FIATAL MŰSZAKIAK TUDOMÁNYOS ÜLÉSSZAKA XIX.

Kolozsvár, 2014. március 20–21.

SZÖVEGFELISMERÉSEN ALAPULÓ AUTOMATIKUS KÖNYVKATEGORIZÁLÓ ALKALMAZÁS

AUTOMATIC BOOK CATEGORIZATION SYSTEM BASED ON TEXT RECOGNITION

MIHÁLYI Martin⁽¹⁾, WINDISCH Gergely⁽²⁾, VÁMOSSY Zoltán⁽³⁾

(1)Óbudai Egyetem, Neumann János Informatikai kar, H-1031 Budapest, Bécsi út 96/b, Magyarország, djozike@gmail.hu

(2)Óbudai Egyetem, Neumann János Informatikai kar, H-1031 Budapest, Bécsi út 96/b, Magyarország, windisch.gergely@nik.uni-obuda.hu

(3)Óbudai Egyetem, Neumann János Informatikai kar, H-1031 Budapest, Bécsi út 96/b, Magyarország, vamossy.zoltan@nik.uni-obuda.hu

Abstract

In this study, we design and implement a book categorization system. Systems based on local image features are not effective in some cases in book spine detection. This is because few book spines contain only text component. The implemented system at first detects the book spines on the image of a bookshelf, then recognizes the text on the book spines by optical character recognition (OCR) and searches in the book database. Finally the implemented system displays the results for the user.

Keywords: image processing, book spine detection, text recognition

Összefoglalás

Ezen cikk egy könyvkategorizáló rendszert mutat be. A képjellemző alapú rendszerek egyes esetekben nem mutatnak jó hatásfokot a könyvek esetében. Ennek oka, hogy néhány könyvgerinc csak szöveget tartalmaz. Ezért adott egy szövegfelismerés alapú rendszer alkalmazása a problémára. A fejlesztett alkalmazás először egy könyvespolcra készült képről könyvgerinceket detektál, majd a rajtuk lévő szöveget szövegfelismerés segítségével felismeri és a könyv adatbázisban keresést végez. Végül a fejlesztett rendszer a felhasználó számára megjeleníti az adott könyvről további információkat.

Kulcsszavak: képfeldolgozás, könyvgerinc keresés, szöveg felismerés

1. Bevezetés

Nagy mennyiségű könyv akár otthoni, akár bolti vagy könyvtári karbantartása, rendszerezése egy napjainkban is egy problémás feladat. A kézi könyvkategorizálás egy nagyon időigényes és rengeteg hibalehetőséget magában hordozó feladat. Léteznek erre már félautomata, rádió-frekvenciás (RFID) vagy vonalkód alapú megoldások, ezeknek különleges eszköz igényük van és használatához először el kell látni megfelelő azonosítóval a könyvek gerincét és ezért nem használatosak az otthoni vagy üzleti könyvespolc esetén. Megfelelő

megoldás lehet egy képfeldolgozás alapú módszer. Ez esetben egy könyvespolcra mobiltelefonnal vagy akár számítógéppel készített kép alapján teljesen automatikusan végzi el a rendszer a könyvek karbantartását. A már meglévő kísérleti rendszerek ezt a képfeldolgozást képjellemző alapú módszerrel (referencia pont meghatározás, társítás) végzik, amely gyenge teljesítményt mutat olyan könyvek esetében, ahol kevés az ilyen meghatározható pont. Ilyen könyvek jellemzően azok, amelyek gerincén csak szöveges információ található. Ezen dolgozatban egy lehetséges megvalósítását mutatjuk be egy olyan rendszernek, amely szövegfelismerés alapján végzi a könyvek kategorizálását.

2. Rendszerfelépítés

A megvalósított alkalmazást három részmodulra került osztásra: egy könyvgerinc felismerő rész, egy szövegfelismerést végző rész, és egy adatbázis keresést végző rész, ahogy az 1. ábrán látható.

1. ábra. A megvalósított rendszer modul szintű felépítése

A könyvdetektáló rész feladata, hogy a könyvespolcra készült képen a könyvgerincek detektálása majd azok szerint az eredeti kép szétdarabolása olyan képekké, amelyeken már csak egy-egy könyv található. Ezt követően a cím felismerő modul előfeldolgozás után detektálja a képen lévő karaktereket, majd ezeken a karaktereken utófeldolgozást végez az adatbázis keresés modul és az így előálló karaktersorozattal leginkább egyező eredményt megjeleníti a felhasználó számára.

3. Könyvdetektálás modul

Ezen modul célja, hogy a könyvespolcra készült képen lévő könyveket meghatározza és könyvenként szétvágja a képet. A könyvdetektálás esetében először a képeken lévő élek kerültek meghatározásra. A legpontosabb él meghatározást a Canny éldetektáló [1] segítségével sikerült elérni. Majd ezekből az élekből Hough-transzformáció [2] segítségével vonalakat állítottunk elő. Ezt követően, mivel ismert, hogy a kép könyveket ábrázol, az egyes irányú élek előfordulásának arányából meghatározásra kerül a könyvek iránya, azaz merre található a legtöbb él, majd ha szükséges a kép elforgatásra kerül, hogy az élek függőleges irányúak legyenek. Utolsó lépésként az eredeti könyvespolcra készült kép detektált élek szerinti szétdarabolása történik. Fontos megjegyezni, hogy a bemeneti képen egyetlen sorban vagy oszlopban kell elhelyezkedni a könyveknek, több sor esetén nem működőképes ez a megvalósítás. Az irodalom se jegyez olyan módszert, amely több sor vagy oszlop esetén is megfelelően működőképes lenne.

4. Címfelismerés modul

A modul célja, hogy a bemenetként érkező egy-egy könyvgerincet ábrázoló képeken meghatározzák a szövegeket. Ehhez először a képen előfeldolgozást kell végezni a jobb hatásfok elérése érdekében, ilyen szükséges előfeldolgozások a binarizáció és a szöveg orientációjának meghatározása. Ezt követően egy online karakter felismerő (OCR) rendszer elvégzi a karakterek felismerését. Majd ezt az eredményt szolgáltatja a következő modul számára.

4.1 Binarizáció

A binarizáció célja, hogy a karakterfelismerő számára kiemelje a képen a fontos részeket, azaz a betűket. Több binarizációs algoritmust is megvizsgálva (Otsu, Bradley, Niblack, Sauvola) az Otsu [3] binarizáció esetén sikerült a legnagyobb hatásfokot elérni a karakterfelismerés terén. Mint ahogy a 2. ábrán is látható Bertsekas: Introduction to Probability című könyvének gerincén végzett Otsu, Bradley [4] binarizációt, illetve ez alatt az adott binarizáció után kapott rendszer által felismert szöveg látható.

2. ábra Az egyes binarizációs algoritmusok eredménye, illetve az adott képen való szövegfelismerés eredménye

4.2 Szöveg orientáció

A szöveg orientációjának meghatározása is egy fontos kérdés, mivel a karakterfelismerő (OCR) csak a megfelelően elhelyezkedő betűket képes detektálni, azaz amelyek nem fejtetőn, vagy adott szöggel elforgatva helyezkednek el. Sajnos a könyv orientációjából nem következik a gerincen lévő betűk orientációja, mivel egyrészt a nemzetközi szabvány szerint a szöveget fentről lefele, a magyar szabvány szerint lentől felfele kell a könyvgerincetre írni, másrészt előfordulhat esztétikai okból ettől a szabványtól való eltérés is, például a szerző neve állítva, a cím elfordítva szerepel a könyvgerincen. Erre a problémára megoldást jelentet, hogy dilatációt végeztünk a binarizált képen, annak érdekében, hogy a detektált betűk összeérjenek. Ezek után egyszerűen meghatározhatóvá váltak az egy-egy összefüggő szövegrészek, amelyek szerint szegmentálható a kép, majd az egyes részeket különböző irányba elforgatva többször is karakterfelismerést végeztünk. Amelyik irányban a legnagyobb egyezést mutatja a az adatbázisból egy rekorddal az adott rész, az a megfelelő irány. Ez a megvalósítás azonban, a könyvek változatos színvilága miatt (sötét háttéren világos szöveg, világos háttéren sötét szöveg) nem javított jelentősen az ennél egyszerűbb és emiatt gyorsabb megoldáson, azaz amikor az egész binarizált képet négyszer végzünk karakterfelismerést (eredeti állapot, 90 fokkal való elforgatás, 180 fokkal való elforgatás, 270 fokkal való elforgatás). Ez esetben időigényesebb a karakterfelismerés (nagyobb területen alkalmazzuk), viszont a dilatáció és szegmentálás elhagyása miatt összességében gyorsabb megoldás.

4.3 Karakterfelismerés

Az alkalmazáshoz egy már létező online karakterfelismerő (OCR) engine-t használtunk. Az irodalom alapján a legmegfelelőbb ilyen rendszernek a Tesseract OCR engine [5] bizonyult, ezért ezt használtuk. Az engine-t csak alfanumerikus karakterek felismerésére korlátoztuk (könyvek esetén kis eséllyel fordul elő más karakter).

4.4 Utófeldolgozás

A mérések során kiderült, hogy sok esetben a karakterfelismerő a valós képen lévő betűkön kívül a megjelenő zajokat is „felismerte”. Ezen fals pozitív betűk kiszűrésére egy szótár került kiépítésre, amely értelmes szavakat tartalmaz. A feldolgozás során a karakterfelismerés eredménye mindig ezzel a szótárral kerül összevetésre és ha egy bizonyos százalékot (80-85%) meghalad az egyezés akkor a szótárból a megfelelő szóra cserélődik a detektált szó. Az összehasonlítást longest common subsequence (LCS) problémaként közelítettük meg és az szerint került megoldásra.

5. Adatbázis keresés modul

Az adatbázis egy szerveren került eltárolásra, a kliens oldal a könyvekről készült képet beküldi a szervernek, majd onnantól minden feldolgozási lépést a szerver végez és az eredményeket visszaküldi a kliensnek, ami azt megjeleníti a felhasználó számára. Ezen adatbázisban a keresés a 4.4 pontban leírt módszerhez hasonlóan történik, az eltérés annyi, hogy nem az eredmény nem egy szótárral kerül összevetésre, hanem az adatbázisban tárolt szerzőkkel és címekkel.

6. Eredmények

A pontosabb összehasonlítás érdekében a tesztelések az irodalomban [6] használt tesztadatbázis alapján történtek. Az eredmények csak a szövegfelismerést tekintve közel 70%-os sikeres felismerést mutattak, azaz az eseték 70%-ában csak a könyvgerincet ábrázoló kép alapján a megfelelő könyv adatokat jelenítette meg hozzá a rendszer. Kiemelendő az adatbázis kereső pontossága. Olyan esetekben, amikor szövegfelismerő motor legalább egy a címben szereplő szót teljesen hibátlanul felismert, az adatbázis keresés 95%-os pontossággal találta meg a hozzá tartozó könyvinformációt az adatbázisban.

7. Összefoglaló

Ezen cikkben bemutattam egy képfeldolgozás alapú könyvkategorizáló rendszer lehetséges megvalósítását, amely a hasonló rendszerekkel ellentétben csak szövegfelismerés alapján végzi a kategorizálást. Részletesen bemutattam a szövegfelismerés menetét, és megállapítottam, hogy ez a megközelítés jól használható, olyan könyvek esetében, amely gerincén kevés referencia pont, csak szöveg található, illetve hogy a címfelismerés hatásfoka nagyban függ a megfelelő binarizációs algoritmustól, illetve paraméter értékektől.

Irodalom

- [1] J. Canny: *A computational approach to edge detection*, IEEE Trans. Pattern Analysis and Machine Intelligence, 1986. pp. 8(6):679–698,
- [2] S. N. Srihari, V. Govindaraju: *Analysis of textual images using the Hough transform*, Machine Vision and Applications, 1989. 2.3, pp. 141-153.
- [3] N. Otsu: *A threshold selection method from gray-level histograms*. IEEE Transactions on Systems, Man and Cybernetics 1979, 9(1), pp. 62-66
- [4] D. Bradley, G. Roth: *Adaptive thresholding using the integral image*. Journal of graphics, gpu, and game tools, 2007, 12.2: 13-21.
- [5] R. Smith: *An overview of the Tesseract OCR engine*, ICDAR 2007. Ninth International Conference on Document Analysis and Recognition, IEEE, 2007. p. 629-633.
- [6] D. Chen, S. Tsai, C.-H. Hsu, K.-Y. Kim, J. P. Singh, and B. Girod: *Building book inventories using smartphones*, ACM Multimedia (MM), October 2010. pp. 651-654