



ISSN:1306-3111
e-Journal of New World Sciences Academy
2009, Volume: 4, Number: 2, Article Number: 1A0016

ENGINEERING SCIENCES

Received: December 2008

Accepted: March 2009

Series : 1A

ISSN : 1308-7231

© 2009 www.newwsa.com

Muammer Nalbant

Mustafa Burunkaya

Yılmaz Eroğlu

nalbant@gazi.edu.tr

University of Gazi

Ankara-Türkiye

OSMANLICA ELYAZISI HARFLERİ ÇEVİRİMİÇİ TANIMA

ÖZET

Bu çalışmada, hazırlanan arayüzdeki sınırlandırılmış alana fare ile çizilen Osmanlıca Harflerin tanınması ve Osmanlıca Metin Editörüne aktarılması sağlanmıştır. Arayüzün tasarlanması için Delphi 6.0 Görsel Programlama Dili kullanılmıştır. Tanınan harfler metin editörüne Unicode Karakter Sisteminde aktarılmıştır. Hazırlanan programda Yapısal Yaklaşım'a ait Yön Kodlaması Yöntemi kullanılmıştır. Harflerin genel doğru tanınma ortalaması %92'dir. Bu çalışmada öznitelik tabloları Nesih Hattına göre hazırlandığı için, bu hat üzerinde doğru sonuçlara ulaşılmaktadır. Fakat öznitelik tabloları diğer hatlara göre oluşturulduğunda bunların da tanınması sağlanabilmektedir. Tanınan karakterler, Osmanlıca olarak metin editörüne aktarılarak düzenlenebilecek hale getirilmiştir.

Anahtar Kelimeler: Çevrimiçi Karakter Tanıma, Osmanlıca Elyazısı, Yön Temelli Öznitelikler, Yön Kodlama, El Yazısı Tanıma

ONLINE HANDWRITTEN OTTOMAN CHARACTER RECOGNITION

ABSTRACT

The aim of this study is online recognition of Ottoman Letters that drawn on program interface by a mouse and transferring to Ottoman Text Editor. It was prepared with Delphi 6.0 Programming Language in the design of interface. The recognized letters were transferred to Ottoman Text Editor in Unicode Character Set. Direction Coding Method Related to Structural Approach was used in the program. General recognition rate of Ottoman Letters is 92%. In this work, as the feature tables are prepared according to Naskh Fonts, the correct results could be found on this font. If the feature tables are designed according to the other fonts, these fonts can also be recognized. Recognized letters are transferred to Ottoman Text Editor in an editable form.

Keywords: Online Character Recognition, Ottoman Handwriting Characters, Direction Based Features, Direction Coding, Handwriting Recognition

1. GİRİŞ (INTRODUCTION)

İnsanoğlu yıllardır kağıt ortamındaki el yazısı veya matbaa çıktısı metinleri ile eş zamanlı olarak bilgisayara girilen el yazılarını tanımak ve bilgisayar ortamında düzenlenebilecek hale getirmek için uğraş vermektedir. Bu tanıma işlemleri içerisinde en zor olan ve araştırmacıları da yıllardır meşgul eden el yazısının tanınmasıdır. Bu alanda birçok çalışma yapılmasına rağmen optimum bir çözüm üretilmiş değildir. El yazısı karakterlerin kullanıcıların yazım tarzlarına göre farklılık göstermesi, harflerin birleşmesi ve matbaa çıktılarına göre tam bir standardının olmayışı el yazılarının daha zor tanınmasının temel sebepleridir. Matbaa çıktıları üzerinde yapılan çalışmalar ise daha tatminkar gözükmetedir. Bunlardan Latin harflerinin matbaa çıktıları üzerinde elde edilen sonuçlar günümüzde yeterli düzeye ulaşmıştır. Fakat Çince, Japonca ve Osmanlıca Karakterler üzerinde sorun hala devam etmektedir. Özellikle Osmanlıca ve Arapça karakterlerin tanıtılması dilin yapısı ve yazım şekli göz önünde tutulduğunda oldukça zorlaşmaktadır. Hatta matbaada hazırlanmış bir Osmanlıca metnin tanınması el yazısı ile yazılmış bir Latince metinden daha zor olabilmektedir [1].

Osmanlıca matbuat metinleri üzerinde bazı çalışmalar olsa da Osmanlıca el yazısı metinler üzerinde şu ana kadar yapılmış tatminkar bir çalışma bulunmamaktadır [2]. Hal böyle olunca Osmanlıca el yazısı karakterlerin bilgisayar ortamında işlenebilecek hale getirilmesi problemi çok büyük önem arz etmektedir.

Bu çalışmanın amacı Delphi 6.0 Görsel Programlama Dili kullanılarak hazırlanan program arayüzüne çizilen Osmanlıca El Yazısı Harflerin çevrimiçi olarak tanınması ve Osmanlıca Metin Düzenleyiciye aktarılabilir hale düzenlenebilecek hale getirilmesidir.

2. ÇALIŞMANIN ÖNEMİ (RESEARCH SIGNIFICANCE)

İngilizce kaynaklar üzerinde yapılan taramada Arapça veya Farsça karakter tanıma konusunda oldukça fazla çalışmaya rastlanmıştır. Fakat Osmanlıca Karakterler konusunda yapılan çalışma azdır. Bu çalışmaların ortak özelliği genelde matbu metinler üzerinde çalışılmış olmasıdır. El yazması metinler üzerinde çok az çalışma vardır. Karakter tanıma konusunda ekseriyetle Saklı Markov Modeli, farklı yapay sinir ağları modelleri (İleri ve Geri Beslemeli Çok Katmanlı Algılayıcı, Adaptif Rezonans Teorisi Ağı, LQV Ağları, Kohonen Ağları vb.) ve Momentler Yardımıyla Optik Karakter Tanıma yöntemleri kullanılmıştır. Matbu metinler üzerinde yapılan bazı çalışmalarda mesafe kat edilmesine rağmen henüz tatminkar bir sonuç elde edilememiştir [2, 3, 4, 5, 6, 7 ve 8]. Bu çalışmada ilk olarak fare ile ekrana çizilen şekle ait yön bilgisi, başlangıç ve bitiş noktaları arasındaki eğim, noktaların harfe ve birbirine olan uzaklıkları gibi karakteristik özellikleri çıkarılmıştır. Daha sonra çıkarılan bu öznitelikler harf tablosundaki her bir harfe ait özelliklerle karşılaştırılmış ve bu öznitelikler hangisine uyuyorsa çizilen şekil o karakter olarak tanınmıştır. Tanımlanan özellikler ilgili harfe ait temel karakteristikler olduğu için genel kurallara göre çizilen her harf tanınmıştır.

3. OSMANLICA'NIN GENEL YAPISI VE LATİN HARFLERİNDEN FARKI (GENERAL STRUCTURE OF OTTOMAN LANGUAGE AND IT'S DIFFERENCE FROM LETTERS OF LATIN)

Osmanlıca'nın temel özellikleri aşağıda özetlenmiştir:

- Osmanlıca yazı (baskı veya el yazısı) bitişiktir ve sağdan sola doğru yazılır. Harfler, normalde birbirine temel çizgi üzerinde birleşir.
- Osmanlıca baskı ve el yazmaları, farsî temelde bitişik ve ayrık yazılır.

Arap alfabesi 28 harften oluşur. Osmanlıcada Arap harflerinin yanı sıra Farsça'daki p (پ), ç (چ) ve (ژ) harfleri de kullanılmıştır. Bu 31 harfin dışında Türkçe'deki ince g ünsüzünü belirtmek için kef harfine bir çizgi eklenerek gef, genizsi n ünsüzü için üç nokta eklenerek nef (sağır kef, kâf-ı nunî), lam ile eliften lamelif, hemze ile h harfinin ünlü şekli olan hâ-i resmiye harfleri oluşturulmuştur.

Dolayısı ile Osmanlıca'da 35 harf vardır. Bunlardan 25 adeti başta, 26 sı ortada ve 34 ü sonda olabilir. Bir harf müstakil yazılır, diğer harflerle birleşmez. Harflerden başka 10 adet rakam ve noktalama işaretleri, boşluklar ve özel semboller kullanılır (Şekil 1).



Şekil 1. Osmanlıca Harfler
(Figure 1. Letters of Ottoman Language)

- Harflerden bazıları, temel çizginin altında yer alır (mesela ر, ز)
- Harflerden bazıları, iki parçadan müteşekkildir (mesela ط, ظ, لا)
- Seslerin gösteriminde diacritic işaretler kullanılır. Ses diacriticlerinin varlığı veya yokluğu aynı kelimenin anlamını farklı kılar.
- Osmanlıca Harflerinde temel ve küçük harf ayrımı yoktur. Noktalama işaretlerinde kesin kurallar bulunmamaktadır. Osmanlıca Harfleri sözcüklerin başında, ortasında ve sonunda farklı biçimde yazılır (Şekil 2).

İsimleri	Harfler	Sonda (sağdan başlık)	Ortada (her iki taraftan başlık)	Başta (soldan başlık)
elif	ا	ا	ا	ا
hemze	ء	ء	ء	ء
be	ب	ب	ب	ب
pe	پ	پ	پ	پ
te	ت	ت	ت	ت
se	ث	ث	ث	ث
cim	ج	ج	ج	ج
çim	چ	چ	چ	چ
ha	ح	ح	ح	ح
hı	خ	خ	خ	خ
dal	د	د	د	د
zel	ذ	ذ	ذ	ذ
re	ر	ر	ر	ر
ze	ز	ز	ز	ز
je	ژ	ژ	ژ	ژ
sin	س	س	س	س
şın	ش	ش	ش	ش
sat	ص	ص	ص	ص
dat	ض	ض	ض	ض
tı	ط	ط	ط	ط
zı	ظ	ظ	ظ	ظ
ayın	ع	ع	ع	ع
gayın	غ	غ	غ	غ
fe	ف	ف	ف	ف
kaf	ك	ك	ك	ك
kef	ك	ك	ك	ك
gef	ك	ك	ك	ك
nef, sağır kef	ك	ك	ك	ك
lam	ل	ل	ل	ل
mim	م	م	م	م
nun	ن	ن	ن	ن
vav	و	و	و	و
he	ه	ه	ه	ه
lamelif	لا	لا	لا	لا
ye	ي	ي	ي	ي

Şekil 2. Osmanlıca Harflerin tek başına, başta, ortada ve sondaki halleri

(Figure 2. Isolated, initial, middle and end positions of Ottoman Letters)



- Osmanlıca Harflerinin Türkçe'deki zengin ünlü sistemini karşılamada yetersiz olduğu düşünülür. Örneğin Osmanlıca Alfabesindeki elif (ا) Türkçe'deki a ve e ünlüsünün karşılığıdır ya da Türkçe'deki u, ü, o, ö ünlülerinin yerine Osmanlıca'da yalnızca (و) harfi vardır, bu aynı zamanda v ünsüzünün de karşılığıdır.
- Herhangi bir konumda aynı büyüklüğe sahip olan çok az harf vardır fakat harfin konumu çizginin (çizginin yukarısında, çizginin aşağısında veya çizginin üstünde) yukarısında da farklı olabilir. Bazı harfler, kelimenin fonetiğine bağlı olarak kendi konumunda farklı şekillere sahip olabilir (mesela و, ا, ؤ, ! و,).
- Farklı Osmanlıca Harfler, tam olarak aynı şekle sahip olabilir ve birinden diğerine sadece tamamlayıcı karakter (harfle bütünleşen noktaların konumu ve sayısı) ilavesiyle ayrılırlar.
Noktalar: Harflerin önemli bir kısmı bir nokta, iki nokta ve üç nokta alır. Bu noktaların sayısı ve harfin üstünde veya altında olmasına göre harfler farklılaşır.
İşaretler: Harflerin üst veya alt kısmına ve bazen isteğe bağlı olarak yerleştirilirler.
- Karakterlerin genişliği, bir karakterden diğerine ve bir hattan diğerine değişir. Birçok karakter (35'in 21'i), karakterin gövdesi ve tamamlayıcı noktalar olmak üzere iki parçadan müteşekkildir. Noktalar, karakter gövdesinin üzerinde veya aşağısında bulunabilir. Noktalar, bir, iki veya üçlük gruplar halinde olabilir. Noktalı harflerin noktasız olanı da vardır. Diğer sesler, üstün, esre, ötre biçiminde çizgilerle nadiren gösterilebilir. Büyük harf veya küçük harf durumu mevcut değildir.
- Osmanlıca yazı bitişiktir ve kelimeler boşluklarla birbirinden ayrılmıştır. Bazı harfler (dal, zel, re, je, vav و, ذ, ر, ج, ج, و) bir sonraki harfle birleşmez. Bundan dolayı bu harflerden biri, kelime içinde mevcutsa, kelime iki alt kelimeye bölünür. Bu karakterler, alt kelimenin sadece sonunda görünür ve bundan sonra gelen harf, bir sonraki alt kelimenin baş harfini biçimlendirir.
- Osmanlıca yazı birçok fonta ve yazı biçimine sahiptir. Aynı fonttaki (hat) harfler farklı büyüklüklere sahiptir. Bundan dolayı sabit genişliğe dayalı bölümlene uygulanamaz.
HAT: Belli kurallara göre güzel yazı yazma sanatına hüsn-i hat (Arapça'da "hat" yazı, "hüsn" güzel) ya da kısaca hat, bu işle uğraşanlara hattat denir. Yüzyıllar boyunca Arap yazısı özellikle İranlı ve Türk sanatkarlarca geliştirilmiştir. Kullanılan kaleme, harflerin büyüklüğü ve şekline göre ya da iki yazının karıştırılmasıyla çeşitli yazı şekilleri ortaya çıkmıştır. Başlıca yazılar arasında küfî, muhakkak, râyhanî, sülüs, nesih, tevkî, rik'a, talîk, divanî sayılabilir. Tablo 1'de bazı hat örnekleri verilmiştir.

Tablo 1. Osmanlıca hat örnekleri
(Table 1. Different styles and fonts for the writing of Ottoman text) [1]

لا إله إلا الله محمد رسول الله
RİK'A (RUQ'A)
لا إله إلا الله محمد رسول الله
TA'LİK (NASTALIQ)
لا إله إلا الله محمد رسول الله
DİVANİ (DIWANI)
لا إله إلا الله محمد رسول الله
CELİ DİVANİ (ROYAL DIWANI)
لا إله إلا الله محمد رسول الله
REYHANİ (RAYHANI)
لا إله إلا الله محمد رسول الله
SÜLÜS (THULUTH)
لا إله إلا الله محمد رسول الله
KUFİ (KUFİ)
لا إله إلا الله محمد رسول الله
NESİH (NASKH)
لا إله إلا الله محمد رسول الله
BİLGİSAYAR ÇIKTISI NESİH (NASKH GENERATED BY COMPUTER)

4. EL YAZISI TANIMA (RECOGNITION OF HANDWRITTEN LETTERS)

Bilgisayar kullanılarak el yazısının tanınması bilgisayar biliminin başlangıcından bu yana araştırmacıları meşgul eden zor problemlerden biridir. Bu konuda çok sayıda çalışma yapılmasına rağmen halen problem devam etmektedir. Yazı içerisinde farklı karakterlerin benzerliği, harflerin birleştirilmesi, bilgilerin bulanıklığı ve çok çeşitliliği bu problemi zor kılan unsurlardandır.

El yazısı tanıma konusunda temel olarak iki sistem bulunmaktadır. Bunlar çevrimiçi ve çevrimdışı sistemlerdir [9]:

Çevrimiçi sistemler, el yazısını yazı yazıldığı sırada tanıyan, özel olarak tasarlanmış sistemlerdir. Genelde elektromanyetik ya da elektrostatik tabletler kullanılır. Kalemin dokunuşları ve hareketlerin devamlılığı göz önünde tutulur. Elektronik ajandalar (PDA) gibi günümüzde çok yaygınlaşan bir yöntemdir. Çevrimdışı sistemler ise genelde kağıt üzerine daha önceden yazılmış bilgilerin sayısallaştırılarak, sonradan tanınmaya çalışılması işlemidir.

Çevrimiçi sistemlerin yazı hızına yetişebilmesi için çok hızlı olması gerekirken, çevrimdışı sistemlerde yazının tanınması için bir



süre kısıtlaması yoktur. Çevrimiçi sistemlerin avantajı harflerin şekil özelliklerinin yanında yazılma sırasındaki hareketlerin de gözlenebilmesidir. Ayrıca kullanıcıyla etkileşimin en büyük faydalarından biri gerektiğinde yanlışların anında düzeltililebilmesidir. Çevrimdışı sistemlerde, yazının yazılması sırasındaki hareketler hakkında hiçbir bilgi olmadığı ve eski belgeler yeterince temiz ve okunaklı olmadığı için hata oranı daha yüksektir.

Sayısallaştırıcılardan kaynaklanan gürültülerin etkisini azaltmak için çok daha detaylı bir ön işlemeye gerek duyulur. Ancak bu sistemlerin avantajı özel bir alete gerek duyulmaması ve böylece yıllardır var olan bütün belgelere uygulanabilmesidir [10].

El yazısı karakterlerin tanınmasında kullanılan yöntemler iki ana grupta toplanabilir:

- Holistik yöntemler, bir harfi analiz etmeden tüm kelimeyi tanımayı amaçlar. İnsanın okuma aktivitesinin incelenmesi yönünde yapılan son çalışmalar, genel tanıma temelinde bir sürecin mümkün olduğunu göstermektedir. Bu yaklaşımlar sözlüğün boyutuna bağlı olarak sistemde bir karmaşıklık artışına yol açabilmektedir. Bu açıdan tanıma genellikle sınırlı sözlükler üzerinde gerçekleştirilmektedir. Holistik yöntemler iki alt grupta toplanır:
 - o Dinamik programlamayı kullanan mesafe ölçmeye dayalı yöntemler.
 - o ii.Probalistik çerçeve çalışmasına dayalı yöntemler (Saklı Markov Modeli) [1].
- Analitik yöntemlerde ise el yazısı karakteri tanıma süreci bir kelimeyi onun harfleri ile tanır ve kelimeyi tanımak için her harfin yerini içerir. Bu yaklaşımda her kelime, sözlük kullanılmadan tanınabildiği için oldukça ilgi çekicidir. Fakat buradaki temel sorun el yazısının dilimlenmesi problemi. Yapılmış birçok çalışma olmasına rağmen dilimleme yalnız sınırlı uygulama alanları için geçerlidir. Son zamanlarda çalışmalar bu iki yöntemin birleştirilerek tanıma oranlarında ilerleme elde edilebileceğini göstermektedir [7].

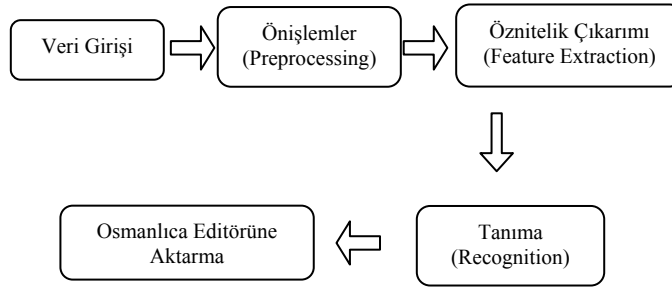
El yazısının taranmasında holistik yöntemlerde genellikle el yazısının genel görünüşünü, sınırlarını belirledikten sonra el hareketinin üst ve alt satır sonlarının (overstrokes) ve kelime uzunluğunun belirlenmesine çalışılır. El yazısı bu hareketlerin yöne ilişkin artması ve azalması şeklinde yorumlanabildiği için, bir kelimeye ait temel bileşenleri olarak kullanılabilir. Bu nedenle söz konusu özellikler kelimeye dayanak (anchor) noktası olarak seçilmektedir. Kelimenin çizim sınır hattı ile ortalama ekseninin kesişme noktası dayanak noktasını belirlemektedir. Önemli bilginin gövdenin alt ve üst kısımları arasındaki geçiş bölgesinde yerleştiği hipotezi olmaktadır. Kelime eksenini bu iki bölümü ayırır ve dayanak noktası bu bölümler arasındaki geçişleri gösterir. El yazısı kelimenin ferdi bileşenleri, genellikle harflere benzerdir [7].

5. OSMANLICA ELYAZISI HARFLERİ ÇEVİRİM-İÇİ TANIMA (ONLINE RECOGNITION OF OTTOMAN HANDWRITTEN LETTERS)

Bu çalışmanın amacı, Delphi 6.0 görsel programlama dili kullanılarak hazırlanan program arayüzüne fare yardımıyla çizilen Osmanlıca el yazısı harflerin tanınması ve Osmanlıca metin editörüne aktararak düzenlenebilecek hale getirilmesidir.

Çalışma aşağıda verilen aşamalar takip edilmiştir:

- Ön işlemler (Preprocessing)
- Öz nitelik çıkarımı (Feature extraction)
- Karakterin tanınması (Recognition)
- Editöre aktarılması (Şekil 3).



Şekil 3. Osmanlıca Elyazısı Karakterleri çevrim-içi tanıma sisteminin aşamaları

(Figure 3. Levels of online recognition Ottoman Script Letters)

Önişlemler aşamasında sınırlandırılan alan içerisinde çizilen şeklin özellikleri çıkarılmıştır. Çıkarılan özellikler harf tablosundaki özelliklerden hangisine uyuyorsa çizilen şekil o karakter olarak tanınmıştır. Tanımlanan özellikler o harfi tanımlayan temel karakteristikler olduğu için genel kurallara göre çizilen her harf tanınmıştır. Tanınan karakter de Osmanlıca olarak editöre aktarılarak düzenlenebilecek hale gelmiştir. Sonuçlarda %100'e varan tanıma oranlarına ulaşılmıştır. Genel harf tanıma oranı ise %92'dir.

5.1. Önişlemler (Preprocessing)

İlk olarak programda harfin girileceği alan, 150x100 piksellik bir alanla sınırlandırılmıştır. Çizilen harflerde hassasiyeti artırmak için farenin hareketi yavaşlatılarak fare hızından dolayı yakalanamayan piksellerin önüne geçilmiştir. Böylece el titremesinden kaynaklanan gürültülerin de önlenmesi sağlanmıştır.

5.2. Öznitelik Çıkarımı (Feature Extraction)

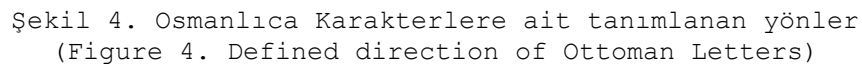
Bu aşamada çizilen şekile ait temel karakteristik özellikleri çıkarılmıştır. Bunlar:

- Çizime ait tüm koordinat değerleri
- Çizime ait dört yönlü sınır değerleri
- Nokta sayısı
- Noktaların birbirine uzaklıkları
- Noktaların harfe uzaklıkları
- Başlangıç ve bitiş noktası arasındaki eğim
- Harfe ait yön bilgileridir.

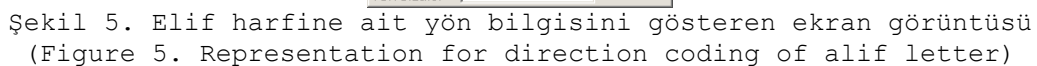
İlk olarak kullanıcı harfi çizerken fare hareketlerine göre piksel bilgileri x ve y koordinatları için ayrı ayrı tutulmuştur. Bu piksel bilgileri ile dört yönlü sınır değerleri, nokta sayısı, noktaların birbirine ve harfe uzaklıkları, başlangıç ve bitiş noktaları arasındaki eğim hesaplanmıştır [11]. Karakteristik özellikler içerisinde en temel özellik ise yön bilgisidir.

5.3. Yön Bilgisinin Hesaplanması (Calculation of Direction Coding)

Yön bilgisinin belirlenmesi için yön kodlaması tekniği kullanılmıştır [9]. Eş zamanlı olarak kullanıcının fare hareketlerinin yönleri belirlenmiştir. Bunlara Şekil 4'de görülen ana yönler için 1, 2, 3, 4; arayönler için ise 5, 6, 7, 8 numaraları verilmiştir.



Yapılan denemelerde 150x100 piksellik bir alanda geri karşılaştırma aralığının 15 birim, fark olarak karşılaştırılacak eşik değerinin ise 7 birim olduğu gözlemlenmiştir. Örneğin Mouse ile ekrana "ELİF" harfi çizildiğinde yön bilgisi 2 olacaktır (Şekil 5).



- x' 'te eşik değerinin üstünde artış var
- y' 'de eşik değerinin üstünde azalış var
- Yön: 8

155

Kullanıcı fareyi hareket ettirirken y değeri 15 piksel öncesine göre eşik değerinin üzerinde azalış göstermiş, fakat x değerinde eşik değeri (7 birim) kadar değişim görülmemiştir ve kullanıcının 2 yönünde hareket ettiği anlaşılmıştır.

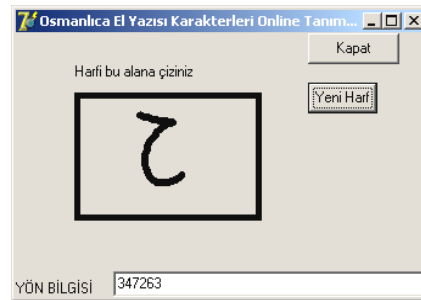
Harflerin yön bilgilerinin x ve y koordinatlarına göre hesaplama kriterleri Tablo 2’de verilmiştir.

Tablo 2. Yön bilgileri hesaplama kriterleri
(Table 2. Calculation criteria of direction coding)

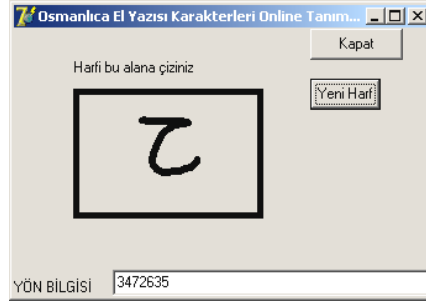
Yönler	Eşik değeri kadar değişim var mı?	
	X	Y
1	Yok	Artış
2	Yok	Azalış
3	Artış	Yok
4	Azalış	Yok
5	Artış	Artış
6	Artış	Azalış
7	Azalış	Azalış
8	Azalış	Artış

Her harfe ait bu şekilde yön bilgileri hesaplandığında çoğu harf için kullanıcı çizim sitillerine göre birden fazla yön bilgisine ulaşılmaktadır. Fakat her harf içinde harflere ait genel karakteristik özelliklerinden kaynaklanan temel yönler bulunmaktadır.

Bu farklı yön bilgilerini karakterin tanınması aşamasında kullanabilmek için 10 farklı kullanıcıya her harf için tatminkar sonuçlar elde edilinceye kadar çizimler yaptırılmıştır. Sonuçlar içinden harfin genel karakteristik özelliklerine uymayanlar elenmiş, doğru karakterlerin yön bilgileri her harf için yön bilgisi tablosuna işlenmiştir. Örneğin Şekil 7 ve Şekil 8’de HA harfinin sırası ile farklı (347263) ve (3472635) yön bilgileri verilmiştir.



Şekil 7. Ha harfine ait (347263) yön bilgilerinden biri
(Figure 7. One of direction coding of Ha letter)



Şekil 8. Ha harfine ait (3472635)yön bilgilerinden bir başkası
(Figure 8. Other one direction coding of Ha letter)

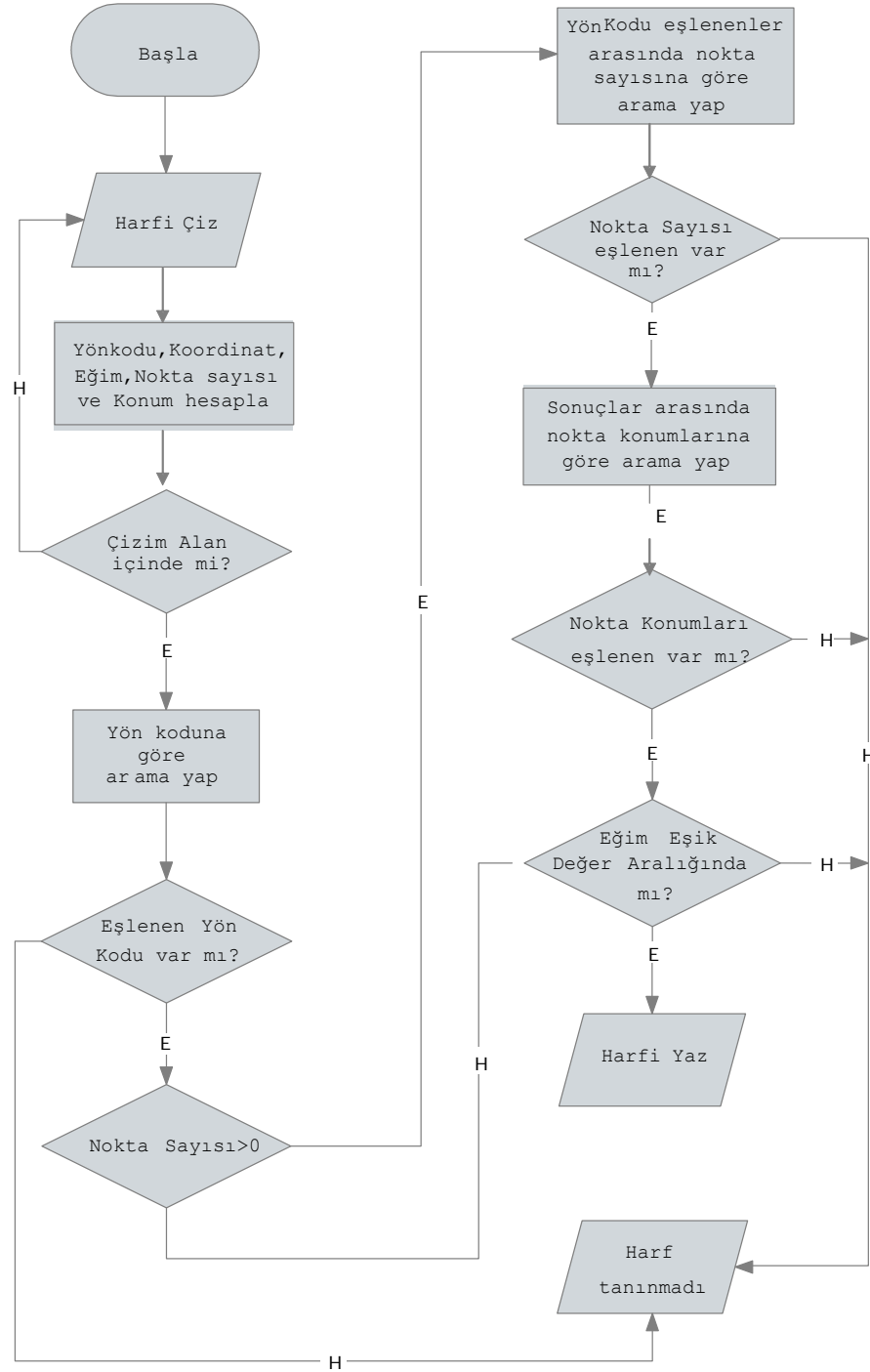
Öznitelik olarak harfe ait yön bilgisinden başka harfin dört yönlü sınır değerleri de hesaplanarak $enbuyukx$, $enbuyuky$, $enkucukx$, $enkucuky$ değişkenlerine aktarılmaktadır. Ayrıca harfte bulunan nokta sayısı ve bunlarında birbirine ve harfe olan uzaklıkları bulunmaktadır. Öznitelik çıkarma aşamasında hesaplanan ve tabloya kaydedilen değerlerin doğruluğu ölçüsünde tanıma aşamasındaki oranlarında yükseldiği görülmektedir.

5.4. Karakter Tanıma (Recognition of Letter)

Bu aşamada çizilen karakterler karakter tablosuna girilen değerlerle karşılaştırılmış, doğru değerlere sahip olanlar tanınmış ve harfin adı formun tanınan harf kısmına yazılmıştır (Şekil 9). Tanınmayan harf için "Tanımsız" ifadesi yazılmıştır. Tekrar harf denemesi yapmak için "YENİ HARF" butonu konulmuştur. Bu butona basıldığında çizimle beraber tüm değişkenler sıfırlanmaktadır.

Farenin sol tuşuna basılıp harf çizilmeye devam edildiği süreçte yön bilgileri hesaplanmıştır. Sol tuş bırakıldığında ise ilk olarak diğer öz nitelikler hesaplanıp sonra harf tanıma aşamasına geçilmektedir.

Tanıma aşamasında ilk olarak noktasız harfler tablodan aranarak sonuç yazılmakta, daha sonra nokta sayısı, noktaların harfe ve birbirine olan uzaklıklarına göre noktalı karakterler arasında arama yapılmaktadır. Uygun sonuç bulunursa ekrana yazılmakta, eşleşen sonuç olmadığı zaman ise ekrana "Tanımsız" uyarısı gelmektedir.



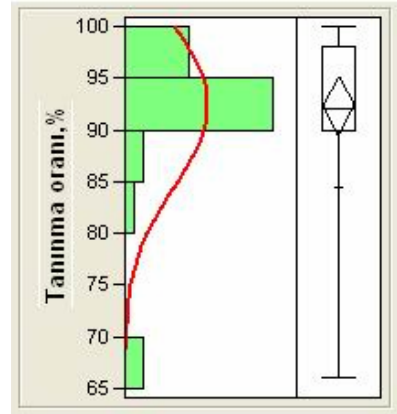
Şekil 9. Osmanlıca Elyazısı Harfleri Çevrimiçi Tanıma Akış Şeması
(Figure 9. Flowchart of Online Handwritten Ottoman Character Recognition)

Tanınma oranlarının belirlenmesi için 10 farklı kullanıcıdan her harf için 10 doğru çizim yapması istenmiştir. Yanlış yapılan çizimler tekrarlatılmış, doğru yapılan çizimlerden ise sadece tanınanlar çizelgeye "Tanındı" olarak kaydedilmiştir. Böylece her harfe ait 100 geçerli girişe göre tanınma oranları belirlenmiştir. Tablo 3'de bu bilgilere göre hazırlanmış harflerin tanınma oranları verilmiştir.

Tablo 3'e ek olarak tanınma oranları Şekil 10'da Dağılım ve Kutu Grafiği ile sunulmuştur. Soldaki bölüm Dağılım; sağdaki bölüm ise Kutu Grafiksel Gösterimini ifade etmektedir. Görüldüğü gibi %65-70 tanınma aralığında sadece iki harf bulunmaktadır. %70-80 tanınma aralığında harf bulunmamaktadır. Diğer 33 harf ise %80'in üzerinde tanınma oranına sahiptir. Harflerin genel tanınma ortalaması %92'dir. Harflerin büyük çoğunluğu bu ortalamanın etrafında tanınma oranları ile yer almıştır.

Tablo 3. Osmanlıca Elyazısı Harflerin tanınma oranları
(Table 3. Recognition rates of Ottoman Handwritten Letters)

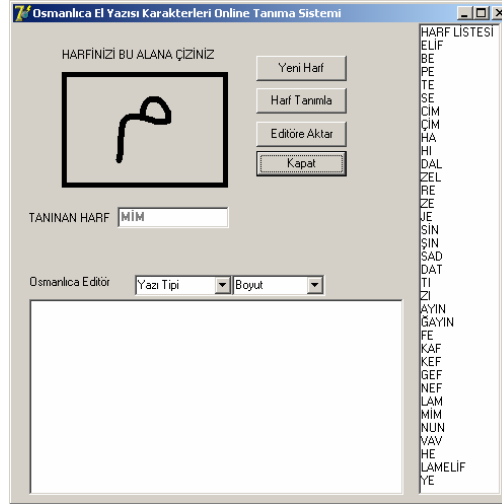
Harf	Tanınma Oranı (Yüzde)	Harf	Tanınma Oranı (Yüzde)
ELİF	100	DAT	87
HEMZE	96	TI	92
BE	98	ZI	93
PE	98	AYIN	90
TE	97	ĞAYIN	90
SE	98	FE	92
CİM	92	KAF	92
ÇİM	92	KEF	91
HA	94	GEF	90
HI	94	NEF	90
DAL	100	LAM	98
ZEL	100	MİM	92
RE	100	NUN	95
ZE	100	VAV	95
JE	100	HE	82
SİN	69	LAMELİF	91
ŞİN	66	YE	92
SAT	89		



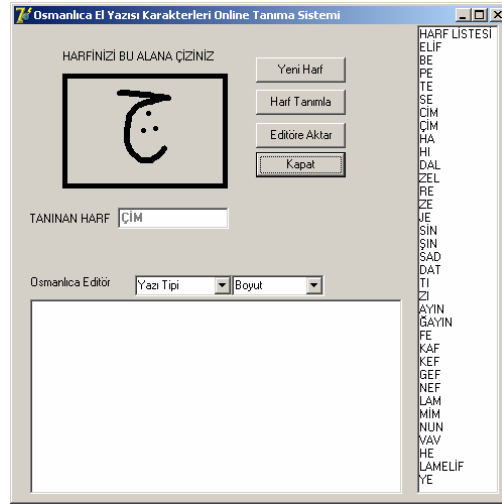
Şekil 10. Harf tanınma oranları ile ilgili dağılım grafiği ve kutu grafiği

(Figure 10. Distribution Graph and Box Plot of Letters Recognition)

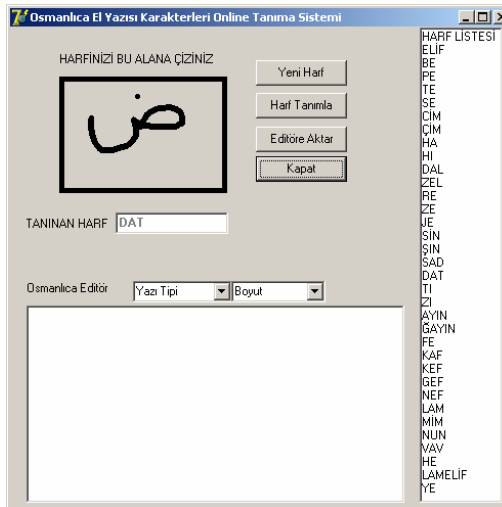
Şekil 11, Şekil 12 ve Şekil 13'de tanınan bazı harflere ait örnekler verilmiştir.



Şekil 11. Mim harfinin tanınması
(Figure 11. Recognition of Miim letter)



Şekil 12. Çim harfinin tanınması
(Figure 12. Recognition of chim letter)



Şekil 13. Dat harfinin tanınması
(Figure 13. Recognition of Dad letter)

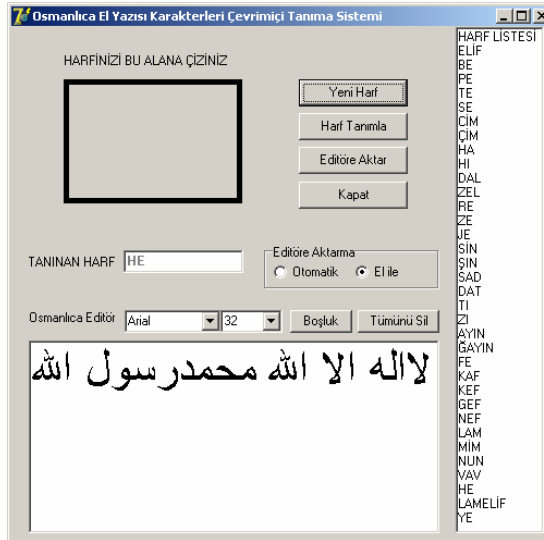
5.5. Harflerin Osmanlıca Metin Editörüne Aktarılması (Transferring of Ottoman Letter to Text Editor)

Bu aşamada da tanınan harfler UNICODE karakter setindeki karşılıklarına göre editöre aktarılmakta ve düzenlenebilecek hale gelmektedir. UNICODE karakter sisteminde karakterlerin sağdan ve soldan birleşme özellikleri tanımlı olduğu için harflerin birleşmesinde problem yaşanmamakta ve editöre eklenen harfler kendi özelliklerine göre sağa veya sola birleşebilmektedir. Şekil 14'de ŞIN harfinin editöre eklenmiş hali görülmektedir.



Şekil 14. Şin harfinin Osmanlıca Metin Editörüne aktarılması
(Figure 14. Transferring of shiin letter to Ottoman Text Editor)

Şekil 15'te cümledeki sırasıyla çizilen harfler editöre eklenmiştir.



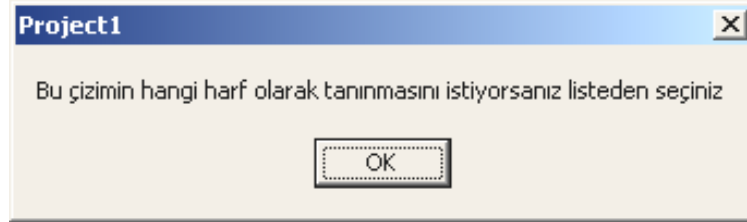
Şekil 15. Bir cümle editöre aktarılmış hali
(Figure 15. A sentence in the editor)

5.6. Program Düğmelerinin Tanıtımı (Presentation of Programme Button)

Şekil 16' da program düğmeleri görülmektedir. Bunların görevleri aşağıda verilmiştir:

- **Yeni Harf:** Çizim alanını temizler ve yeni harf girişi için tüm değişkenleri sıfırlar.

- **Harf Tanımla:** Çizilen şeklin bir harfe ait özelliklere eklenmesini sağlar. İlk olarak tanımlanacak harf çizilir. Daha sonra "Harf Tanımla" butonuna basıldığında Şekil 15'deki uyarı gelir. Bundan sonra tanımlanacak harf listeden seçilerek işlem gerçekleştirilir.



Şekil 16. Harf tanımlama uyarısı
(Figure 16. The warning of letter definition)

- **Editöre Aktar:** Tanınan karakterin Osmanlıca editöre aktarılmasını sağlar.
- **Kapat:** Programı sonlandırır.

6. SONUÇ VE ÖNERİLER (RESULT AND SUGGESTIONS)

Kullanıcının çizim yapacağı arayüzdeki alan 150x100 piksel ile sınırlandırılmıştır. Kullanıcının el titremeleri ve ani hareketlerinden etkilenmemesi ve gürültülere neden olmaması için farenin hızı yavaşlatılmıştır. Böylece farenin hızından dolayı yakalanamayan ve atlanan piksel değerlerinin de önüne geçilmiştir. Kullanıcı fare ile arayüze çizimini yaparken eş zamanlı olarak hangi yönler gittiği hesaplanmıştır. Bunun için kullanıcının takip ettiği x ve y koordinatları bir diziye aktarılmıştır. Yön değişimlerinin bulunabilmesi için aktif x ve y değerleri geri değerlerle karşılaştırılmış, belirli bir eşik değerini aştığında ise yön değiştirildiği anlaşılmıştır, bu değer de harfin yön bilgisine kaydedilmiştir. Yapılan denemeler sonucunda 150x100 piksellik bir alana çizilen harfe ait uygun geri karşılaştırma aralığının 15 birim; aktif değer ile karşılaştırılan değer arasındaki uygun eşik değerinin ise 7 birim olduğu gözlenmiştir.

Harfe ait yön bilgisinden başka başlangıç ve bitiş pikselleri arasındaki eğim, noktaların konumu, adedi, birbirine ve harfe olan uzaklık değerleri tutulmuştur. Çizim anında hesaplanan değerler daha önce hazırlanan harf tablosundaki değerlerle karşılaştırılmış, uygun eşleşmeler sağlandığında harf tanıma gerçekleşmiştir.

Sistemin önemli özelliklerinden biri de harflere ait karakteristik özelliklerin tutulduğu tablonun yine hazırlanan arayüz ile oluşturulmasıdır. Burada en temel özellik harfin yön bilgisidir. Harf karakteristik tablosunun hazırlanması için Osmanlıca yazmayı ileri derecede bilen 10 farklı kişiye tüm harfler, doğru tanıma oranına ulaşıncaya kadar farklı şekillerde arayüze çizdirilmiştir. Harf tablosu Nesih hattına göre oluşturulmuştur. Çizilen her doğru şeklin özellikleri tabloya işlenmiştir. Tabloya girilen değerler o harfe ait genel karakteristikler olduğu için genel yazım kurallarına göre Nesih Hattında çizilen her harf tanınmıştır.

Karakteristik tablosu oluşturulurken çizilen harfte yön bilgisi karmaşıklaştığında ve çok fazla yön değişimine sahip olan harflerde tanıma oranı azalmaktadır. Tanımlamada en çok karıştırılan harf Sin ve Şin harfleri olmuştur. Sin ve Şin harflerinin tanıma ortalaması %67,5 tir. Elif, Dal, Zel, Re, Ze, Je harflerinin tanıma ortalaması %100 dür. Tüm harflerin genel tanıma ortalaması ise %92'dir. Harflere ait tanıma oranları özet olarak Tablo 4'de verilmiştir.

Tablo 4. Osmanlıca Elyazısı Harflerin tanınma oranları
(Table 4. The recognition rates of Ottoman Handwritten Letters)

Tanınma Oranları (%)	Harfler
100	Elif, Dal, Zel, Re, Ze, Je
95-99	Hemze, Be, Pe, Te, Se, Lam, Nun, Vav
90-94	Cim, Çim, Ha, Hı, Tı, Zı, Aydın, Ğayın, Fe, Kaf, Kef, Gef, Nef, Mim, Lamelif, Ye
85-89	Sad, Dat
80-84	He
65-70	Sin, Şın

Sonuç olarak öznitelik çıkarımında harfe ait karakteristik özelliklerin doğruluğu arttıkça, yön bilgileri gereği kadar artırıldığında ve fare ile çizilen piksellerin yakalanma oranı artırıldığında daha yüksek tanınma oranları elde edilebilecektir.

Hazırlanan programa kelime bazında da tanınma özelliği eklenebilmesi için, öznitelik tablolarında harflerin başta, ortada ve sondaki halleri de girilmeli veya kelime bölütleme özelliği eklenmelidir. Program mevcut haliyle kişisel bilgisayarlarda (PC), Elektronik tahtalarda, Tablet PC'lerde kullanılabilir durumdadır. Cep telefonları ve avuçiçi bilgisayarlarda (PDA) ise gerekli uyarlamalar yapılarak kullanılabilir.

KAYNAKLAR (REFERENCES)

1. Amin, A., (1998). Offline Arabic Character Recognition: The State Of Art. School of Computer Science and Engineering: 31(5):520-521.
2. Öztürk, A., (1998). Osmanlıca Karakterlerin Bilgisayar Destekli Tanınması. Yüksek Lisans Tezi. Gebze Yüksek Teknoloji Enst.ss:8-15.
3. Özçilingir, M., (1990). An Ottoman Character Recognizer. Yüksek Lisans Tezi. ODTU Bilgisayar Mühendisliği, Ankara, ss:1-42.
4. Atıcı, A., (1994). Segmentation, Feature Extraction and Recognition of Ottoman Script. Yüksek Lisans Tezi. ODTU Bilgisayar Mühendisliği, Ankara, ss:90-98.
5. Amin, A., Al-Sadoun, H., (1994). Handprinted Arabic Character Recognition System. Proc.12th Int.Conf.on Pattern Recognition, pp:536-539.
6. Amin, A. and Mansoon, W., (1997). Recognition of Printed Arabic Text Using Neural Networks. Proc.4th Int.Conf.on Document Analysis Recognition, Germany.
7. Nabiye, V.V., (2005). Yapay Zeka. Seçkin Yayınevi, pp:534-570.
8. Amin, A., Kaced, A., Haton, J., Moher, R., (1980). Handwritten Arabic Character Recognition by the I.R.A.C System. In Proc.5th Int.Joint Conf.Pattern Recognition, pp:729-731.
9. Plamondon, R. and Srihari, S.N., (2000). Online and Off-line Handwriting Recognition: A Comprehensive Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1):63-84.
10. Duygulu, P., Ören, T. ve Çölkesen R., (2006). El Yazısı Tanıma. Türkiye Bilişim Ansiklopedisi, İstanbul: Papatya Yayıncılık, ss: 369-372.
11. Aizawa, H., Wakahara, T., and Odaka, K., (1997). Realtime Handwritten Character String Segmentation Using Multiple Stroke Features. IEICE Trans.Inf. & Syst, 80(5): 1178-1185.



12. Okamoto, M. and Yamamoto, K., (1997). On-line Handwriting Character Recognition Method with Directional Features and Direction-Change Features. Fourth International Conference Document Analysis and Recognition (ICDAR'97). pp:926.
13. Blumenstein, M., Liu, X.Y., and Verma, B., (2007). An Investigation of the Modified Direction Features for Cursive Character Recognition. Pattern Recognition: 40(2): 376-388.