тачност, међутим, мала промјена у величини или фонту је јако смањивала тачност оваквих система јер су кориштени прости алгоритми.

У данашње вријеме се користе вишефонтовски системи утемељени на особинама штампе. Информације о фонту су садржане у његовим знаковима, а поступци засновани на особинама су неосјетљивији према промјени фонта. Постоје два главна правца развоја оваквих система. Први омогућава кориснику да научи систем било ком фонту или скупу фонтова, након чега је могуће распознавање знакова из скупа научених фонтова. Други за циљ има потпуну независност од фонта.

Помоћни поступци при препознавању знакова

Одређивање особина знакова

Овај корак је неопходан у контексту анализе документа гдје бројни узроци, као што су: геометријске трансформације због ниског квалитета података, нагнутост и промјена дебљине знакова због промјене фонта узрокују различита одступања у односу на основни изглед знакова. Из тог разлога се траже особине које се не

мијењају са промјеном фонта. То омогућава да за сваки симбол постоји само један образац (овде треба обратити пажњу да за једно слово може постојати више различитих симбола, нпр. за ћирилично велико л (Л) или (Λ)). Такође се примјењују трансформације које врше нормализацију ради смањења појединих добро проучених одступања у изгледу. Преостала одступања се уче статистичким прилагођавањем.

Учење симбола

Најбитније ствари при учењу штампаних знакова су скуп знакова из ког се учи и прилагодба система за препознавање новим знаковима и фонтовима. Скуп може бити задан од стране корисника или створен из примјерака докумената. У првом случају корисник задаје симболе и њима одговарајуће кодове и евентуално усмјерава систем при стварању образаца а у другом се симболи налазе из докумената окупљањем сличних и налажењем „симбола нагомилавања“ за сваки скуп. То су симболи којима су симболи из њима одговарајућих скупова јако слични (сличност је приближно иста као и међусобна сличност симбола из једног скупа), а они се међусобно јако разликују. Проблем с овим приступом је што није сигурно да се сви потребни симболи налазе у узорцима докумената.

Контекстна обрада

Сврха контекстне обраде је исправљање грешака направљених због одлучивању на основу локалних особина посматрањем односа међу знаковима у ријечи. Већина техника комбинује геометријске и језичке информације. Описаћемо једну од техника: користе се скривени Марковљеви модели за моделисање знакова и ријечи. Знакови се спајају у низове за које се провјерава да ли представљају ријечи из ријечника користећи „Ratcliff/Obershelp“ метод налажења пандана обрасцу низа. Уколико се не нађе прихватљива ријеч списак несигурних знакова се пропушта кроз Витербијеву мрежу и узима се највјероватнија ријеч с излаза. При овом се користе вјероватноће двознака и мјеста знака добијене статистичком обрадом циљног ријечника. У примјеру рађеном за Француски језик користећи ријечник од 190000 ријечи кориштењем овог метода добијена је тачност препознавања ријечи преко 98%.

Очекивања у области оптичког препознавања знакова

Оптичко препознавање знакова је суштински дио домена анализе докумената. Препознавање знакова у окружењу више фонтова се не може поуздано извести без закључака о размјештају текста и фонтова у тексту. Такође, да би се добио систем употребљиве тачности неопходан је ослонац на језички контекст. Само комбинација више различитих техника оптичког препознавања знакова омогућава високу тачност препознавања.

Технике сегментације представе свјетлине

Основе сегментације свјетлине

Сегментација свјетлине је претварање представе¹ свјетлине (gray-level image) у двосвјетлинску представу (bilevel, monochrome, black and white image). Двосвјетлинска представа садржи много мање информација али треба да задржи све битне информације о броју, мјесту и облику објеката. Основни разлог разврставања тачака по свјетлини је што блиске тачке приближне свјетлине обично припадају истом објекту, а смањење сложености података поједностављује многе поступке препознавања и разврставања. Сегментација је неопходна прије поступака стањивања, векторизације и морфолошких поступака.

Најједноставнији начин претварања представе свјетлине у двосвјетлинску представу је избором јединствене граничне свјетлине. Све тамније тачке се сврставају у црне а свјетлије у бијеле. У том случају проблем разграничења свјетлина се своди на избор прага свјетлине T . Да би овај поступак био дјелотворан свјетлине тачака у представи I морају припадати једној од двије могуће врсте, црној или бијелој.

За црне тачке вриједи:

$$I(i, j) < T,$$

док су бијеле оне за које је:

$$I(i, j) \geq T.$$

¹ Преводимо ријеч image ријечју представа а не слика јер ријеч представа означава информациони објект док ријеч слика означава материјални објект који садржи информациони објект представе у облику просторног и/или временског распореда молекула или физичких поља у основном медиуму (нпр. молекула у фотопапиру, електро-магнетних поља у полупроводничкој меморији, тврдом диску или простору).

Ова претпоставка не вриједи у свим случајевима, јер се у стварним примјенама јављају шум и неједнако освијетљење. Због овога се у општем случају не може употријебити јединствен праг, али се може употријебити у довољном броју случајева да би се употријебио као почетна претпоставка. На примјер, документи скенирани било којим данашњим просјечним скенером могу бити сегментирани јединственим прагом. Праг се одређује на основу свјетлина тачака у представи. Врши се одређивање вриједности одређеног скупа особина представе и на основу њих и познатих особина дате врсте представа се израчунава праг. Најједноставнији начин одређивања прага је рачунањем средње свјетлине тачака у слици, међутим овај начин даје добре резултате једино код слика код којих је приближно једнак број тамних и свијетлих тачака. У општем случају не може се очекивати да је удио тамних тачака сталан.

Ипак, постоје врсте представа код којих је удио тамних тачака приближно сталан, тако да је могуће вршити сегментацију подешавајући праг тако да одговарајући дио тачака постане црн. Овакав праг се може одредити на једноставан начин користећи хистограм свјетлина у представи. Креће се од најтамнијег дијела хистограма, вршећи сабирање броја тачака одговарајуће свјетлине све док однос збира према укупном броју тачака не буде приближно једнак очекиваном удјелу црних.

Кориштење хистограма при одређивању прага присутно је код многих метода сегментације. У случајевима када је праг очигледан он се налази у области малих вриједности хистограма између двије области великих вриједности. У овом случају се одређивање прага своди на два корака: одређивања врхова области великих вриједности и одређивања дна области малих вриједности између њих. Налажење првог врха је једноставно. То је врх са највећом вриједношћу у хистограму. Међутим, следећа вриједност је најчешће у тачки поред прве, што није врх друге области. Једноставан начин којим се често успијева превазићи ова тешкоћа је множење хистограма квадратом удаљености од првог врха и налажење врха добијеног хистограма. Оваква сегментација се назива сегментација између два врха хистограма.

Бољи начин за налажење врхова, с обзиром да до растурања свјетлина у одговарајућој врсти представа од двије основне (које представљају бијелу и црну) долази због шума, био би апроксимација хистограма збиром двије Гаусове криве.

Међутим, пошто нам није позната њихова стандардна девијација, овакав поступак би био изузетно рачунски захтјеван.

Сегментација кориштењем ивичних тачака

Ивичне тачке се налазе на граници објекта и позадине или на граници два објекта. Због тога су свјетлине ивичних тачака много мање промјењиве од свјетлина тачака у слици. Њихов хистограм је много равнији од укупног хистограма. На основу овог је сачињен поступак за налажење прага употребом дигиталног лапласијана, који је неусмјерени оператор за одређивање ивица. Први корак при одређивању прага је рачунање лапласијана улазне представе. Један од начина да се то обави је конволуцијом представе следећом маском.

$$\begin{array}{ccc} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{array}$$

Након овог се налази хистограм само оних тачака у којима је лапласијан велик, нпр. оних чија је вриједност изнад 85% осталих вриједности и на основу њега одређује праг.

Поступак итеративног бирања

Итеративно бирање је поступак код ког се почетна претпоставка свјетлине прага побољшава током вишеструких пролазака кроз представу. При том се не користи хистограм већ се при сваком проласку тачке дијеле на тачке објекта и тачке позадине, користећи свјетлине оба дијела за побољшање прага.

За почетну претпоставку прага се узима средња свјетлина слике. Затим се користећи овај праг одређују бијеле и црне области и њихове статистичке особине (у овом случају само средње свјетлине). Средња свјетлина црних области се означава са T_b , а средња свјетлина бијелих области са T_o . Следећа претпоставка прага је $(T_b+T_o)/2$, тј. средња вриједност просјечне свјетлине црних и просјечне свјетлине бијелих области. Даље се поступак понавља користећи у сваком следећем кораку праг одређен у претходном кораку. Када се добије праг једнак прагу добијеном у непосредно претходном кораку поступак се прекида.

Овај поступак је намијењен склоповском извршавању, при ком се у првом кораку сматра да четири угла слике припадају областима позадине, а остатак

областима објеката. У случају програмског извршавања могуће је израчунати исти праг користећи хистограм. Програмско остварење се знатно убрзава рачунањем преко хистограма умјесто претходно описаним поступком јер је хистограм једнодимензионални низ мале и непромјенљиве дужине. Крећући од почетне претпоставке прага T_0 , k -та претпоставка прага се добија као:

$$T_k = \frac{\sum_{i=0}^{T_{k-1}} i \cdot h(i)}{2 \cdot \sum_{i=0}^{T_{k-1}} h(i)} + \frac{\sum_{i=T_{k-1}+1}^N i \cdot h(i)}{2 \cdot \sum_{j=T_{k-1}+1}^N h(i)}$$

гдје је h хистограм свјетлина у представи. И овај поступак завршава када се добије праг једнак прагу добијеном у непосредно претходном кораку.

Поступак анализе варијансе хистограма свјетлина

Поступци засновани на проналажењу дна између два врха у хистограму крећу од претпоставке да свјетлине објеката и свјетлине позадине имају нормалну расподелу са различитим средњим вриједностима и дисперзијама (варијансама).

Укупна варијанса свјетлина σ^2 у представи се рачуна или директно из представе или из њеног хистограма. За сваки праг T могу се посебно израчунати варијанса свјетлина објеката и варијанса свјетлина позадине. Њих означавамо са σ_{w1}^2 и σ_{w2}^2 и називамо унутарврским варијансама. Варијансу просјечне свјетлине објеката и позадине у односу на просјечну свјетлину читаве представе означавамо са σ_b^2 и називамо међуврсном варијансом. Најбољи праг код поступка анализе варијансе је праг који даје најмањи однос међуврсне варијансе према укупној варијанси. Овај однос се дефинише са:

$$\eta(T) = \frac{\sigma_b^2(T)}{\sigma^2}.$$

Просјечна свјетлина укупне представе се означава са μ . Међуврсна варијанса се рачуна као:

$$\sigma_b^2(T) = \omega_0 \omega_1 (\mu_0 \mu_1)^2,$$

гдје је:

$$\omega_0 = \sum_{i=0}^T p_i, \quad \omega_1 = 1 - \omega_0,$$

а p_i је вјероватноћа свјетлине i , тј. вриједност хистограма у тачки i подијељена укупним бројем тачака. Такође је :

$$\mu_0 = \frac{\mu_T}{\omega_0}, \quad \mu_1 = \frac{\mu - \mu_T}{1 - \omega_0}, \quad \mu_T = \sum_{i=0}^T i \cdot p(i),$$

Све ове вриједности се рачунају из хистограма представе, а затим се рачуна $\eta(T)$ и тако за све могуће вриједности T па се узима она вриједност T која даје најмање $\eta(T)$ као коначни праг.

Сегментација употребом ентропије

Овде се говори о ентропији у смислу теорије информација. Нека извор симбола X даје на свом излазу симболе x_i из скупа од N могућих са вјероватноћама p_i . Тада је ентропија извора X дата са:

$$H(X) = -\sum_{i=1}^N p_i \text{ld}(p_i)$$

и изражава се у битима по симболу или Шенонима по симболу.

Представа се може посматрати као извор симбола (нивоа свјетлине). Ентропија црних тачака сегментираних прагом T је:

$$H_b = -\sum_{i=0}^T p_i \text{ld}(p_i)$$

гдје је p_i вјероватноћа нивоа i , док је ентропија бијелих тачака дата са:

$$H_w = -\sum_{i=T+1}^{N-1} p_i \text{ld}(p_i)$$

гдје је N број различитих нивоа свјетлине. Поступак се састоји у налажењу оне вриједности прага T за коју је укупна ентропија $H = H_b + H_w$ највећа, што је исто што и тражење T за које је израз:

$$f(T) = \frac{H_T}{H} \frac{\text{ld} P_T}{\text{ld}(\max_{i=1,T} \{p_i\})} + \left[1 - \frac{H_T}{H} \right] \frac{\text{ld}(1 - P_T)}{\text{ld}(\max_{i=T+1, N-1} \{p_i\})}$$

највећи, при чему је:

$$H_T = -\sum_{i=0}^T p_i \text{ld} p_i$$

ентропија црних тачака на основу прага T ,

$$H = -\sum_{i=0}^{N-1} p_i \text{ld} p_i$$

је укупна ентропија, а

$$P_T = \sum_{i=0}^T p_i$$

је вјероватноћа за било коју тачку у представи да ће бити црна. Величине зависне од T се могу израчунати из хистограма, а H је независно од T и рачуна се на почетку, на основу читавог хистограма.

Други начин (Капуров поступак) сегментације употребом ентропије прво дефинише расподелу вјероватноћа свјетлина објеката $A(i)$ и свјетлина позадине $B(i)$ на следећи начин:

$$A(i) = \frac{p_i}{P_T}, i = \overline{0, T}$$

$$B(i) = \frac{p_i}{1 - P_T}, i = \overline{T+1, N-1}$$

Из чега се ентропија црних и бијелих тачака рачуна на основу одговарајућих формула, употребљавајући ове расподеле вјероватноћа:

$$H_b = - \sum_{i=0}^T \frac{p_i}{P_T} \text{ld} \left(\frac{p_i}{P_T} \right)$$

$$H_w = - \sum \frac{p_i}{1 - P_T} \text{ld} \left(\frac{p_i}{1 - P_T} \right)$$

И овде је најбољи праг онај за који је збир ове двије ентропије највећи. У овом случају пробавамо све могуће прагове да би смо нашли најбољи.

Трећи начин (Јохансенев поступак) је да се свјетлине подијеле у два скупа тако да је њихова међусобна зависност најмања. Ово се своди на налажење прага T за који је збир $S_b(T) + S_w(T)$ најмањи, гдје су:

$$S_b(T) = \text{ld} \left(\sum_{i=0}^T p_i \right) + \frac{1}{\sum_{i=0}^T p_i} \left[E(P_T) + E \left(\sum_{i=0}^{T-1} p_i \right) \right]$$

и

$$S_w(T) = \text{ld} \left(\sum_{i=T}^{N-1} p_i \right) + \frac{1}{\sum_{i=T}^{N-1} p_i} \left[E(P_T) + E \left(\sum_{i=T+1}^{N-1} p_i \right) \right]$$

и гдје је

$$E(x) = -x \cdot \text{ld}(x)$$

ентропијска функција. При овом треба обратити пажњу да се прагови за које је $P_T = 0$ не узимају у обзир.

Сегментација помоћу фази скупова

Код фази² скупа припадност скупу је одређена мјером или вјероватноћом припадности. Кажемо да је вјероватноћа да елемент x припада скупу S једнака u_x . При сегментацији представе сврставамо тачке или у скуп тачака позадине или у скуп тачака објеката, па је очигледна могућност примјене фази скупова.

Постоји мноштво поступака сегментације изведених на основу фази скупова. Један од њих ради на следећи начин: дефинише се неодређеност као мјера удаљености између представе свјетлина и двосвјетлинске представе. Затим се одређује праг за који је неодређеност најмања.

Неодређеност се дефинише преко функције припадности, која одређује вјероватноћу сврставања поједине тачке као тачке објекта или позадине. Нека је просјечна свјетлина позадине μ_0 , а објеката μ_1 . Што је мања разлика свјетлине тачке x и просјечне свјетлине скупа у који је сврстана то је већа вриједност функције припадности $u_x(x)$. Одговарајућа функција припадности се може дефинисати нпр. као:

$$u_x(x) = \begin{cases} \frac{1}{1+|x-\mu_0|/C}; x \leq T \\ \frac{1}{1+|1-\mu_1|/C}; x > T \end{cases}$$

за дати праг T . Овде је C стална вриједност и представља разлику између највеће и најмање свјетлине у представи. Степен у ком нека тачка припада скупу тачака објеката или тачака позадине, зависно од односа њене свјетлине према прагу, је дат функцијом $u_x(x)$ која даје вриједности између $1/2$ и 1 .

Постоји више начина да се одреди неодређеност на основу функције припадности свјетлина тачака за дати праг који испуњавају захтјев да неодређеност треба да узима вриједности између 0 и 1 . Један од начина Ентропијски начин) је заснован на ентропији фази скупа, која се израчунава употребом Шенонове функције дате следећим изразом:

² fuzzy – нејасан, помућен, неразаберив; fuzzy sets – неразабериви или неразмеђиви скупови (јер се не може са сигурношћу рећи коме од њих елемент припада)
fuzziness – нејасност, помућеност, несигурност, неразаберивост

$$H_f(x) = -x \cdot \text{ld}(x) - (1-x) \cdot \text{ld}(1-x),$$

тако да је ентропија читавог фази скупа представе дата са:

$$E(T) = \frac{1}{M \cdot N} \sum_{g \in G} H_f(\mu_x(g)) \cdot h(g)$$

гдје је G скуп свих свјетлина, N и M бројеви редова и колона, а h хистограм свјетлина. Вриједност прага T за коју је $E(T)$ најмање се узима за коначну вриједност прага.

Други начин (Јагеров начин) за дефинисање мјере неодређености је заснован на идеји да за разлику од нормалних скупова код којих елемент припада или скупу A или његовом комплементу код фази скупова елемент припада и једном и другом с одређеним вјероватноћама.

Степен у којем су A и његов комплемент неразлучиви представља мјеру његове неодређености. Ово се може израчунати на следећи начин:

$$D_p(T) = \left[\sum_{g \in G} |\mu_x(g) - \mu_x^-(g)|^p \right]^{\frac{1}{p}}$$

гдје је p цио број већи или једнак 1 и $\mu_x^-(g)$ представља $\mu_x(g)$ рачунато по формули за свјетлине са супротне стране прага.

Вриједност p одређује врсту даљине. За $p = 2$ добијамо еуклидску даљину. За p се могу узимати и нецијели бројеви већи од 1, али то само успорава рачун а не даје битно различит исход од од оног добијеног употребом најближег цијелог броја.

На који год начин да одређујемо неодређеност морамо прво одредити просјечне свјетлине објеката и позадине за праг T . Ове вриједности су дате следећим формулама:

$$\mu_0(T) = \frac{\sum_{g=0}^T g \cdot h(g)}{\sum_{g=0}^T h(g)}$$

за позадину (ако је позадина тамна а објекти свијетли, иначе је обрнуто) и:

$$\mu_1(T) = \frac{\sum_{g=T+1}^{N-1} g \cdot h(g)}{\sum_{g=T+1}^{N-1} h(g)}$$

за објекте. Праг се налази испробавањем свих могућих вриједности за T .

Сегментација најмањим одступањем

Хистограм представе се може посматрати као измјерена функција расподеле вјероватноћа свјетлина у представи помножена бројем тачака у представи. Ако сматрамо да сви објекти требају бити једне свјетлине, а позадина друге онда се одступања од тих свјетлина могу посматрати као шум, па је ову функцију расподеле вјероватноћа могуће посматрати као збир двије функције расподеле вјероватноћа, једне за објекте а друге за позадину.

Пошто се обично узима да шум има нормалну расподелу, хистограм се може приближно представити као:

$$p(g) = \frac{A_1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(g-\mu_1)^2}{2\sigma_1^2}} + \frac{A_2}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(g-\mu_2)^2}{2\sigma_2^2}}$$

гдје су: A_1 и A_2 приближно бројеви тамних и свијетлих тачака, σ_1 , σ_2 , μ_1 и μ_2 стандардне девијације и средње вриједности тамних и свијетлих тачака. На основу овог је могуће добити непознате величине поступком најмање средњеквадратне грешке уз помоћ неког од поступака за нумеричко рјешавање нелинеарних једначина или неуронском мрежом. При овоме треба обратити пажњу на утицај ограниченог опсега свјетлина у представи због чега долази до одсијецања дијелова расподеле свјетлина изван опсега и пресликавања у граничне вриједности свјетлина, због чега је потребно прије поступка за одређивање непознатих извршити одсијецање хистограма испод најмање свјетлине која постоји у представи, укључујући и ту свјетлину и изнад највеће, укључујући и њу.

С обзиром да је претходни поступак сложен предложено је умјесто тога налажење најмање вриједности функције:

$$J(T) = 1 + 2(P_1(T) \ln \sigma_1(T) + P_2(T) \ln \sigma_2(T)) - 2(P_1(T) \ln P_1(T) + P_2(T) \ln P_2(T))$$

гдје су:

$$P_1(T) = \sum_{g=0}^T h(g), \quad P_2(T) = \sum_{g=T+1}^{N-1} h(g),$$

$$\mu_1(T) = \frac{\sum_{g=0}^T g \cdot h(g)}{P_1(T)}, \quad \mu_2(T) = \frac{\sum_{g=T+1}^{N-1} g \cdot h(g)}{P_2(T)},$$

$$\sigma_1^2(T) = \frac{\sum_{g=0}^T h(g) \cdot (g - \mu_1(T))^2}{P_1(T)}$$

и

$$\sigma_2^2(T) = \frac{\sum_{g=T+1}^{N-1} h(g) \cdot (g - \mu_2(T))^2}{P_2(T)}.$$

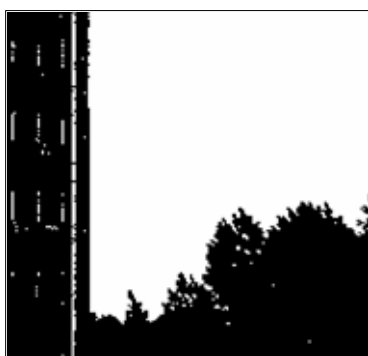
За праг се узима она вриједност T за коју је функција $J(T)$ најмања.

Поређење поступака сегментације једним прагом

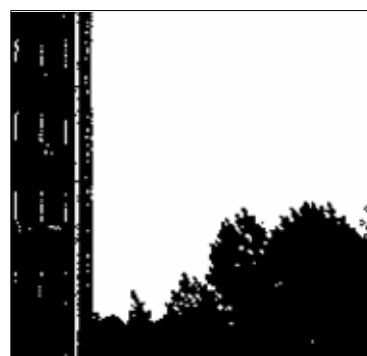
Описани поступци су примијењени на три слике. Испод сваке слике је исписан назив поступка на основу којег је добијена.



Оригинална слика – Небо



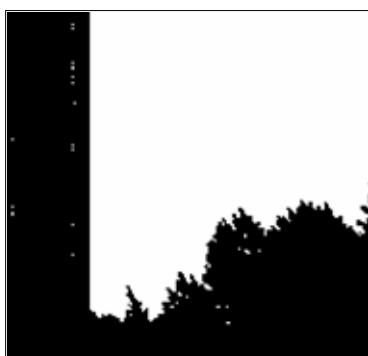
Анализа варијансе



Итеративно бирање



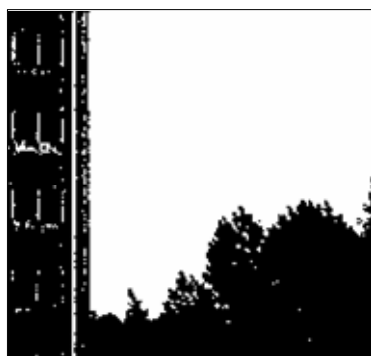
Између два врха



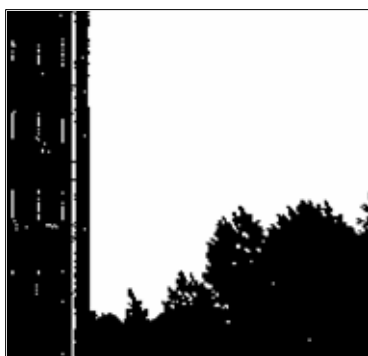
Најмање одступање



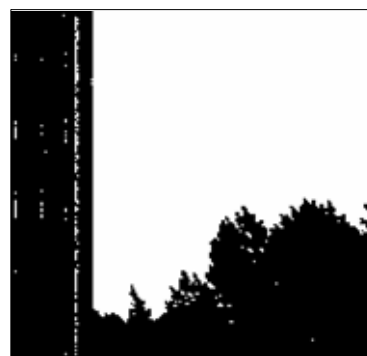
По постотку (10%)



Фази скупови – ентропија



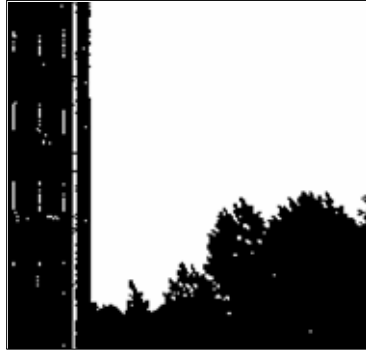
Фази скупови – Јагер



Средња вриједност



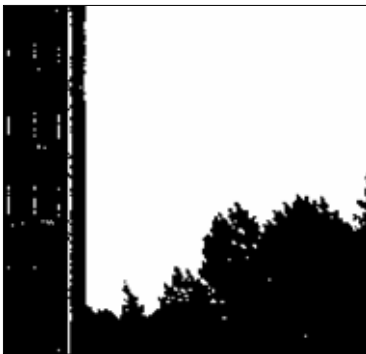
Свјетлином ивичних тачака (10%)



Ентропијом, Јохансенов поступак



Ентропијом, Капуров поступак



Ентропијом, Пунов поступак

```
program test Liaput, o
var i, j: integer;
    xy1, xz2: real;
begin
  for i:= 1 to 10 d
  begin
    j:= j+i;
    xy1:= xy1 +
  end;
  writeln ('Result
```

Оригинална слика – Паскал

```
program test Liaput, o
var i, j: integer;
    xy1, xz2: real;
begin
  for i:= 1 to 10 d
  begin
    j:= j+i;
    xy1:= xy1 +
  end;
  writeln ('Result
```

Анализа варијансе

```
program test Liaput, o
var i, j: integer;
    xy1, xz2: real;
begin
  for i:= 1 to 10 d
  begin
    j:= j+i;
    xy1:= xy1 +
  end;
  writeln ('Result
```

Итеративно бирање

```
program test Liaput, o
var i, j: integer;
    xy1, xz2: real;
begin
  for i:= 1 to 10 d
  begin
    j:= j+i;
    xy1:= xy1 +
  end;
  writeln ('Result
```

Између два врха

```
program test Liaput, o
var i, j: integer;
    xy1, xz2: real;
begin
  for i:= 1 to 10 d
  begin
    j:= j+i;
    xy1:= xy1 +
  end;
  writeln ('Result
```

Најмање одступање

```
program test Liaput, o
var i, j: integer;
    xy1, xz2: real;
begin
  for i:= 1 to 10 d
  begin
    j:= j+i;
    xy1:= xy1 +
  end;
  writeln ('Result
```

По постотку (10%)

```

program test Linput, 0
var i, j: integer;
    xyl, xz2: real;
begin
  for i:= 1 to 10 do
    begin
      j:= j+i;
      xyl:= xyl + 1;
    end;
  writeln ('Result

```

Фази скупови – ентропија

```

program test Linput, 0
var i, j: integer;
    xyl, xz2: real;
begin
  for i:= 1 to 10 do
    begin
      j:= j+i;
      xyl:= xyl + 1;
    end;
  writeln ('Result

```

Фази скупови – Јагер

```

program test Linput, 0
var i, j: integer;
    xyl, xz2: real;
begin
  for i:= 1 to 10 do
    begin
      j:= j+i;
      xyl:= xyl + 1;
    end;
  writeln ('Result

```

Средња вриједност

```

program test Linput, 0
var i, j: integer;
    xyl, xz2: real;
begin
  for i:= 1 to 10 do
    begin
      j:= j+i;
      xyl:= xyl + 1;
    end;
  writeln ('Result

```

Свјетлином ивичних тачака (10%)

```

program test Linput, 0
var i, j: integer;
    xyl, xz2: real;
begin
  for i:= 1 to 10 do
    begin
      j:= j+i;
      xyl:= xyl + 1;
    end;
  writeln ('Result

```

Ентропијом, Јохансенов поступак

```

program test Linput, 0
var i, j: integer;
    xyl, xz2: real;
begin
  for i:= 1 to 10 do
    begin
      j:= j+i;
      xyl:= xyl + 1;
    end;
  writeln ('Result

```

Ентропијом, Капуров поступак

```

program test Linput, 0
var i, j: integer;
    xyl, xz2: real;
begin
  for i:= 1 to 10 do
    begin
      j:= j+i;
      xyl:= xyl + 1;
    end;
  writeln ('Result

```

Ентропијом, Пунов поступак



Оригинална слика – Лице



Анализа варијансе



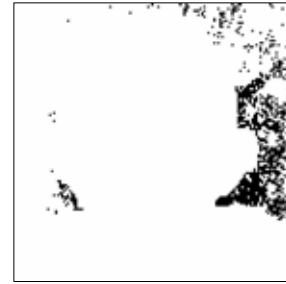
Итеративно бирање



Између два врха



Најмање одступање



По постотку (10%)



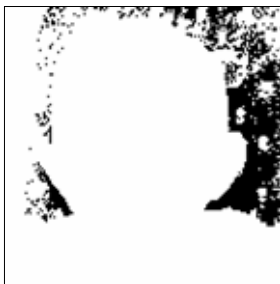
Фази скупови – ентропија



Фази скупови – Јагер



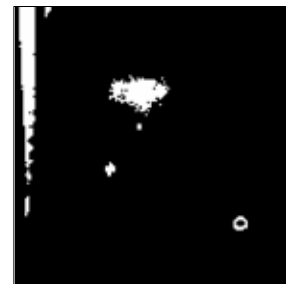
Средња вриједност



Свјетлином ивичних тачака (10%)



Ентропијом, Јохансенов поступак



Ентропијом, Капуров поступак



Ентропијом, Пунов поступак

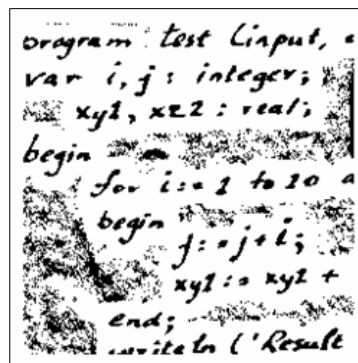
Употреба обласних прагова

Код претходно наведених поступака се подразумијевало да се свјетлине објекта и свјетлине позадине не преклапају ни једним дијелом. Међутим, код великог броја представа се јавља неједнакомјерно освјетљење услед чега је то истина само у оквиру довољно малих области, гдје је освјетљење приближно стално, али не и у читавој представи. Да би се извршила сегментација овакве

представе потребно је за сваку од тих области посебно одредити њен праг и сегментирати је помоћу њега.

При овом је најзначајније питање одређивања броја области и њихових величина. Када се то уради онда се може употријебити било који од претходно наведених поступака за сегментацију појединачних области. Могуће је једноставно задати број области, па ако је тај број довољно велик области ће бити довољно мале да би се у оквиру њих могао примијенити неки од претходних поступака, што је понекад брже од употребе неког компликованог поступка за одређивање области, иако такав поступак може дати мањи број области.

Показаћемо како ово ради на примјеру. Сегментираћемо представу користећи 21 x 21 област, које се међусобно преклапају, поступком итеративног бирања. Исход оваквог поступка на представе неба и паскал програма је приказан овде:



Исход сегментације поступком обласног итеративног бирања на представе Небо и Паскал

Исход је слабији од очекиваног јер овај поступак настоји да у свакој области нађе и тачке објаката и тачке позадине, тако да у областима гдје су све тачке једне врсте проглашава шум за тачке друге врсте, из чега се види да код употребе обласних прагова треба прво провјерити постоје ли двије врсте тачака, па ако не постоје не вршити сегментацију већ читаву област прогласити оном врстом тачака која је сусједна овој области.

Поступак Чоуа и Канека

Овај поступак дијели представу у 49 преклопљених области величине 64 x 64 тачке. Ово је подјела за представе величине 256 x 256, међутим подјелу је могуће извршити и на неки други начин, посебно уколико је представа различите величине. Налази се хистограм сваке области и врши провјера двоврсности хистограма. Пошто двоврстан хистограм има двије врсте свјетлина праг се налази између њихових

врхова. За сваки такав хистограм се налазе двије одговарајуће Гаусове криве поступком најмањих квадрата. Овако се тачније одређују врхови што омогућује избор најповољнијег прага за сваку област.

Прагови области са једноврстним хистограмом се процјењују из прагова сусједних области двоврстним хистограмом. На крају се за сваку тачку појединачно врши процјена прага на основу прагова сусједних области.

Двоврстни хистограм се може представити као збир двије Гаусове функције. При том је потребно наћи средњу вриједност, стандардну девијацију и множилац величине за обије Гаусове функције, што се врши поступком најмањих квадрата. Код сваке области прво налазимо хистограм који затим умекшавамо на основу следећег израза:

$$F_s(i) = \frac{F(i-2) + 2F(i-1) + 3F(i) + 2F(i+1) + F(i+2)}{9}$$

Умекшавањем се смањује утицај шума. Након овог се хистограм дијели на два дијела по најмањој умекшаној тачки, чију свјетлину са v . Сада сматрамо да је једна Гаусова функција лијево а друга десно од ње. Почетне вриједности њихових параметара процјењујемо из одговарајућих дијелова хистограма. Процјене параметара су:

$$N_1 = \sum_{i=0}^v F(i), \quad N_2 = \sum_{i=v+1}^{N-1} F(i)$$

$$\mu_1 = \sum_{i=0}^v F(i) \cdot i, \quad \mu_2 = \sum_{i=v+1}^{N-1} F(i) \cdot i$$

$$\sigma_1 = \sqrt{\frac{1}{N_1} \sum_{i=0}^v F(i) \cdot (i - \mu_1)^2}, \quad \sigma_2 = \sqrt{\frac{1}{N_2} \sum_{i=v+1}^{N-1} F(i) \cdot (i - \mu_2)^2}$$

$$P_1 = \frac{\sigma_1 N_1}{\sum_{i=0}^v e^{-\frac{(i-\mu_1)^2}{2\sigma_1^2}}}, \quad P_2 = \frac{\sigma_2 N_2}{\sum_{i=v+1}^{N-1} e^{-\frac{(i-\mu_2)^2}{2\sigma_2^2}}}$$

Овде су Гаусове функције дефинисане као: $G(x) = \frac{P}{\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Коначне вриједности параметара се добијају налазећи најмању вриједност израза:

$$R(P_1, \mu_1, \sigma_1, P_2, \mu_2, \sigma_2) = \sum_{i=0}^{N-1} (G_1(i) + G_2(i) - F(i))^2.$$

Мјерила двоврстности хистограма су следећа: средње вриједности се морају разликовати бар за $N/64$, однос стандардних девијација мора бити близак јединици (узима се $0,5 < \sigma_1/\sigma_2 < 2,0$) и однос дна хистограма и мање од вриједности хистограма у средњим вриједностима расподјела мора бити мањи од 0,8.

Уколико хистограм није двоврстан даљи поступак у тренутној области се прекида, а иначе се одређује праг рјешавањем квадратне једначине:

$$\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)t^2 + 2\left(\frac{\mu_2}{\sigma_2^2} - \frac{\mu_1}{\sigma_1^2}\right)t + 2\log\left(\frac{P_2\sigma_1}{P_1\sigma_2}\right) = 0,$$

при чему се узима оно рјешење које је између средњих вриједности. За области у којима нису одређени прагови они се одређују линеарном интерполацијом прагова сусједних области ако су код свих сусједних одређени. Након овог се врши конволуција прагова области следећом маском:

$$\begin{array}{ccc} \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} \\ 1 & 2 & 1 \\ \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} \end{array}$$

а затим се одређују прагови за сваку тачку појединачно билинеарном интерполацијом из прагова сусједних области, а ако праг неке од сусједних области није одређен онда се за праг тачке узима праг најближе области са одређеним прагом. Овако се одређују и прагови тачака на ивици представе.

Узорковање освијетљености помоћу ивица

Уколико је освијетљеност објекта позната онда је задатак сегментације тачака објекта много једноставнији. Свјетлина тачке у представи је пропорционална производу освијетљености и рефлексивности објекта у тој тачки. На основу познате промјене освијетљености објекта у представи се може израчунати и поништити њен утицај, па се може извршити сегментација једним прагом. Ово је у суштини једнако одређивању посебног прага за сваку тачку на основу обласних особина представе. Код поступка узорковања из ивица се одређују ивичне тачке, које се сматрају за сигурне тачке објекта, и на основу њих се одређује локална освијетљеност. Освијетљеност (и прагови) осталих тачака се процјењују из освијетљености ивичних тачака, сматрајући процјену непоузданиом с порастом удаљеност. Тачке ивица се

одређују Шен-Кастановим поступком, затим се провлачи површ кроз свјетлине ивица па на основу ње се процјењује освијетљеност и у осталим тачкама. Површ се провлачи кроз свјетлине ивица поступком помичних најмањих квадрата. При овом поступку се у свакој тачки рјешава проблем најмањих квадрата са тежинским множитељима. Нека је:

$$J(x, y) = \sum_{i=1}^N w_i(x, y)(I(x_i, y_i) - S(x_i, y_i))$$

функција одступања, гдје је N број ивичних тачака, $I(x_i, y_i)$ свјетлине у њима, $S(x, y) = ax + by + c$ раван свјетлине кроз тачку (x, y) , $w_i(x, y)$ тежине обрнуто зависне од удаљености тачке (x_i, y_i) од тачке (x, y) , при чему ова зависност не мора бити линеарна. Обично се узима да је $w_i(x, y)$ једнако нули изнад неке удаљености d , што омогућава да се утицај даљих тачака не узима у обзир због чега се убрзава поступак. Недостатак сталне вриједности d је што се може десити да се на мањој даљини не налазе бар три тачке ивице, што онемогућава извршење поступка, а може се десити и да се на мањој даљини нађе јако много ивичних тачака што успорава поступак. Мана овог поступка је што је јако спор.

Релаксациони поступци

Релаксација је итеративни низ поступака. При одређивању прагова, прагови у било којој итерацији су функција прагова у истој околини у претходној итерацији. Поступак се састоји у следећем:

1. Сачини се почетна претпоставка сегментације и процјена поузданости те претпоставке у свакој тачки.
2. За сваку тачку измијени се сегментација и процјена поузданости на основу блиских тачака. Довољно је околних осам тачака.
3. Понавља се корак 2. док се сегментација не заврши. Обично је то када нема промјена између два сусједна корака.

Процјене поузданости имају облик вјероватноћа, иако не морају бити тачне, битно је да су приближно тачне у односу на остале тачке у околини.

Мана релаксационих поступака је што су јако спори, док им резултати нису на задовољавајућем нивоу, због чега се овде неће у танчине обрађивати.

Помични просједи

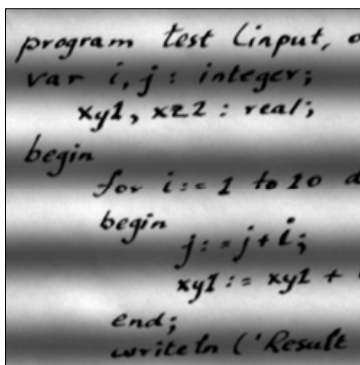
Овај поступак је сачињен у сврху сегментације текста. Помични просјек је средња свјетлина задњих n посматраних тачака. Представа се може посматрати као једнодимензионалан низ тачака, док просјек може да се израчуна тачно или процијени на основу:

$$M_{i+1} = M_i - \frac{M_i}{n} + g_{i+1}$$

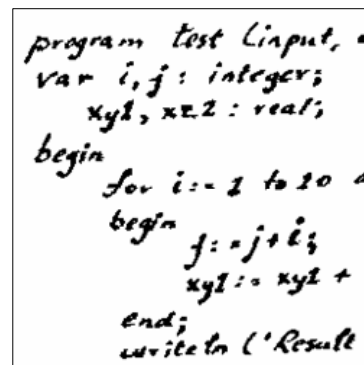
гдје је M_{i+1} процјена помичног просјека $i+1$ - ве тачке чија је свјетлина g_{i+1} , а M_i је помични просјек претходне тачке. Све вриједности овде су помножене са n . Свака тачка чија је свјетлина мања од унапријед заданог постотка њеног просјека се проглашава црном, иначе бијелом. Редови се прелазе у измјеничним смјеровима да би се спријечила склоност ка једној страни. За вриједност n се обично узима осмина дужине реда. Сегментирана слика се добија на основу:

$$V = \begin{cases} 0; & \text{за } n \cdot g_i < M_i \frac{100-p}{100} \\ 255; & \text{иначе} \end{cases}$$

Гдје је V сегментирана вриједност тачке а p задани постотак. Исход овог поступка на представу паскал програма на коју је додано синусно освјетљење је приказан на следећој страни:



Оригинална представа



Сегментирана представа

Одређивање нагнутости

Уколико се документ који се читава поставља пажљиво у скенер он неће бити нагнут више од 3 степена, међутим корисник који врши читавање обично није увијек бити тако пажљив, а велика нагнутост знакова смањује тачност препознавања код већине система за препознавање (не смањује код неуронских мрежа пројектованих тако да буду отпорне на нагнутост знакова). Да би се достигла

највећа тачност препознавања потребно је одредити угао нагнутости текста у представи и уврнути представу за тај угао у супротном смјеру.

Постоји мноштво поступака за одређивање нагнутости, од којих ће неки овде бити описани.

Поједини поступци се заснивају на следећој чињеници: код већине знакова су доње линије описаног правоугаоника колинеарне, тако да би средње тачке доње линије требале бити колинеарне за сваки ред текста. Из овог следи следећи поступак (Бејрдов поступак):

1. Одреди се све повезане области, претпостављајући да свака представља знак уколико не прелази задани праг величине.
2. Одреди се описани правоугаоник сваке области и тачка средине његове доње ивице.
3. За сваки претпостављени угао θ одреди се пројекције тачака одређених у кораку 2, што даје једнодимензионални низ пројекција $P(\theta)$. Бејрд је користио подиок величине $1/3$ величине знака од 6 типографских тачака у резолуцији представе. $P_i(\theta)$ је вриједност i -тог подиока одређеног за угао θ .
4. Одреди се θ за које је вриједност функције:

$$A(\theta) = \sum_{i=1}^n P_i^2(\theta)$$

највећа. Овај угао представља нагнутост представе.

Уколико је позната процјена угла вријеме претраге ће бити значајно смањено. Могуће је прво извршити грубу претрагу а затим фину да би се смањило вријеме претраживања.

Умјесто свега овог могуће је након корака 1. примијенити Хафов поступак. Он функционише на следећи начин. Свака права се може представити у облику:

$$p : \mathcal{L}(\vec{v}) + r \cdot \vec{v}_\perp,$$

гдје је \vec{v} било који јединични вектор, $\mathcal{L}(\cdot)$ ознака за линеал, r реалан број а \vec{v}_\perp јединични вектор окомит на \vec{v} . Јединични вектор \vec{v} се може претставити у облику:

$$\vec{v} = \begin{bmatrix} \cos(\varphi) \\ \sin(\varphi) \end{bmatrix},$$

док је угао њему окомитог јединичног вектора већи или мањи за $\pi/2$. Овде ћемо усвојити да је већи за $\pi/2$, тако да је окомити вектор дат са:

$$\vec{v}_\perp = \begin{bmatrix} \cos(\varphi + \pi/2) \\ \sin(\varphi + \pi/2) \end{bmatrix} = \begin{bmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{bmatrix}.$$

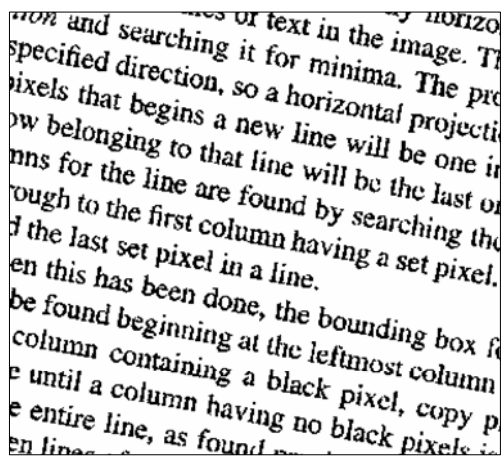
Сада узимамо средину представе за исходиште координатног система, одређујемо корак квантизације угла $\Delta\varphi$ и удаљености од средине Δr и посматрајући претставу као функцију двије промјењиве (вриједност функције у општем случају представља произвољан број придружен одређеној координати (x, y) , у специјалном случају Хафове трансформације тај број је један ако је тачка битна (тачке одређене у кораку 1. горе описаног поступка), иначе нула, а у случају Радонове трансформације тај број је вриједност свјетлине тачке на тој координати) рачунамо њену трансформацију (пројекцију) у Хафов простор (φ, r) .

Означимо средину претставе (исходиште) са (x_0, y_0) . За сваку тачку у којој је вриједност различита од нула усвајамо редом вриједности дозвољених углова $\varphi_1, \varphi_2, \dots, \varphi_q$; и рачунамо одговарајуће r_1, r_2, \dots, r_q ; притом заокружујући сваки r на најближи дозвољени, и вриједност у тачки додајемо претходној вриједности у тачки (φ_i, r_i) . Знајући координате тренутне тачке (x, y) , r_i рачунамо на следећи начин:

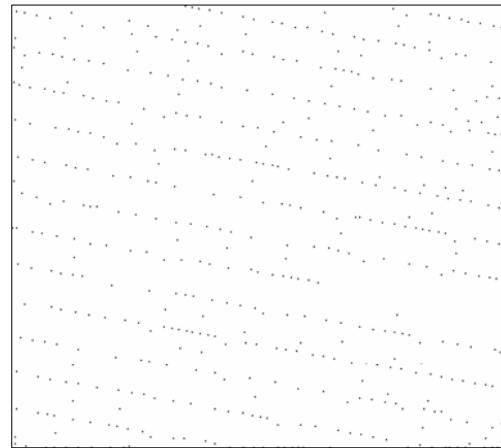
$$r_i = \lceil [(y - y_0) \cos(\varphi_i) - (x - x_0) \sin(\varphi_i)] / \Delta r \rceil \cdot \Delta r,$$

гдје је са $\lceil \cdot \rceil$ означена функција која врши заокруживање на најближи цијели број, а у случају једнаке удаљености од два цијела броја заокружује на парни. Након одређивања Хафове трансформације претражујемо Хафов простор тражећи тачку са највећом вриједношћу. Угао праве која је претстављена том тачком се узима за угао нагнутости текста у претстави.

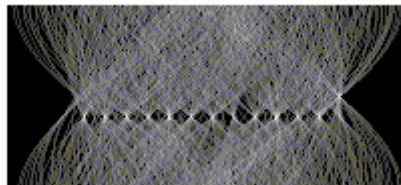
У прилогу су дати програм за припремни дио Бејрдовога поступка који је преузет из прилога литературе [1] и прилагођен компајлирању Мајкрософтовим компајлером „Visual C++ 2005“ и програм за Хафову трансформацију из истог извора, при чему је сама функција која обавља Хафову трансформацију изнова написана на основу горњих једначина. Поред тога је у нову функцију додана могућност задавања почетног угла, корака угла, броја углова и корака удаљености од исходишта. Прва слика на следећој страни је претстава текста нагнутог 10 степени, поред ње је слика добијена примјеном Бејрдовога поступка на њу а на слици испод њих је претстава Хафове трансформације десне слике.



Претстава текста нагнутог 10 степени



Претходна претстава након примјене
Бејрдовог поступка



Претстава Хафове трансформације
слике десно

Још једна могућност за одређивање угла нагнутости је да се сваком знаку одреде два најближа, затим се одреде њихове средине и међусобно се повежу линијом. Такве линије ће приближно одређивати праве под углом једнаким углу нагнутости текста.

Препознавање знакова

Након сегментације, процјене угла нагнутости, увртања оригиналне слике за супротни угао и сегментације добијене слике долази се до самог поступка препознавања. Прво се врши расчлањивање странице на редове текста, а затим се сваки ред расчлањује на појединачне знакове, који представљају повезане области објекта. Расчлањивање на редове се врши рачунањем хоризонталне пројекције (сабирање свјетлина свих тачака у једном реду претставе (ово не треба мијешати са редом текста)) и означавањем области без тачака објекта међупросторима, а сваке остале повезане области редом текста. Прије овог поступка потребно је одстранити изоловане тачке објекта из претставе. Након одређивања редова врши се раздвајање сваког појединог реда на знакове поступком вертикалне пројекције. При овом се може догодити да се више знакова међусобно додирују па да буду проглашени

једним знаком, због чега је потребно вршити раздвајање таквих знакова. Постоји више поступака за раздвајање:

1. Израчуна се вертикална пројекција. Затим се врше пробна раздвајања у тачкама гдје је вертикална пројекција најмања и препознавање тако добијених знакова. Затим се на основу вјероватноће тачности препознавања појединих знакова и процјеном вјероватноће одређеног низа знакова на основу ријечника закључује која ријеч је стварно била написана. Ово се често врши користећи поступак дрвета одлуке, помоћу којег се тражи највјероватнија комбинација препознатих знакова која претставља постојећу ријеч. Без употребе ријечника често долази до замјене при препознавању сличних знакова, чак и када су они раздвојени.
2. Често се умјесто вертикалне пројекције користи „цијена раздвајања“, па се даље врши поступак као у претходној тачки, употребљавајући „цијену раздвајања“ умјесто вертикалне пројекције. Цијена раздвајања у одређеној колони је једнака броју тачака објекта у датој колони којима је сусједна тачка у истом реду претходне колоне такође тачка објекта.
3. Могуће је мјеста пробног раздвајања одредити и на основу препознавања само лијевог дијела издвојеног знака, онолике величине колики је образац с којим упоређујемо. Уколико величина фонта није стална и једнака величини фонта на основу које су направљени обрасци знакова потребно је извршити одговарајућу промјену величине било обрасца било знака. При препознавању само лијевог дијела знака одређује се највећа висина знака до одређене удаљености од лијевог руба и на основу њиховог односа и односа одговарајућих димензија обрасца процјењује се на коју величину је потребно довести било образац било знак, у зависности од тога шта нам је промјењиво а шта не. Након препознавања лијевог дијела знака за пробна мјеста раздвајања се узимају ширине знакова чија је вјероватноћа велика. Затим се уклања подручје претпостављеног знака и врши препознавање лијевог дијела остатка, и тако све док ширина препознатог низа знакова не постане приближно једнака ширини нераздјеленог знака. За вјероватноћу тачности препознавања појединог овако добијеног низа знакова се узима геометријска средина тачности препознавања појединих знакова из низа. Овакав поступак повећава и тачност препознавања појединачног знака, јер се при препознавању, уколико се врши препознавање само оноликог дијела с лијеве

стране колико захвата образац, догађа да ужи знак који је приближно дио ширег има исту или већу вјероватноћу од ширег који стварно одговара издвојеном знаку, док се примјеном претходног поступка умјесто ужег знака добија одређени низ знакова, чија ће вјероватноћа бити мања од вјероватноће исправног, ширег знака.

Примјеном ријечника, лексичких и синтаксних правила и узимањем у обзир контекстних информација се вјероватноћа грешке при овим поступцима своди на најмању могућу мјеру.

Поступак придруживања образаца

Постоје разни поступци препознавања. Најједноставнији (и често најтачнији) поступак јесте одређивање сличности препознатог знака са свим могућим обрасцима и придруживање врсти оног који је најсличнији. Недостатак овог поступка је рачунска захтјевност, уколико има много образаца, тако да је практично непримјенив за кинеско и слична писма са великим бројем знакова или у случају великог броја фонтова. На овај поступак значајно утиче ниво шума у претстави, посебно уколико се прво врши сегментација, а затим придруживање образаца јер тачке које су због шума погрешно сегментирани уносе велико одступање у односу на исправан образац, па може лако да се догоди да знак буде препознат као њему сличан, јер су се погрешно сегментирани тачке нашле на одговарајућим мјестима (нпр. када се у средини слова с нађу погрешно сегментирани тачке оно лако може да буде препознато као слово е). У случају да се придруживање образаца врши несегментираној области знака, одређеној из граница сегментираниог знака, утицај шума је мањи али је и поступак рачунски сложенији јер више не могу да се врше битске операције над пакованим битима који претстављају тачке већ се морају рачунати збирови и производи вриједности свјетлина. Тачност у случају сегментирани области се може повећати употребом већег броја образаца добијених вишеструким скенирањем и сегментациом исте врсте знака.

Сам поступак придруживања обрасца се може вршити на различите начине. Једна од могућности, у случају рада са сегментираним претставом, је да се посматрају само тачке објекта, па да се од броја тачака објекта које се поклапају у препознатом знаку и у тренутном обрасцу одузме број тачака објекта које се не

поклапају. Тачке позадине се при овом поступку не користе јер би то узроковало наклоност ка давању великих оцјена слагања обрасцима који имају велике позадинске области. Ипак, шупљине у знаку се могу сматрати градивним дијеловима, па се при рачунању мјере слагања могу сматрати тачке шупљина које се међусобно поклапају као слагање. Да би се ово извршило морају се тачке шупљина и у знаку и у обрасцу означити посебним бројем, тако да нису довољни само један и нула (објект и позадина), па се више не могу примијенити бинарне операције над пакованим битима, што значајно успорава поступак. Још један проблем у овом случају је што шум може у једном дијелу изазвати прекидање знака, па шупљина постаје вањска област, тако да се више не слаже са шупљином у одговарајућем обрасцу, тако да је овакав поступак осјетљивији на шум од претходног. Боља мјера слагања него у претходним случајевима се добија ако се мјера слагања из претходних поступака нормализује бројем тачака објекта (и шупљина у другом случају) у обрасцу. Нормализовани индекс слагања (NMI) се одређује из броја битних тачака у обрасцу које се слажу (M^+) и броја битних тачака у обрасцу које се не слажу (M^-) на основу следеће формуле:

$$NMI = \frac{M^+ - M^-}{M^+ + M^-}.$$

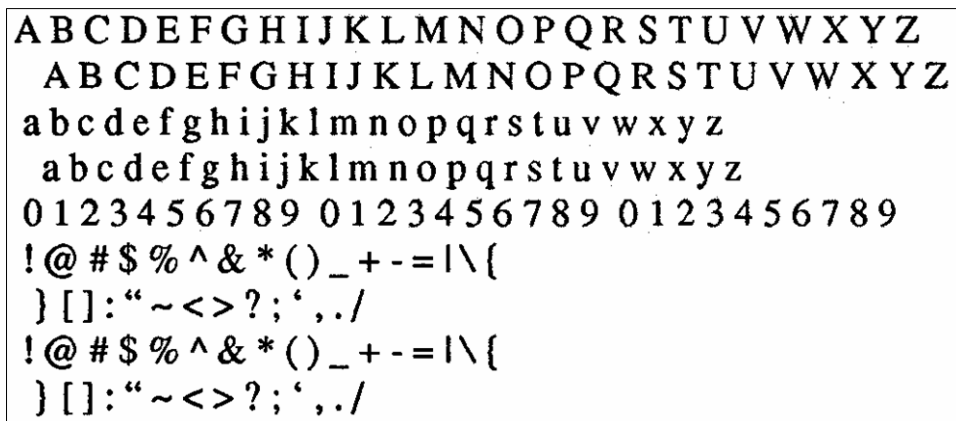
Нормализовани индекс слагања узима вриједности из интервала $[-1, 1]$ што је јако погодно.

Још једна могућност је да за мјеру слагања узмемо коефицијент корелације између области препознаваног знака и обрасца које се поклапају или само између тачака објекта у таквој области обрасца и њима одговарајућих тачака знака или да помножимо ове двије мјере или нађемо геометријску средину њихових апсолутних вриједности, за предзнак узимајући минус ако бар један од њих има предзнак минус.

На ивицама сегментираног знака се усљед шума појављују тачке због којих он има погрешне димензије описаног правоугаоника. Ово се може десити и због спајања знакова, кад се споје нижи и виши знак постојаће празан простор испод или изнад нижег знака. Због тог је потребно код поступка придруживања рачунати слагање са појединачним обрасцем помјереним по хоризонтали за неколико тачака у оба смјера, а исто тако по вертикали бар за исти број тачака, а боље је за онолики број тачака колика је разлика у висинама између обрасца и знака. Ово знатно увећава број поређења, пошто је укупан број поређења једног знака једнак умношку

броја могућих помјераја по хоризонталном и броја могућих помјераја по вертикалном правцу.

У програму који је дат у прилогу, направљеном на основу програма из [1], при чему је функција за рачунање слагања промијењена, кориштен је коефицијент корелације између области препознаваног знака и обрасца које се поклапају. Програм користи базу података образаца сачињену помоћу програма за учење образаца из њихове представе, који је преузет из литературе [1] и прилагођен. База образаца је добијена примјеном програма за учење на следећу слику:



Обрасци знакова

У програму за учење је унапријед одређен овакав распоред знакова, тако да је за сваки знак позната врста (слово) којој припада. Ова слика је добијена скенирањем и сегментациом, тако да више узорака појединог слова треба да омогући већу тачност препознавања у реалним условима.

Програм за препознавање чита обрасце из базе и слику која му је задана као улазни параметар и пише кодове препознатих знакова у теку која му је задана као излазни параметар. На стандардном излазу за сваки знак даје листу могућих кодова и њихове вјероватноће, поредане у опадајућем редоследу вјероватноћа. Примјеном програма на следећу слику:

Since the orientation is perfectly horizontal, the first step is to determine the position and extent of the lines of text in the image. This can be done by constructing a horizontal *projection* and searching it for minima. The projection is simply the sum of the pixel values in a specified direction, so a horizontal projection is the sum of the pixels in each row. The row of pixels that begins a new line will be one in which some of the pixels are black, and the last row belonging to that line will be the last one having any black pixels. The start and end columns for the line are found by searching the rows that belong to that line, from column zero through to the first column having a set pixel. The same is done, but in the reverse direction, to find the last set pixel in a line.

Слика на којој је вршено препознавање, преузета из [1]

добијен је следећи излаз:

Since the orientation is perfectly horizontal, the first step is to determine the position and extent of the lines of text in the image. This can be done by constructing a horizontal projection and searching it for minima. The projection is simply the sum of the pixel values in a specified direction, so a horizontal projection is the sum of the pixels in each row. The row of pixels that begins a new line will be one in which some of the pixels are black, and the last row belonging to that line will be the last one having any black pixels. The start and end columns for the line are found by searching the rows that belong to that line, from column zero through to the first column having a set pixel. The same is done, but in the reverse direction, to find the last set pixel in a line.

Црвеном бојом су исписана слова која су прескочена јер су била састављена са претходним словима, а зеленом бојом су маркирана слова која су погрешно препозната. Прескакање је смањило поузданост препознавања и за први лијеви знак. На основу поступка из тачке 3. међу поступцима за раздвајање знакова се добија и већа тачност препознавања лијевог знака а препознају се и остали знаци. У овом програму је додано изабарање ширег обрасца уколико су сличности више образаца знаку веће од границе сигурности препознавања. На овај начин се дјелимично умањују посљедице недостатка препознавања свих спојених знакова на тачност препознавања лијевог знака. Види се да су у ријечи *projection* тачно препозната свега три слова а уз то су унутар ријечи додане двије тачке које нису постојале у препознаваном тексту. Ово је посљедица непостојања образаца за нагнута слова. Изузимајући ову ријеч постоји укупно 628 слова, од тога су 44 прескочена због спајања, па је препознавано само 584 слова, од чега је 16 погрешно препознато. Одавде је тачност препознавања 97,26%. Из теке са листом вјероватноћа за сваки препознавани знак дате у прилогу види се да би додавањем једноставног синтаксног правила да велико слово не може да се нађе у средини ријечи још 5 слова било тачно препознато, па би тачност била 98,12%, а додавањем провјере у ријечнику би број грешака био смањен бар за ред величине.

Такође је могуће посматрати руб или рубове (уколико знак има шупљину или се састоји из више дијелова) знака као функцију једне промјењиве, па вршити одређивање сличност облика руба. На овај начин се комплексност смањује са квадратне $O(M \cdot N)$ на линеарну $O(M + N)$, гдје су M и N димензије знака, међутим мана овог поступка је што је далеко осјетљивији на шум, усљед којег руб знака бива искрзан а такође може доћи и до потпуне промјене облика руба усљед прекидања знака. Један од таквих поступака је и претварање руба знака у векторски облик.

Поступци вјештачке интелигенције

У новије вријеме се све више користе поступци вјештачке интелигенције, међу којима највише генетски алгоритми и неуронске мреже.

Генетски алгоритми су у ствари алгоритми за тестирање и селекцију одређеног скупа алгоритама истог облика, зависних од одређеног скупа параметара. Сваки појединачни алгоритам се добија изабирањем вектора параметара. Поступак се састоји из одређивања почетног скупа вектора параметара случајним избором, примјене њима одређених алгоритама на задани проблем и оцјене њихове доброте. Што је доброта одређеног алгоритама већа то је већа шанса да ће његов вектор параметара бити укључен у стварање следеће генерације алгоритама. Вектори параметара следеће генерације се добијају комбиновањем случајно изабраних вектора параметара претходне генерације, узимајући у обзир да је вјероватноћа избора неког вектора параметара тим већа што му је доброта већа. Битно је да популација вектора параметара буде довољно велика да би се дошло до рјешења, али и да не буде превелика, јер би проце трајао предуго. Обично је величина популације стална, мада има појединих поступака при којима се популација дјелимично повећава у току тражења рјешења, а са повећањем доброте (тј. приближавањем рјешењу) поново се смањује.

Неуронска мрежа је масивно паралелни дистрибуисани процесор који има природну склоност за чување искуственог знања и омогућавање његовог кориштења. Слична је мозгу по два питања:

1. Знање стиче процесом учења.
2. За смјештај знања се користе јачине међунеуронских веза (синаптичке тежине).

Постоји више врста неуронских мрежа, од којих се за препознавање знакова највише користе мреже са повратном пропагациом и самоорганизујуће мреже (мреже са Хебиановим учењем, мреже са компетитивним учењем и мреже са очувањем максималне количине информација). Предност неуронских мрежа над поступком придруживања образаца је што неуронске мреже у току учења могу одредити из више примјерака једног знака (на примјер исто слово из различитих фонтова) особине које га стварно одређују, тако да долази до генерализације, па чак и кад се на улазу појави знак који представља исто слово из фонтова који није кориштен при учењу он ће бити препознат уколико му није суштински различит

облик. Проблем неуронских мрежа је брз раст броја неурона и још бржи раст броја синапси са бројем суштински различитих знакова које треба препознати. Због тог се неуронске мреже у свом најпростијем облику (нпр. мрежа са пропагациом уназад са једним скривеним слојем) користе само у случају словних писама у вишефонтовском окружењу, а у случају писама са великим бројем знакова се употребљава нека од самоорганизујућих мрежа, која ствара одређену подјелу знакова на врсте по међусобној сличности, а затим се за сваку врсту прави и тренира посебна једноставна неуронска мрежа. При процесу препознавања прво самоорганизујућа мрежа одређује врсту којој знак припада, а затим се само препознавање врши мрежом намијењеном за ту врсту, тј. њом се одређује знак у врсти.

Закључак

Оптичко препознавање знакова има велики распон различитих примјена. У неким од њих су постигнути значајни резултати, тако да се у тим областима може сматрати да је то практично ријешен проблем. Међутим, у истраживачкој области остају проблеми сложених докумената, нарочито у вишејезичком окружењу и докумената са разним математичким, хемијским и другим формулама и цртежима. У будућност ће се тежити све већој интеграцији процеса сегментације слике, препознавања знакова и контекстне обраде, при чему ће у блиској будућности у оквиру контекстне обраде бити посматрани само синтаксички и лексички ниво док се у даљој будућности очекује укључење семантичког и прагматичког нивоа, што је условљено напретком у области вјештачке интелигенције. Ова интеграција драстично повећава комплексност обраде, што захтијева или изналажење нових, ефикаснијих алгоритама, или повећање процесорске моћи рачунара за неколико редова величине а највјероватнији правац развоја ће водити поступном напретку на оба поља (већа процесорска моћ ће омогућити развој и употребу нових врста алгоритама).

Литература

1. J. R. Parker, "Algorithms for Image Processing and Computer Vision"
2. "Survey of the State of the Art in Human Language Technology",
cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html
3. Vassilios Chatzis and Ioannis Pitas, "A Generalized Fuzzy Mathematical Morphology and its Application in Robust 2-D and 3-D Object Representation"
4. Ravikanth Malladi and James A. Sethian, "A Unified Approach to Noise Removal, Image Enhancement, and Shape Recovery"
5. Shan Mo and V. John Mathews, "Adaptive, Quadratic Preprocessing of Document Images for Binarization"
6. Jezekiel Ben-Arie, Zhiqian Wang and K. Raghunath Rao, "Affine Invariant Shape Representation and Recognition using Gaussian Kernels and Multi-Dimensional Indexing"
7. Oscar E. Agazzi, Shyh-shiaw Kuo, Esther Levin and Roberto Pieraccini, "Connected and Degraded Text Recognition Using Planar Hidden Markov Models"
8. Andy Tsai, Anthony Yezzi, Jr., and Alan S. Willsky, "Curve Evolution Implementation of the Mumford-Shah Functional for Image Segmentation, Denoising, Interpolation, and Magnification"
9. Hungwen Li, "Fast Hough Transform for Multidimensional Signal Processing"
10. Majid Mirmehdi, Phil L. Palmer, Josef Kittler, and Homam Dabis, "Feedback Control Strategies for Object Recognition"
11. Haiyan Wang and Bijoy Ghosh, "Geometric Active Deformable Models in Shape Modeling"
12. Scott T. Acton and Dipti Prasad Mukherjee, "Image Edges from Area Morphology"
13. Pui-Kin Ser and Wan-Chi Siu, "Invariant Hough Transform With Matching Technique For The Recognition Of Non-Analytic Objects"
14. Jean-Bernard Martens, "Local Orientation Analysis In Images By Means Of The Hermite Transform"
15. Alexander S. Sherstinsky and Rosalind W. Picard, "M-Lattice: From Morphogenesis to Image Processing"
16. Alexander C. P. Loui, Anastasios N. Venetsanopoulos and Kenneth Carless Smith, "Morphological Autocorrelation Transform: A New Representation and Classification Scheme for Two-Dimensional Images"
17. Ping-Fai Yang and Petros Maragos, "Morphological Systems For Character Image Processing And Recognition"

18. Yassin M. Y. Hasan and Lina J. Karam, "Morphological Text Extraction from Images"
19. Zhi-Pei Liang and David C. Munson, Jr., "Partial Radon Transforms"
20. K. C. Wong, H. C. Sim and J. Kittler, "Recognition of Two Dimensional Objects Based on a Novel Generalized Hough Transform Method"
21. Avanindra and Subhasis Chaudhuri, "Robust Detection of Skew in Document Images"
22. Hans Knutsson Mats Andersson, "Robust N-Dimensional Orientation Estimation using Quadrature Filters and Tensor Whitening"
23. Do Hyun Chung and Guillermo Sapiro, "Segmentation-free Skeletonization of Gray-scale Images via PDE's"
24. George J. Grevera and Jayaram K. Udupa, "Shape-Based Interpolation of Multidimensional Grey-Level Images"
25. Sai Prasad Raja and Jayaram K. Udupa, "Shape-Based Interpolation of Multidimensional Objects"
26. Shy-Shyan Chen and Frank Y. Shih, "Skeletonization for Fuzzy Degraded Character Images"
27. Michael Unser, "Splines: A Perfect Fit for Signal and image Processing"
28. K. P. Philip, E. L. Dove, D. D. McPherson, N. L. Gotteiner, W. Stanford, and K. B. Chandran, "The Fuzzy Hough Transform-Feature Extraction in Medical Images"
29. W. W. Cindy Jiang, "Thresholding and Enhancement of Text Images for Character Recognition"
30. Tony F. Chan and Chiu-Kwong Wong, "Total Variation Blind Deconvolution"
31. Joseph M. Reinhardt and William E. Higgins, "Toward Efficient Morphological Shape Representation"
32. Zhanyi. Hu and Song De Ma, "Towards a New Framework of the Hough Transform"
33. Dong-Gyu Sim and Rae-Hong Park, "Two-Dimensional Object Alignment Based on the Robust Oriented Hausdorff Similarity Measure"
34. Francine R. Chen, Lynn D. Wilcox, and Dan S. Bloomberg, "Word Spotting In Scanned Images Using Hidden Markov Models"