

OCR API'leri ve Gerçek Zamanlı OCR Uygulaması

Hüseyin KUTLU¹, Ersan YAZAN²

¹ Adıyaman Üniversitesi, Besni Meslek Yüksekokulu, Adıyaman

² Adıyaman Üniversitesi, Besni Meslek

Yüksekokulu, Adıyaman

hkutlu@adiyaman.edu.tr,

eyazan@adiyaman.edu.tr,

Özet: Optik Karakter Tanıma veya OCR, taranmış kâğıt evrakları, PDF dosyaları veya dijital bir kamerayla çekilen resimler gibi değişik belge türlerini düzenlenebilir ve aranabilir verilere dönüştürmenize olanak sağlayan bir teknolojidir. Bir tarayıcının tüm yapabileceği şey, belgenin resmini oluşturmaktır ki bu raster görüntüsü olarak bilinen siyah beyaz veya renkli noktalar topluluğundan başka bir şey değildir. Taranmış belgelerden, kamera resimlerinden verileri çıkartmak ve başka bir amaç için düzenlemek amacıyla resimlerdeki harfleri seçip ayıracak, onlardan kelimeleri ve kelimelerden de cümleleri oluşturup bu sayede orijinal belgenin içeriğine erişmenize ve düzenlemenize olanak tanıyacak bir OCR yazılımları mevcuttur. Söz konusu yazılımlar çeşitli amaçlar için kullanılmak üzere API'lere dönüşmüşlerdir. Yapılan çalışmada gerçek zamanlı görüntüden OCR API'leri (OCR Motoru - OCR Engines) ile metin okuması gerçekleştirilmiş, OCR API'lerinin kullanım alanları belirtilmiş ve performansları karşılaştırılmıştır.

Anahtar Sözcükler: OCR, OCR API, OCR Motoru, Gerçek Zamanlı OCR uygulaması, EmguCV, Tesseract.

OCR API's and Real Time OCR APPLICATION

Abstract: Optical Character Recognition, or OCR scanned paper documents, PDF files , such as pictures taken with a digital camera or a different document types into editable and searchable data is a technology that allows you to convert . All the things you can do with a scanner , the document is to create a picture of this raster image in black and white or colored dots , known as the community is not something other than . Scanned documents, camera picture data from the extract, and for no other purpose to arrange photos in the letters glean that from them, words and of words in sentences create and thus the original document to access its content and allows you to edit a OCR software is available. The software in question to be used for various purposes have turned into APIs . In a study conducted in real-time APIs from the image OCR (OCR Engine - OCR Engines) has been carried out with the text reading , OCR APIs specified areas and their performances are compared .

Keywords: OCR, OCR API's, OCR Engines, Real Time OCR Application, EmguCV, Tesseract.

1. Giriş

Optik Karakter Tanıma veya OCR, taranmış kâğıt evrakları, PDF dosyaları veya dijital bir kamerayla çekilen resimler gibi değişik belge türlerini düzenlenebilir ve aranabilir verilere dönüştürmenize olanak sağlayan bir teknolojidir. Bilgisayarların hızlanması, görüntü alma cihazlarının gelişmesi ile görüntü işlemeye olan ilgiyi daha da arttırmış dolayısı ile gelişen görüntü işleme teknikleri ile bir çok uygulamalar gelişmiş ve gelişmektedir. OCR de söz konusu uygulamalar arasında önemli bir alandır.

Api, Application Programming Interface kelimelerinin baş harflerinden oluşan bir kısaltmadır. Uygulama Programlama Ara yüzü anlamına gelen Api, herhangi bir uygulamanın belli işlevlerini diğer uygulamalarında kullanabilmesi için oluşturulmuş bir modüldür. Optik Karakter Tanıma işlemi, yani elektronik görüntüler üzerindeki karakterlerin ya da metin bilgilerinin okunarak ASCII koda dönüştürülmesi işlemi ile makineler tarafından yazılmış karakterler ve işaretler kolaylıkla okunup ASCII koda dönüştürülebilirler bu işlemi yapan birçok API vardır [1]. Tesseract, ExperVision TypeReader & RTK, ABBYY FineReader, AnyDoc Software, Aquaforest OCR SDK, LEADTOOLS, CuneiForm/OpenOCR, Transym OCR, Image to OCR Converter, SimpleOCR, Dynamsoft OCR SDK, OmniPage, Microsoft Office OneNote 2007, FreeOCR, GOOCR, Ocrad, SmartScore, Microsoft Office Document Imaging (MODI), Puma.NET, ReadSoft, Scantron, OCRFeeder, OCRopus uluslar arası bilinen API'lerden bazılarıdır. API kullanarak bir resimden metin okumak kolay bir işlemdir. Yapılan çalışmada gerçek zamanlı bir video görüntüsünden

yani kameradan OCR API'leri kullanılarak metin okuma işlemi gerçekleştirilmiştir. Bu uygulama geliştirilmesi durumunda birçok yerde kullanılan tarayıcıların yerini her bilgisayarda bulunan kameraların alacağı düşünülmektedir. Aynı zamanda metin okuyan APIler yardımı görme engelli birinin doğru bir şekilde kameraya tuttuğu kâğıdı okuması sağlanabilir. [8].

2. OCR API'leri Nedir? Nasıl Çalışır? Nerelerde kullanılır?

OCR API'leri genel olarak amaçları olan elektronik görüntüler üzerindeki karakterlerin ya da metin bilgilerinin okunarak ASCII koda dönüştürülmesi işlemlerinde kullanılırlar. OCR metodolojileri kullanılarak makineler tarafından yazılmış karakterler ve işaretler kolaylıkla okunup ASCII koda dönüştürülebilirler. OCR, makineler tarafından yazılmış karakterlerin okunup tanılırlar. OCR iki sınıfa ayrılır: Otomatik Karakter Tanıma (ADC-Automatic Character Recognition) ve metin tanıma (TR-Text Recognition). ADC ile her karakter doğru olarak tanınır ve karşılığında bir ASCII kod atanır.

Karakterlerin tanınmasında birkaç OCR metodolojisi kullanılır. Bunlar arasında topolojik tabanlı olan, yapay sinir ağlarına dayanan, matris eşleme ile gerçekleşen, öznitelik analizi ile çalışan ve eğrileri izleyerek karakter tanıma işlemi gerçekleştiren metodolojiler yer almaktadır. Karakterlerin daha büyük doğrulukla tanınması için, okunan verilere ön işleme ve son işleme algoritmaları uygulanmaktadır. Bu algoritmalar, formların belirlenmesinde, formlar üzerindeki belirli alanların silinmesinde, form ve karakterlerin düzeltilmesinde ve görüntünün iyileştirilmesinde kullanılırlar. Karakterlerin büyük doğrulukla

okunmasında, sisteme sunulan elektronik görüntünün mümkün olduğunca keskin, temiz ve düzgün olması oldukça büyük önem taşımaktadır. Buna rağmen, günlük yaşamda kullanılan dokümanlar ve formlar genellikle kirli, yıpranmış ve katlanmışlardır. Tanıma işleminden önce uygulanacak ön işleme algoritmaları ile formun elektronik görüntüsü temizlenir ve karakterler mümkün olduğunca iyileştirilir.

OCR yazılımlarının benzer amaçlı farklı uygulamalarda kullanılabilmesi için API'leri oluşturulmuştur. Söz konusu API'ler birçok yazılım geliştirme ortamını desteklemektedirler.

Günümüzde, doküman işleme işlemleri, üzerine çalışmalar yapılan bir alandır. OCR uygulamaları sıklıkla şu alanlarda kullanılmaktadır:

- Doküman İşleme
- Masaüstü Yayıncılık
- Mahkemelerle İlgili Alanlar
- Vergilendirme ve Tahsilat
- Personel Kayıt Yönetimi
- Nüfus Sayımı Formlarının İşlenmesi
- Çek İşleme
- Ödeme İşleme
- Emekli Fonu İşleme
- Sipariş İşleme
- Text CAPTCHA Aşma

Gibi uygulama alanlarında OCR API'lerinin kullandıklarını görebiliriz.

3.Yaygın OCR API'lerinin İncelenmesi

Yapılan uygulama kapsamında yaygın olarak bilinen Abbyy FineReader, Tesseract, MODI (Microsoft Office Document Imaging), Office 2007 CR API'leri incelenmiştir.

3.1. Tesseract OCR API

Açık kaynak kodludur. Hewlett-Packard firması tarafından yaratılmıştır ve sponsoru Google firmasıdır. Tesseract OCR API bellekte az yer kaplar ve hızlıdır. Bir çok dili destekler, ücretsizdir.

Açık kaynak kodlu olmasının avantajlarının yanı sıra Web platform desteğinin olmaması ve emsallerine göre daha yavaş gelişmesi ve daha az performans göstermesi dezavantajları arasında gösterilebilir.

3.2. Abbyy FineReader OCR API

En çok bilinen OCR API'lerindedir. Bu alanda bir çok ödül almıştır. Tüm geliştirme ortamlarını destekler. Kurulumu kolaydır. Bulanık ve dönük metinleri OCR edebilmiştir. Mobil platform desteği vardır. Geniş bir dil desteği bulunmaktadır.

Lisanslama hizmeti zahmetli ve pahalıdır. Performansı iyi fakat yavaştır.

3.3. MODI (Microsoft Office Document Imaging), Office 2007 OCR API

MODI (Microsoft Office Document Imaging), Office 2007 ile birlikte ücretsiz olarak dağıtılmıştır. .NET kütüphanesi ile uyumlu çalışmaktadır.

Microsoft tarafından artık desteklenmemektedir.

