



artificial intelligence studies

ARTIFICIAL INTELLIGENCE STUDIES YAPAY ZEKA ÇALIŞMALARI

2018, 1(1):39-46

www.aistudies.org

Gazi AKADEMİK
YAYINCILIK

Osmanlı Alfabeti ile Yazılan Kelimelerin Türk Alfabetine Çevrilmesi Üzerine Bir Yöntem

Önder ÖZBEK^a

^a Gazi Üniversitesi Bilgisayar Mühendisliği Bölümü, ANKARA 06500, TÜRKİYE

MAKALE BİLGİSİ

Alınma: 04.04.2018
Kabul: 23.06.2018

Anahtar Kelimeler:

Optik Karakter
Tanıma, Osmanlı
Alfabeti, Türk
Alfabeti, Osmanlı
Türkçesi, Levenshtein
Mesafesi

***Sorumlu Yazar:**

e-posta:
oozbek@gazi.edu.tr

ÖZET

Osmanlı alfabesi, Arap harflerle Türkçe yazı dili olarak kullanılmıştır. Türk alfabesi, Latin harflerle Türkçe yazı dili olarak kullanılmaktadır. Arşivlerde Osmanlı alfabesi ile yazılan sayısız belge bulunmaktadır. Bu çalışmada, Osmanlı alfabesi ile yazılmış kelimelerin görüntüsünün, optik karakter tanıma yöntemi ile düzenlenebilir metne dönüştürülmüştür. Bu sayede kelimeler metne çevrilmiştir ve daha sonrada bu kelimelerdeki karakterler Türk alfabesindeki karşılıklarını kullanarak anlaşılır hale getirilmiştir. Türk alfabesine çevrilmiş kelimeler, Türkçe sözcüklerin olduğu tabloda kıyaslanarak doğruluk oranı artırılmaya çalışılmıştır. Kıyaslamak işlemi için benzerlik değeri veren bir algoritma kullanılmıştır.

DOI: 10.30855/AIS.2018.01.01.05

A Method on The Translation of The Words Written by Ottoman Alphabet to Turkish Alphabet

ARTICLE INFO

Received:
04.04.2018
Accepted:
23.06.2018

Keywords:

Optical Character
Recognition,
Tesseract, Ottoman
Alphabet, Turkish
alphabet, Levenshtein
Distance

***Corresponding**

Authors

e-mail:
oozbek@gazi.edu.tr

ABSTRACT

The Ottoman alphabet was used as a writing language in Turkish with Arabic letters. The Turkish alphabet is used as a writing language in Turkish with Latin letters. There are numerous documents written in the Ottoman alphabet in the archives. In this study, the image of the words written in Ottoman alphabet was converted into editable text by optical character recognition method. In this way the words are translated into text. Later, the characters in these words were made understandable by using their equivalents in the Turkish alphabet. It was tried to increase the accuracy rate by comparing the words translated into Turkish alphabet and the table where Turkish words were found. An algorithm that gives a similarity value is used for the comparison process.

DOI: 10.30855/AIS.2018.01.01.05

1. GİRİŞ (INTRODUCTION)

Osmanlı alfabesi 10. yüzyılda İslamiyeti kabuledilmesi ile yazı dili olarak Arap Alfabeti alınmış ve Türk Dilinin özelliklerine göre düzenlemiştir. Osmanlı alfabesinin Türkçe'nin ünlü seslerini ifade etmekte yetersiz kalmasından ötürü imla kargaşası meydana gelmiştir. Yazılı basın ve resmi okul kitapların yaygınlaşması ile bu

durum giderek daha çok hisedilir hale gelmiştir. 19. yüzyıl ile beraber yeni bir alfabe düzenlemesi istenmiştir. Bu istekler iki türlü olmuştur. Biri Osmanlı Alfabesi'nin güncellenmesi diğeri ise Latin harflerinin kabuledilmesi şeklindedir. Bunun sonucunda 1 Kasım 1928 yılında "Yeni Türk harflerinin kabul ve tatbiki hakkında Kanun" yasanın kabul edilmesi ile Harf Devrimi gerçekleştirilmiştir. Bu yasanın kabulüyle Arap harflerin kullanıldığı Osmanlı alfabesi yerine Latin harflerin kullanıldığı Türk alfabesi yürürlüğe konulmuştur [1].

Günümüzde Osmanlı alfabesi ile yazılmış eserler ve belgeler hala bulunmaktadır. Osmanlı alfabesi ile yazılan kelimeler arap harfleri ile yazıldığından türkiyede çoğu kişi için yazılması ve anlaşılması güç bir hale gelmiştir. Bunu gidermek için Osmanlı alfabesi ile yazılmış bir kelimenin karakterlerini tek tek girmek yerine onun görüntüsü üzerinden işlemlerle Türk alfabesindeki karşılığını almak daha kolay olabilir. Osmanlı belgelerin, elle yazılmış veya tarayıcılarla içi içe girmiş yazıların anlaşılabilmesi için veya ayrıştırılabilmesi için yöntemler geliştirilmiştir. Bu belgelerin anlaşılabilmesi için OCR(Optical Character Recognition- Optik Karakter Tanıma) sistemleri kullanılmıştır. Bu sayede yazılanlar bir metin olarak elde edilmesi sağlanmıştır. Bunun yanında Osmanlı alfabesindeki karakterler için bir klavye yapmakta mümkündür.

İlerideki literatürdeki çalışmalar bölümünde, bu çalışma konusuna benzer başka diğerk çalışmalar hakkında bilgilere değinilmiştir. Medot ve materyal bölümünde, bu çalışma ile ilgili kullanılan metod ve materyaller hakkında bilgiler verilmiştir. Deneysel çalışmalar bölümünde, bu çalışmadaki test yöntemlerine, elde edilen sonuçlara ve diğerk benzer çalışmalar ile kıyaslanmasına yer verilmiştir. Sonuç ve öneriler bölümünde, bu çalışma ile ilgili elde edilen sonuçlara ve bu çalışma için geliştirilebilir önerilere değinilmiştir. Kaynaklar bölümünde, literatürdeki çalışmalar ve materyaller ile ilgili bilgilerin kaynağı sunulmuştur.

2. LİTERATÜRDEKİ ÇALIŞMALAR (LITERATURE STUDIES)

2016 yılında Mustafa Doğru tarafından "Ottoman-Turkish Optical Character Recognition and Latin Transcription" çalışmasında Osmanlı alfabesi ile yazılan belgelerin günümüzde anlaşılabilir olması için optik karakter tanıma motoru kullanılmıştır. Tanınamayan karakterleri Osmanlı alfabesi ile yazmak için Osmanlıca klavye geliştirilmiştir. Toplam 2762 karakter, optik karakter tanıma motoru ile %76,87 lik başarı sağlanmıştır [2].

Bilkent Üniversitesinde Ali Öztürk ve arkadaşları tarafından Osmanlı alfabesindeki karakterlerin elle veya tarayıcılar ile içi içe girmiş karakterleri ayrıştırılması ve tanınması için yapay sinir ağları ile çalışma yapılmıştır. Ağ, eğitim setindeki karakterleri yaklaşık %95'lik doğruluğu ile sınıflandırmayı başarmıştır [3].

Zeynep Kurt ve arkadaşları tarafından LDA (Linear Discriminant Analysis- Doğrusal Diskriminant Analizi) tabanlı Osmanlı Karakter Tanıma sistemi önerilmiştir. Sistem 99 örnekte %88 başarı sağlamıştır [4].

Ayşe Onat ve arkadaşları tarafında elle yazılan Osmanlı yazıların tanınması, bölünmesi ve özellik çıkartılması için OCR çalışması yapılmıştır. Bölünme ve özellik çıkarma aşamaları geometrik özellik analizine dayanır. HMM (Hidden Markov Model - Gizli Markov Modeli) önerilmiştir. Sistem yaklaşık %65 doğruluk sağlamıştır [5].

A. Enis Çetin ve arkadaşları tarafından Osmanlı Arşivleri için CBR (Content-Based Retrieval- İçeriğe Dayalı Erişim) sistemi önerilmiştir. Bu çalışmada, belgelerden çıkarılan semboller, denetimli bir şekilde oluşturulan sembol kütüphanesinde en benzer olanla eşleştirilmiştir. Doğruluk oranı %92 olarak sağlanmış [6].

Ertuğrul Başar ve arkadaşları tarafında yapay sinir ağları ile Osmanlı alfabesi ile yazılan karakterleri tanıma yöntemi önerilmiştir. Önerilen tanıma sistemi %85.5'lik sınıflama doğruluğu elde edilmiştir [7].

Esra Ataer ve Pinar Duygulu tarafından Osmanlı belgelerinin tanınması için karakter tanıma gerektirmeden bir yöntem önerilmiştir. Osmanlı kelimelerinin imgeler olduğu düşüncesiyle problemi görüntü alma problemi olarak ele alınmıştır ve görüntü eşleme tekniklerine dayalı bir çözüm önerilmiştir. Küçük basılı veri setiyle ilgili değerlendirmelerde %100 çizgi segmentasyonu ve %82 kelime segmentasyon performansları elde edilmiştir [8].

İsmet Zeki Yalnız ve arkadaşları tarafından Osmanlı belgelerinde bağlı karakterler için bağlara duyarlı bölme ve tanıma yöntemi önerilmiştir. Test sonuçlarında %90'dan daha yüksek doğruluk sonucu elde edilmiştir [9].

Tablo 1. Osmanlı Alfabeti-Türk Alfabeti (*Ottoman Alphabet-Turkish Alphabet*)

Tek başına	Sözcük sonunda	Sözcük ortasında	Sözcük başında	Adı	Günümüz Türkçesiyle
ا	ا	—		elif	a, e
ء		—		hemze	', a, e, i, u, ü
ب	ب	ب	ب	be	b
پ	پ	پ	پ	pe	p
ت	ت	ت	ت	te	t
س	س	س	س	se	s
ج	ج	ج	ج	cim	c
چ	چ	چ	چ	çim	ç
ح	ح	ح	ح	ha	h
خ	خ	خ	خ	hi	h
د	د	—		dal	d
ذ	ذ	—		zel	z
ر	ر	—		re	r
ز	ز	—		ze	z
ژ	ژ	—		je	j
س	س	س	س	sin	s
ش	ش	ش	ش	şin	ş
ص	ص	ص	ص	sat, sad	s
ض	ض	ض	ض	dat, dad	d, z
ط	ط	ط	ط	ti	t
ظ	ظ	ظ	ظ	zi	z
ع	ع	ع	ع	ayın	'a, h
غ	غ	غ	غ	gayın	g, ğ
ف	ف	ف	ف	fe	f
ق	ق	ق	ق	kaf	k
ك	ك	ك	ك	kef	k, g, ğ, n
گ	گ	گ	گ	gef	g, ğ
ن	ن	ن	ن	nef, sağır kef, nazal ne	n
ل	ل	ل	ل	lam	l
م	م	م	م	mim	m
ن	ن	ن	ن	nun	n
و	و	—		vav	v, o, ö, u, ü
ه	ه	ه	ه	he	h, e, a
لا	لا	—		lamelif	la
ی	ی	ی	ی	ye	y, i, î

İsmet Zeki Yalnız ve arkadaşları tarafında tarihsel, basılı ve el yazısı belgelerinde karakter bölme ve tanıma aşamaları için çeşitli yöntemler araştırılmıştır. Test sonucunda %90'lık doğruluk değeri elde edilmiştir [10].

Walter G. Andrews ve arkadaşları tarafından mevcut Osmanlı metinler için transkripsiyon yöntemlerine eleştirel bir gözle yaklaşılmıştır ve geri dönüştürülebilir transkripsiyon sistemi önerilmiştir [11].

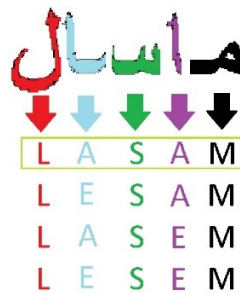
3. METOD VE MATERYAL (METHOD AND MATERIAL)

Bu bölümde, Osmanlı alfabesi ile yazılan kelimelerin Türk alfabesine çevrilmesi ile ilgili kullanılan metod ve materyallardan bahsedilmiştir.

Tesseract: Tesseract, çeşitli işletim sistemleri için özgür bir OCR (Optical Character Recognition - Optik Karakter Tanıma) motoru olarak geliştirilmiştir [12]. Tesseract motoru ilk olarak 1985-1994 yılları arasında Bristol, İngiltere ve Greeley, Colorado'da Hewlett Packard laboratuvarlarında tescilli bir yazılım olarak geliştirilmiştir. 1996'da Windows'a yapılan bazı değişiklikler ve 1998'de C'den C++'a bazı geçişlerle geliştirilmiştir [13]. Daha sonra 2005 yılında Hewlett Packard ve Nevada Üniversitesi, Las Vegas (UNLV) tarafından açık kaynak olarak piyasaya sürülmüştür. Tesseract'ın geliştirilmesi, 2006'dan beri Google tarafından desteklenmektedir [14]. 1995 yılında ilk üç OCR motorları içerisinde karakter tanıma doğruluğu sağlanmıştır [15]. Tesseract, Linux, Windows ve Mac OS X işletim sistemlerinde kullanılabilir [13]. Tesseract, 100'den fazla dil ve dil varyantı için, metin tanıma verileri, ek modüllerde mevcuttur. Geliştirici sürümü sadece Latin yazı tiplerini değil, aynı zamanda kırılma fontunu, Devanagari'yi (Hint betiği), Çince, Arapça, Yunanca, İbranice, Kiril ve diğer yazıları da desteklemektedir. [16]

Osmanlı Alfabesi: Osmanlı alfabesi (الفبا elifbâ), 10. ve 20. yüzyılına arasında Arapça tabanlı türkçe yazıdili olarak kullanılmıştır [1]. Osmanlı alfabesindeki her bir karakter bir veya birden çok Türk alfabesindeki harfe karşılık gelmektedir ve bu alfabedeki karakterler konum (başta, ortada, sonda ve tek başına) olarak başka şekiller alırlar [17]. Osmanlı alfabesindeki karakterlerin Türk alfabesindeki harflerin karşılıkları Tablo 1'de gösterilmiştir.

Yukarıdaki tabloda da görüldüğü gibi her karakter tek başına kullanıldığında veya kelime içerisinde başında, ortasında ve sonunda kullanıldığında farklı şekiller alırlar. Ayrıca her bir karakter latin alfabede birden çok harfe denk gelebilirler. Örneğin elif karakteri hem a hemde e harfi olarak kullanılabilir. Bu nedenle bu çalışmada çevrilmesi istenen Osmanlı alfabesi ile yazılmış bir kelime birden çok anlamlı veya anlamsız Türk alfabesi ile yazılmış kelimelere denk gelebilir. Örneğin Osmanlı alfabesi ile “masal” yazan bir kelime Türk alfabesine dört farklı anlamlı veya anlamsız kelimeye denk gelebilir. Bu kelimeler “masal”, “masel”, “mesal”, “mesel”dir. Osmanlı alfabesi ile yazılmış bir kelimenin karakterlerinin Türk alfabesindeki harflere karşılık gelebilecek kelimeler için örnek Şekil 1'de gösterilmiştir.



Şekil 1. Örnek masal yazan kelimeler dizisi (Example Tale Written Words Sequence)

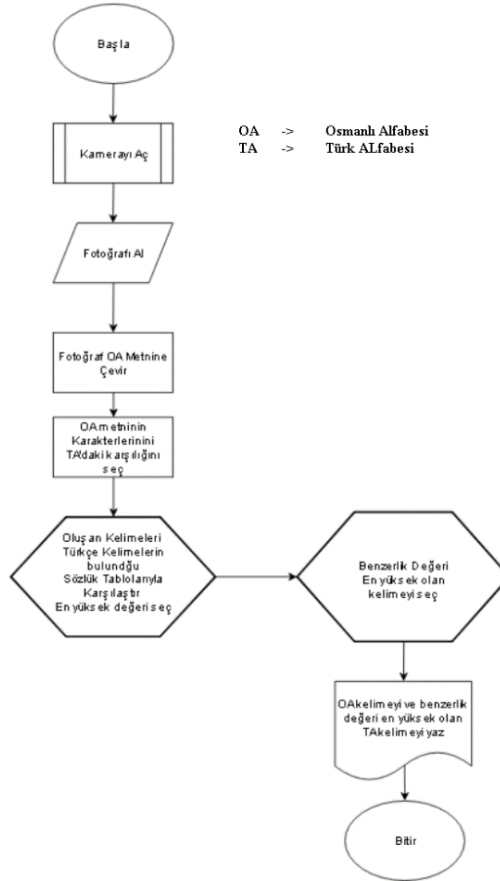
Bu çalışmada, görüntüsü düzenlenebilir metin halini alması sağlanmış bir Osmanlı alfabesi ile yazılmış kelimenin, Türk alfabesindeki karşılığının alına bilmesi için ilk önce arapça kelimenin içerisindeki karakterlerden birden fazla harfe karşılık gelenler belirlenmiş ve kaç farklı kelimeye denk gelebilecekleri hesaplanmıştır. Hesaplama işleminden sonra elde edilen değerde bir dizi oluşturulmuştur. Ardından arapça kelimedeki her bir karakterin Türk alfabesindeki harf karşılıkları ile benzersiz bir şekilde farklı kelimeler oluşturulmuştur. Sonunda birçok Türk alfabesi ile yazılmış anlamlı veya anlamsız kelimeler elde edilmiştir.

Sözlük Tablosu: İçerisinde nerdeyse tüm türkçe kelimelerin bulunduğu bu tablo, doğru veya doğruya en yakın kelimelerin bulunması için bu çalışmada kullanılmıştır. Bu sözlük tablosundaki kelimeler 1 den 15+ a kadar harf sayısına ve baş harflerine göre ayrı ayrı öbeklenmiştir.

Levenshtein Mesafesi: İki dizi arasındaki farkı ölçmek için kullanılır. İki kelime arasındaki Levenshtein mesafesi, bir kelimeyi diğerine dönüştürmek için gerekli olan en az sayıda tek karakterli düzenleme işlemleri olan ekleme, silme ve değiştirme ile sağlanır [18].

Kullanılan Materyaller: Bu çalışmada Android Studio ortamında Java kodları aracılığıyla programlama yapılmıştır. Osmanlı Alfabeti, optik karakter tanıma işlemi için Tesseract ve sözlük tablosundan benzerlik değerleri elde etmek için Levenshtein mesafesi kullanılmıştır.

Akış Diagramı: Bu çalışmada kullanılan akış diagramı Şekil 2’de sunulmuştur.



Şekil 2. Akış diagramı (Flow diagram)

Akış diagramında da görüldüğü gibi kamera açılır, kamera ile Osmanlı alfabesi ile yazılan kelimenin fotoğrafı çekilir, çekilen fotoğraf alınır, fotoğraf OCR işlemi ile kelime düzenlenebilir metin haline gelir, Osmanlı alfabesi ile yazılmış bu kelimedeki karakterlerin türk alfabesindeki harflerine karşılık gelenler ile Türk alfabesi ile yazılan kelime veya kelimeler elde edilir. Daha sonra bu kelimeler türkçe sözlükler tablosundan benzerlik değeri en yüksek olan seçilir ve bu kelime ekrana yazılmasıyla işlem tamamlanır.

Çeviri Yönteminin Görüntüsü: Bu çalışmada hazırlanan çeviri yöntemi ile denenmiş “kaynar kazan yakar” kelimelerinin fotoğrafının çekilmiş görüntüsü Şekil 3’de sunulmuştur.

قاینار قازان یاقار

Şekil 3. Fotoğrafi çekilmiş Osmanlı alfabesi ile yazılmış kelimeler (Photographs written with the Ottoman alphabet)

“kaynar kazan yakar” kelimelerinin bu çalışmada hazırlanan yöntemden geçtikten sonra elde edilmiş sonuç Şekil 4’te sunulmuştur.



Şekil 4. Çeviri yöntemi sonrası (After translation method)

Şekil 4’te de görüldüğü gibi Şekil 3’deki kelimelerin OCR işlemi gördükten sonra düzenlenebilir metin elde edilmiş ve “Eski Türkçe” bölümünde yazılmıştır. Bu düzenlenebilir kelimeler gerekli işlemleri gördükten sonra Türk alfabesindeki karşılıkları elde edilmiştir ve “Latin Harflerle” bölümüne yazılmıştır.

4. DENEYSEL ÇALIŞMA (EXPERIMENTAL STUDY)

Bu çalışmada fotoğrafı çekilmek istenen kelimeler beyaz zemin üzerine siyah renkte ve Osmanlı alfabesi yazılmış kelimeler test edilmiştir. Tesseract, karakter tanıma işlemleri gerçekleştirdikten sonra arapça metin elde edilmiştir. Osmanlı alfabesindeki karakterler Türk alfabesinde birçok harfe karşılık geldiğinden anlamlı anlamsız birçok kelime elde edilmiş ve o kelimeleri tutan dizi değişkeni kullanılmıştır. Bu kelimelerin içindeki doğru veya doğruya en yakın kelimeyi bulmak içinde türkçe kelimelerin olduğu bir sözlük tablosunda araştırılarak gerçekleştirilmiştir. Sözlük tablosu içerisindeki kelimeler, harf sayılarına göre ve baş harflerine göre farklı tablolara ayrıştırılmıştır.

Baş Harfe Göre Araştırma: Baş harfe göre araştırma yapıldığında sadece baş harfi ile aynı olan tablolarda araştırma yapılmıştır.

Harf Sayısına Göre Araştırma: Harf sayısına göre araştırma işlemi için harf sayısına göre 1’den 14’e kadar olanlar ayrı ayrı ve 15 ve üzeri olanlar ayrı bir şekilde tablolara ayrıştırılmıştır. Harf sayısına göre araştırma esnasında kelimenin uzunluğuna göre hangi tablolarda araştırma yapacağı Tablo 2’de sunulmuştur.

Tablo 2. Harf sayısına göre araştırma yapılacak tablolar (Tables to be researched by number of letters)

Kelime Uzunluğu	Araştırılacak Dosyalar
1-	1, 2, 3
2-	1, 2, 3
3-	2, 3, 4
4-	3, 4, 5
5-	4, 5, 6
6-	5, 6, 7
7-	6, 7, 8
8-	7, 8, 9
9-	8, 9, 10
10-	9, 10, 11
11-	10, 11, 12
12-	11, 12, 13
13-	12, 13, 14
14-	13, 14, 15+
15-	14, 15+
15+	15+

Bu şekilde araştırma işlemi yapılmasıyla birlikte optik karakter tanıma işlemleri esnasında oluşabilecek karakterlerde eksik algılanma veya fazla algılamaya durumların önüne geçildiği gözlenmiştir.

Araştırmaların Sonuçları: Araştırmaların sonuçları Tablo 3’de sunulmuştur.

Tablo 3. Araştırma sonuçları (Research results)

Toplam Kelime	Yöntem	Ortalama Süre	Başarı
100	Baş harfe göre	4,87 sn	80%
100	Harf sayısına göre	9,76 sn	81%

Başarımı ölçmek için yüz adet Osmanlı alfabesi ile yazılan kelimeler test edilmiştir. Baş harfe göre araştırma sonuçlarında, 4,87 saniye süre ile %80 lik doğruluk oranı elde edilmiştir. Harf sayısına göre ise 9,76 saniye süre ile %81.57 lik doğruluk oranı elde edilmiştir. Ortalama süre değeri, kelimelerin toplam sürelerinin, kelimelerin toplamına bölünmesi işlemi ile elde edilmiştir ve ortalama başarı değeri, kelimelerin başarı değeri toplamının, kelimelerin toplamına bölünmesi işlemi ile elde edilmiştir. Bu işlemler her iki yöntem için kullanılmıştır. Harf sayısına göre araştırma yönteminde birden çok tabloda araştırma yapıldığından geçen sürenin daha fazla olduğu gözlenmiştir. Baş harfe göre araştırma işleminde optik karakter tanıma işlemi esnasında baş harfin yanlış tanınması durumlarından ötürü başarı değerlerinin düşük kaldığı gözlenmiştir.

Metot Karşılaştırmaları: Osmanlı alfabesindeki karakterlerin tanınmasıyla ilgili bazı araştırmalar Tablo 4’te sunulmuştur.

Tablo 4. Metod karşılaştırmaları (Method comparisons)

Araştırmacılar	Yöntem	Başarı
Mustafa Doğru	OCR	77%
Ali Öztürk ve arkadaşları	Yapay Sinir Ağları	94%
Zeynep Kurt ve arkadaşları	LDA	88%
Ayşe Onat ve arkadaşları	HMM	65%
A. Enis Çetin ve arkadaşları	CBR	92%
Ertuğrul Başar ve arkadaşları	Yapay Sinir Ağları	86%
Esra Ataer ve Pinar Duygulu	Görüntü eşleştirme	82%
İsmet Zeki Yalnız ve arkadaşları	Bağlara duyarlı bölme ve tanıma	90%
İsmet Zeki Yalnız ve arkadaşlar	CBR	90%
Önder Özbek	OCR, Kelimelerin baş harfine göre	80%
Önder Özbek	OCR, Kelimelerin harf sayısına göre	81%

Osmanlı alfabesi ile yazılmış metinlerin içindeki karakterlerin anlaşılabilmesi için yapılan çalışmalarda, OCR sistemi dışında bazı yöntemlerle önerilmiştir. Bu metinler, elle yazılmış veya tarayıcılarla iç içe girmiş karakterler olmasından dolayı anlaşılması zor bir hal almıştır. Bu durumu giderebilmek sadece OCR yöntemi yetersiz hal almıştır. Bunun için, yapay sinir ağları, LDA (Linear Discriminant Analysis - Doğrusal Diskriminant Analizi), HMM (Hidden Markov Model - Gizli Markov Modeli), CBR Content-Based Retrieval - İçeriğe Dayalı Erişim) ve görüntü eşleştirme gibi yöntemleri önerilmiştir.

5. SONUÇ VE ÖNERİLER (CONCLUSION AND RECOMMENDATION)

Yüz adet örnek kelimenin testi sonunda harf sayısına göre, %81’lik başarı ve üç farklı tabloda araştırma yapıldığı için 9.76 saniyelik süre değerleri elde edilmiştir. Baş harfe göre, %80’lik başarı ile 4.87 saniyelik süre değerleri elde edilmiştir.

Fotoğrafi çekilmek istenen Osmanlı alfabesi ile yazılan kelime beyaz zemin üzerine siyah renkte yazılanlarda olumlu sonuçlar alınmıştır. Bu nedenle bu çalışma, çekilen fotoğrafın siyah beyaz renklere dönüştürülmesi yöntemi ile geliştirilebilir. Fotoğrafi çekilen kelimeler daha sonra türkçe kelimelerin olduğu sözlük tablolarında araştırılıp oradaki en benzer kelime seçilmiştir. Tablolarda bulunan kelimeler ek almamış kelimeler olduğundan

fotoğrafi çekilen kelimenin ek almış olması durumu doğruluk değerini etkilediği gözlenmiştir. Bu nedenle bu çalışma, ilk önce kelimeyi köküne ayırıp daha sonra tabloda araştırıp ve bulma yöntemi ile geliştirilebilir. Ayrıca tablolara, türkçe kelimelerin ek almış halleri şeklinde eklenmesi ile de bu çalışma geliştirilebilir.

KAYNAKLAR (REFERENCES)

- [1] Z. Korkmaz, “Alfabe Devriminin Türk Toplumunu Üzerindeki Sosyal Ve Kültürel Etkileri” International Periodical For the Languages, Literature and History of Turkish or Turkic Volume 4/3 Spring 2009
- [2] M. Doğru, “Ottoman-Turkish Optical Character Recognition and Latin Transcription” Yıldırım Beyazıt Üniversitesi, Ankara / Türkiye, 2016.
- [3] A. Öztürk, S. Güneş, Yüksel Özbay, “Multifont Ottoman Character Recognition”, Bilkent University, Computer Center, 65530, Ankara/Türkiye, Selçuk University, Electrical and Electronics Engineering ,42031, Konya / Türkiye.
- [4] Z. Kurt, H. İ. Türkmen, M. E. Karslıgil, “Linear Discriminant Analysis in Ottoman Alphabet Character Recognition”, Yıldız Teknik Üniversitesi İstanbul / Türkiye.
- [5] A. Onat, F. Yıldız, M. Gündüz, “Ottoman Script Recognition Using Hidden Markov Model”, World Academy of Science, Engineering and Technology 14 2006.
- [6] A. E. Çetin, E. Şaykol, A. K. Sinop, U. Güdükbay, Ö. Ulusoy, “Content-Based Retrieval (CBR) System for Ottoman Archives”, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 13, NO. 3, MARCH 2004.
- [7] E. Başar, N. Kılıç, P. Görel, B. Uçan, “Ottoman Character Recognition with Artificial Neural Networks and Development of Automatic Intelligent Translation System into Turkish”, Electrical, Electronics, Computer, Biomedical Engineers, 12th. National Conference, Eskişehir, Turkey, 2007
- [8] E. Ataer, P. Duygulu, “Matching Ottoman Words: An image retrieval approach to historical document indexing” in Proc. of ACM Int. Conf. on Image and Video Retrieval, ACM, 2007, New York, pp. 341–347, 2007
- [9] İ. Z. Yanlız, İ. Ş. Altıngövde, Ö. Ulusoy, “Integrated Segmentation And Recognition Of Connected Ottoman Script”, Bilkent Üniversitesi, Ankara / Türkiye, 2008.
- [10] İ. Z. Yanlız, İ. Ş. Altıngövde, Ö. Ulusoy, “Ottoman Archives Explorer: A Retrieval System for Digital Ottoman Archives” Bilkent Üniversitesi, Ankara / Türkiye, 2008.
- [11] W. G. Andrews, M. İnan, S. Kebeli, S. Waters, “Rethinking The Transcription Of Ottoman Texts. The Case For Reversible Transcription” , Turkish Studies International Periodical for the Languages, Literature and History of Turkish or Turkic Volume 5/2 spring 2010.
- [12] K. Anthony, "Tesseract: an Open-Source Optical Character Recognition Engine" Linux Journal: "linuxjournal.com", 2007. [Online]. Available: <https://www.linuxjournal.com/article/9676>.
- [13] V. Luc, "Announcing Tesseract OCR" "archive.org", 2006. [Online]. Available: "https://web.archive.org/web/20061026075310/http://google-code-updates.blogspot.com/2006/08/announcing-tesseract-ocr.html"
- [14] A. Tesseract OCR - The official Google blog
- [15] Rice Stephen V., F. R. Jenkins, and T. A. Nartker "The Fourth Annual Test of OCR Accuracy", "expervision.com/wp-content/uploads/2012/12/1995.The_Fourth_Annual_Test_of_OCR_Accuracy.pdf", 2013.
- [16] E. Bärwaldt: Buchstabensalat. Texterkennung mit Tesseract. In: LinuxUser. Nr.5. Linux New Media AG, 2011.
- [17] H. Develi, “Osmanlının Türkçe Klavuzu”, Genel Dağıtım Kitabevi, 2003.
- [18] Владимир И. Левенштейн (1965). Двоичные коды с исправлением выпадений, вставок и замещений символов [Silme, ekleme ve değiştirme ile düzeltebilen ikili kodlar]. Доклады Академий Наук СССР (Rusça). 163