

Görüntü işleme teknikleri kullanılarak multimedya görüntülerinde alt yazı bölgelerinin tespit edilmesi

Erdal Güvenoğlu ^{*1}

¹Maltepe Üniversitesi, Meslek Yüksekokulu, Bilgisayar Programcılığı Bölümü, İstanbul, *ORCID: 0000-0003-1333-5953*

Geliş Tarihi:02.11.2023

Kabul Tarihi:25.12.2023

Özet

Mobil cihazların ve multimedya teknolojilerinin yaygın olarak kullanımı ile birlikte görüntülerin elde edilmesi çok daha kolay hale gelmiştir. Bununla birlikte görüntüler içerisinde yer alan alt yazıların elde edilmesi ve bunların farklı amaçlar için kullanımı bir problem olarak ortaya çıkmıştır. Bu çalışmada multimedya görüntülerinde yer alan altyazıların bulunduğu bölgelerin tespit edilmesi için kullanımı basit ve etkili bir yöntem önerilmiştir. Yöntem birbirini takip eden farklı adımlardan oluşmaktadır. Giriş görüntüsü olarak verilen 24 bit renkli bir görüntüler üzerinde görüntü işleme teknikleri yardımıyla metin olan alt yazı bölgesine ait koordinatlar belirlenmektedir. Ardından renkli görüntü üzerinde işaretlenmektedir. Birbirinden farklı özellik ve ölçülerde görüntüler üzerinde deneysel çalışmalar gerçekleştirilmiştir. Çalışmanın gerçekleştirilmesinde köşe noktaların işaretlenmesi amacıyla Harris köşe saptama algoritması, gürültülerin giderilmesi için gauss filtreleme ve morfolojik görüntü işleme teknikleri kullanılmıştır. Gerçekleştirilen çalışmalarda %94 oranında bir başarımla elde edilmiştir. Süre ölçüm testlerinde ise ortalama olarak 1.56 sn gibi iyi bir başarımla süresine ulaşılmıştır. Süre ölçümleri literatürdeki diğer çalışmalar ile karşılaştırılmıştır. Önerilen yöntemin, süre bakımından oldukça iyi bir performansa sahip olduğu görülmüştür.

Anahtar Kelimeler: Altyazı tespiti, Görüntü işleme, Kenar belirleme, Köşe saptama, Morfolojik görüntü işleme

Detecting subtitle regions in multimedia images using image processing techniques

Abstract

With the widespread use of mobile devices and multimedia technologies, the acquisition of images has become much easier. However, obtaining the subtitles in the images and using them for different purposes have emerged as a problem. In this study, a simple and effective process is proposed for detecting the regions where subtitles are present in multimedia images. The process consists of different successive steps. The coordinates of the subtitle region, which is the text, are determined with the help of image processing techniques on a 24-bit colour image given as an input image. Then, it is marked on the colour image. Experimental studies were carried out on images with different features and sizes. Harris corner detection algorithm was used to mark the corner points, Gaussian filtering and morphological image processing techniques were used to remove noise. A success rate of 94% was achieved in the studies performed. In time measurement tests, a good performance time of 1.56 seconds was achieved on average. The time measurements were compared with other studies in the literature. It has been observed that the proposed process has an excellent performance in terms of time.

Keywords: Subtitle detection, Image processing, Edge detection, Corner detection, Morphological image processing.

*Sorumlu yazar (Corresponding author): Erdal Güvenoğlu, erdalguvenoglu@maltepe.edu.tr.

1. Giriş

Bilgisayarlı görme teknikleri gündelik yaşantımızda ve sorunlarımızın kolaylaştırılmasında önemli bir rol oynamaktadır. Görüntü işleme, nesne tanıma ve algılama, yapay zekâ ve çeşitli desenler kullanılarak özel alanlara girişin engellenmesi, sinyal işleme, nesnelerin interneti gibi pek çok teknik bilgisayarlı görmenin önemli oyuncuları haline gelmiştir[1]. Dolayısı ile belgelerin dijitalleştirilmesinde ve ardından çeşitli uygulama alanlarında kullanılmasında bilgisayarlı görme en önemli adımlardan biridir. Belgelerin dijitalleştirilmesinde yaygın olarak optik karakter tanıma sistemleri (OCR) kullanılmaktadır. OCR görüntünün bir kısmını veya tamamını dikkate alarak metin verilerinin elde edilmesini sağlamaktadır. Ancak elde edilen sonuçlar bazen çok tatmin edici olmamakta ve desteklenen karakter setleri ile sınırlı olabilmektedir. Bununla birlikte, günümüzde yaygın olarak kullanılan video teknolojileri ile birlikte OCR sistemleri de altyazıların metinlere dönüştürülmesinde yaygın olarak kullanılmaktadır [2]. Sosyal medya platformları ve dijital eğlence sektörünün de yaygınlaşması ile birlikte video dosyalarındaki alt yazıların elde edilmesi ve işlenmesi önemli bir konu haline gelmiştir. Dijital eğlence sektöründe video materyalleri genellikle iki farklı türde altyazı seçeneği ile sunulmaktadır. Bunlar sabit kodlu ve harici altyazılar olarak bilinmektedir [3]. Sabit kodlu altyazılar videonun bir parçasıdır ve video üzerinde resim gibi görünmektedirler. Harici altyazılar ise video dosyasından ayrı bir halde harici bir dosyada bulunmakta ve gerektiğinde ortadan kaldırılabilirler. Sabit kodlu altyazıya sahip video dosyalarından bu yazıların elde edilmesi oldukça güçtür. Piyasada sabit kodlu altyazıların elde edilmesi amacıyla birbirinden farklı yazılımlar bulunmaktadır. Ancak bu yazılımların pek çoğu tatmin edici kaliteli sonuçların elde edilmesinde yetersiz kalabilmektedir. Bu yazılımlarda sabit altyazıların bulunduğu bölgenin belirlenmesi elde edilecek sonuçların daha kaliteli olabileceğini sağlayacaktır. Görüntü çerçeveleri birbirinden farklı özelliklere sahip olduğundan OCR yazılımları tüm çerçevede yer alan metinleri elde etmektedir. Bu sonuç kaliteli altyazıların elde edilmesini etkileyen bir sebeptir.

Son yıllarda mobil cihazların performanslarındaki artış nedeniyle görüntü ve video dosyalarında metnin elde edilmesi problemi büyük ilgi görmüştür. Bu cihazlarda gerçek zamanlı çeviri uygulamalarının da kullanılması bu probleme olan ilgiyi daha da arttırmıştır [4]. Bunun bir diğer nedeni ise yüksek performanslı ve hesaplama kapasitesi yüksek mobil cihazların giderek yaygınlaşmaya başlamasından ileri gelmektedir [5]. Bilgisayarlı görme teknolojilerindeki ilerlemeler bu problem üzerindeki çalışmalarında artmasına önayak olmuştur. Karakter tanımadaki en önemli araç OCR teknolojisidir. Her ne kadar düz belgelerde başarı oranı yüksek olsa da karmaşık bir arka plana sahip görüntülerde başarıyı düşüktür. Bu nedenle aydınlatma, metin düzeni, düşük çözünürlük ve yazı tiplerindeki farklılıklardan dolayı karmaşık görüntüler üzerinde bilgisayarlı görme tekniklerinin uygulanması gerekmektedir.

Literatürde sabit ve hareketli görüntü dosyalarından altyazıların elde edilmesi ile ilgili pek çok çalışma yapılmıştır. Bu çalışmaların temel amacı görüntü içerisinde metnin varlığının belirlenmesi [4, 6, 7], yerleştirilmesi [8-12] ve tanınması [6, 13] işlemlerini kapsamaktadır. Görüntü çerçevelerinde yer alan metinlerin elde edilmesi amacıyla bir çalışma [14] tarafından yapılmıştır. Çalışmada görüntü çerçevesinin tamamı üzerinde durulmuştur. Yöntemde Sobel kenar belirleme operatörü kullanılmış elde edilen sonuç görüntüsü üzerinde de dört farklı komşuluktaki renk değişimleri ile metinler tespit

edilmeye alıřılmıřtır. Benzer bir alıřma [15] tarafından yapılmıřtır. Benzer bir biimde kenar belirleme operatrleri kullanılmıř ve grnt bloklara ayrılmıřtır. Her bir blokta grnt histogramları oluřturularak karakter olabilecek koordinatlar iřaretlenmiřtir. Yntem arka planı tek dze renklerden oluřan grntlerde iyi alıřmakta ancak karmařık doku yođunluđunun yksek olduđu grntlerde hatalı sonular vermektedir. Metnin algılanması ve tanınması amacıyla bađlı bileřen analizi (CCA, Connected Component Analysis) [16, 17] ve kayan pencereler [16, 18, 19] metotları da tercih edilen yntemler arasındadır. CCA grnt ierisinde metin olabilecek alanların belirlenmesi amacıyla bir grntnn renk, kenar, piksel gradyanı ve doku gibi zelliklerini kullanmaktadır. Bu zellikler dıřında kalan alanlar ise eřitli filtreler yardımıyla grntden elenmektedir. İřlenebilecek bileřen sayısı az olduđundan etkili bir yntem olarak bilinmektedir. Ancak metin tipine karřı duyarlı olduđundan bazı metin ve karakterlerin kaybedilmesi sz konusu olabilmektedir [20]. Kayan pencereler ynteminde ise farklı leklerde pencereler yardımı ile ilgili pikselin tm komřuluklarına bu pencerelerin kaydırılması ile metin tanımlama iřlemi gerekleřtirilmektedir. Bu sre ok kapsamlı bir arama gerektirmekte ve hesaplama maliyeti olduka yksektir.

Grnt ierisinde bir metnin konumunun belirlenmesi iin yapılan bařka bir alıřmada [21], kkk lekli aday kutuları kullanılmıřtır. Aday kutuları yardımıyla metnin uzunluđu elde edilmesi amacıyla Bađlantıcı Metin Teklif Ađı (CTPN, Connectionist Text Proposal Network) tekniđi kullanılmıřtır. Grntlerde metnin uzunluđunun sabit olmaması nedeniyle yatay ynde CTPN ile metnin uzunluđu tahmin edilmektedir. Tespit edilen uzunluktaki metni elde etmek iin tekrarlayan sinir ađı (RNN, Recurrent Neural Network) [22] ađı kullanılmıřtır. Yatay olmayan metinlerin algılanmasında iyi bir bařarıma sahip deđildir. Yapılan benzer bir alıřmada [23] CTPN ile birlikte bađlantı segmentler kullanılmıřtır. Elde edilen tm metin paraları birleřtirilmiřtir. İlk olarak eđik olan metinler dzeltilmeye alıřılmıř ve birleřtirme iřlemi gerekleřtirilmiřtir. Ancak her ne kadar eđik metinler dzeltilmeye alıřılsa da kavis oranı yksek belgelerde metnin algılanmasında bařarım oranı dřktr. Tm satırların blmlere ayrılarak metni tespit edilmesine ynelik bir alıřma [24] tarafından yapılmıřtır. Metnin algılanma sreci birka ařamaya blndđnden daha fazla zaman gerektirmektedir. alıřmada kullanılan verimli ve dođru sahne metin dedektr (EAST, Efficient and Accurate Scene Text Detector) metnin bulunduđu blgeyi dođru tahmin edebilmekte ancak uzun metinlerin algılanmasında bařarımı dřktr.

Son yıllarda bir grntde metnin algılanması ve tanınması amacıyla derin đrenme teknikleri de kullanılmıřtır [25-28]. Grntlerde yer alan metinleri ve metin olmayan nesnelere sınıflandırılmak amacıyla ok katmanlı evriřimsel sinir ađları (CNN, Convolutional Neural Networks) kullanılarak ilgili ađın derin zellikleri hesaplanmaktadır. Long vd. tarafından derin ađlar kullanılarak grntden metnin elde edilmesini sađlayan bir yntem nerilmiřtir [28]. Yntemde nceden etiketlenmiř grntler eđitime verilmiř ve ardından tanıma iřlemi gerekleřtirilmiřtir. Bu alıřma, tanımanın yanı sıra diđer alıřmalarda kullanılan metin algılama metotlarının bir karřılařtırmasına da yer vermektedir. Etiketleme ve eđitim sresi ciddi zaman almaktadır. Etiketleme srecinde tanımda yardımcı olacak nesnelere dikkatli bir biimde sınırlarının belirlenmesi gerekmektedir. Derin ađların verimli bir Őekilde alıřabilmesi iin de ok fazla miktarda grntye ihtiya bulunmaktadır. Grnt ierisinde nesnelere sınırlarını tespit etmek amacıyla kullanılan yazılımlar ise nesne sınırlarının dođru tespit edilmesinde grnt arka planında ok karmařık bir zemin olması durumunda ok etkili deđillerdir. Tm bu sebeplerden dolayı etiketleme sresi el yordamıyla yapıldıđından uzun bir zaman gerektirmektedir.

Görüntü işleme teknikleri kullanılarak multimedya görüntülerinde alt yazı bölgelerinin tespit edilmesi

Benzer bir biçimde görüntüler çok fazla sayıda piksel öge barındırdığından ve görüntü sayısı çok fazla miktarda olduğundan eğitim süresi de uzun süreler gerektirebilmektedir. Bir diğer zorluk ise tanıma öncesinde düzensiz metin veri kümelerini normal metinlere dönüştürmek gerekmektedir. Derin öğrenmeye dayalı metin tanıma modelleri regresyon tabanlı ve segmentasyon tabanlı yöntemler olarak iki sınıfa ayrılmaktadır [29]. Derin öğrenmeye dayalı metin tanıma algoritmaları temel olarak görüntü düzeltme, özellik çıkarma, dizi tahmini vb. içermektedir. Bu sınıflarda yer alan bazı çalışmalar Tablo 1’ de özetlenerek verilmiştir.

Tablo 1. Derin öğrenme teknikleri ile metin algılama teknikleri [29]

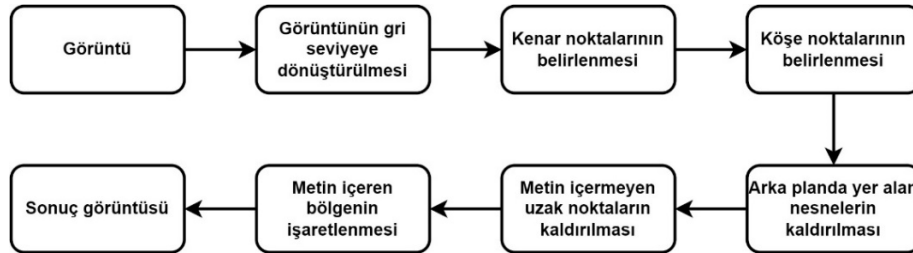
Tip	Model	Avantaj ve Dezavantajları
Regresyon	CTPN [21]	Hem evrişimli hem de tekrarlayan sinir ağlarını kullanan uçtan uca eğitim yöntemini benimsemektedir. Tahmin hızı oldukça düşüktür.
Regresyon	Seglink [23]	Metnin daha iyi algılanması için döndürme işlemi uygulanmaktadır. Geniş aralıklı ve kavisli metin satırlarını algılayamamaktadır.
Regresyon	EAST [24]	Doğruluğun sağlanması sırasında yöntem hızlı ancak uzun metin tespitinde iyi değildir.
Regresyon	ATRR [30]	Yöntem isteğe bağlı metin bölgelerini algılayabilmek için sınırlı sayıda nokta kullanmaktadır. Nokta sayısının artırılması durumunda işlem süresi artmaktadır.
Regresyon	LOMO [31]	Daha uzun metinlerin algılanması için kullanılmaktadır. Yinelemeli bir optimizasyon metodunu kullandığından süre açısından performansı düşük kalabilmektedir.
Regresyon	SBD [32]	Yöntemde yatay ve dönük metinleri algılayabilmekte ancak rastgele biçimlerden oluşan metinleri algılayamamaktadır.
Segmentasyon	CRAFT [33]	Yöntemde metnin içeriğinin tanınması için küçük bir alıcı alana ihtiyaç duymaktadır. Ancak bitişik karakterlerin algılanmasında iyi değildir.
Segmentasyon	PAN [34]	Yöntemde metnin algılanması için benzerlik vektörleri kullanılmaktadır. Kavisli yüzeylerde performansı düşüktür.
Segmentasyon	PSENet [35]	Rastgele biçimlere sahip metinlerin algılanmasında kullanılmaktadır. Yöntemin iyi çalışabilmesi için koordinat parametrelerini iyi seçilmesi gerekmektedir.

Literatürde yer alan alıřmalar dikkate alındığında derin öğrenme tekniklerini kullanan yöntemlerde önemli sayıda verinin yer aldığı veri setlerine ihtiyaç duyulmaktadır. Bununla birlikte veri setleri içerisinde yer alan metin ifadelerinin eğitim aşaması öncesinde etiketlenmesi gerekmektedir. Bu süreç çok ciddi zaman gerektirmektedir. Bu alıřmanın literatüre ana katkıları řunlardır:

- Derin öğrenme ve diđer tekniklerin kullanımı gerektiğinde etiketleme süresini kısaltmaktadır.
- Karmařık arka plana sahip görüntüler içerisinde yer alan metinlerin bulunduğu bölgelerin de tespit edilmesini sađlanmaktadır.
- Karmařık arka plana sahip görüntülerde OCR yazılımları etkili sonuçlar vermemektedir. Bundan dolayı önerilen yöntem ile metin verilerin bulunduğu bölgelerin tespit edilmesi OCR sürecinde de dođru sonuçların elde edilmesine yardımcı olmaktadır.
- Önerilen yöntem ile karmařık arka planların yer aldığı görüntülerde metin bulunan bölgelerin tespit edilmesinin ardından, metin karakter sınırlarının tespitini de kolaylařtırmaktadır.
- Görüntü işleme teknikleri kullanıldığından altyazı tespit süresi ortalama 1.56 sn' dir. Daha yüksek özelliklere sahip cihazlarda bu süre daha da kısaltmakta ve sektör kullanımına uygundur.
- alıřma ile birbirinden farklı özelliklere sahip görüntülerden oluşan ve gelecek alıřmalarda kullanılabilcek bir veri seti oluşturulmuřtur.

2. Önerilen yöntem

Önerilen yöntem birbirini takip eden adımlardan meydana gelmektedir. Bu adımlar řekil 1' de gösterilmiş ve yöntemin ayrıntılı detayları alt bölümlerde açıklanmıştır.



řekil 1. Önerilen yöntemin blok diyagramı

2.1. Kenar noktalarının belirlenmesi

Bu adımda öncelikle renkli görüntü gri seviye görüntüye dönüřtürölmektedir. Ardından kenar noktalarının belirlenmesi amacıyla biri yatay yönde biri de dikey yönde olmak üzere iki ayrı gradyan matrisi tanımlanmıştır. Yatay ve dikey yönde tanımlanan gradyan matrisleri řekil 2' de verilmiştir.

Görüntü işleme teknikleri kullanılarak multimedya görüntülerinde alt yazı bölgelerinin tespit edilmesi

$$G_x = \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \end{bmatrix} \quad G_y = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$$

Şekil 2. Gradyan matrisleri

Çalışmada bilinen kenar belirleme operatörleri ile de çeşitli testler yapılmış ancak etkili sonuçlar elde edilememiştir. Bu nedenle Şekil 2’ de verilen gradyan matrisleri tanımlanmış ve daha etkili olduğu görülmüştür. Belirtilen gradyan matrisleri çok sayıda deneysel çalışma yapılarak elde edilmiştir. Her bir gradyan matrisi görüntü üzerinde yatay ve dikey yönde ayrı ayrı gezdirilmiştir. Elde edilen gradyan görüntüde piksel değerlerinin daha baskın olarak elde edilmesi için sonuç görüntüsüne ait pikseller üzerinde orijinal görüntü I , kenarları tespit edilmiş görüntü I_{xy} , gradyan matrisleri gezdirilmiş görüntüler I_x ve I_y olmak üzere Denklem 2.1, 2.2 ve 2.3 ile verilen işlemler uygulanmıştır.

$$I'_x = I_x^2 \quad (2.1)$$

$$I'_y = I_y^2 \quad (2.2)$$

$$I_{xy} = I'_x \times I'_y \quad (2.3)$$

Kenarları belirlenmiş görüntüde çeşitli gürültüler meydana gelebilmektedir. Görüntüde yer alan gürültüler elde edilecek çalışma sonucunu etkileyen önemli bir faktördür. Görüntüde gürültülerin yer alması alt yazı bölgelerinin yanlış tespit edilmesine neden olabilmektedir. Bu nedenle kenarları belirlenmiş görüntüye gauss filtresi uygulanarak gürültü seviyesi metin konumunun daha başarılı tespit edilmesi amacıyla en aza indirgenmiştir.

2.2. Köşe noktalarının belirlenmesi

Köşe olabilecek noktaların belirlenmesi amacıyla Harris köşe saptama algoritması kullanılmıştır. İlk defa 1988 yılında C. Harris ve M. J. Stephens tarafından ortaya atılmıştır. Herhangi bir yönde hareket ettirilen pencere bölgelerindeki gri renk değişimlerini incelemektedir. Tüm yönlerin kontrolü sırasında birden fazla yöndeki pencerelerde gri renk değerlerinde önemli ölçüde değişiklik meydana geliyorsa bu nokta köşe noktası olarak değerlendirilmektedir [36]. Çalışmada kullanılan örnek bir görüntüde Harris algoritmasının uygulanması ile elde edilen sonuç görüntüsü Şekil 3’ de verilmiştir.



Şekil 3. Örnek köşe noktaları

Harris köşe saptama algoritma ile elde edilen görüntülerde metin bölgesi dışında yer alan bölgelerde de köşe noktaları olduğu görülmektedir. Bu noktalar metin bölgelerinin doğru tespit edilmesini engellemektedir. Bu nedenle metin ile ilgisi olmayan noktaların ortadan kaldırılması gerekmektedir. Görüntü arka planında yer alan nesnelerin ortadan kaldırılması istenmeyen bu noktaların da ortadan kaldırılmasına yardımcı olmaktadır.

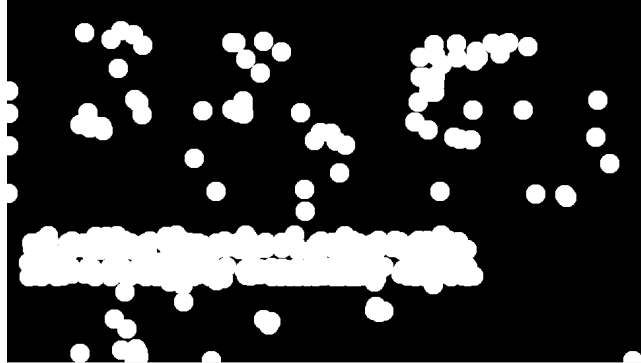
2.3. Arka planda yer alan nesnelerin kaldırılması

Köşe noktalarının koordinatlarının tespit edilmesinden sonra görüntü arka planında yer alan karmaşık nesnelerin ortadan kaldırılması gerekmektedir. Arka planda yer alan karmaşık nesnelere yazı olabilecek alanların belirlenmesinde sıkıntılar oluşturabilmektedir. Köşe olarak tespit edilen noktaların y eksenini koordinatı P_y ve x eksenini ile olan koordinat P_x olmak üzere arka plan nesnelerinin ortadan kaldırılması işlemi Denklem 2.4 ile gerçekleştirilmektedir.

$$I_B(x, y) = \sqrt{(x - P_x)^2 + (y - P_y)^2} , \quad eğer \begin{cases} 1 & I_B(x, y) < 15 \\ 0 & I_B(x, y) > 15 \end{cases} \quad (2.4)$$

Burada n görüntünün genişliğini ve m yüksekliğini ifade etmek üzere, $n \times m$ büyüklüğünde bir görüntüde tüm koordinatlar üzerinde gerekli hesaplamalar yapılarak köşe koordinatları dışındaki pikseller ortadan kaldırılmaktadır. Denklem 2.4' te verilen 15 değeri bir eşik değeri görmekte ve köşe koordinatlarına ait piksel değerlerinin yayılmasını sağlamaktadır. Bu eşik değeri görüntü içerisinde yer alan metinlerin büyüklüklerine göre ayarlanabilmektedir. Elde edilen örnek bir görüntü Şekil 4' de verilmektedir.

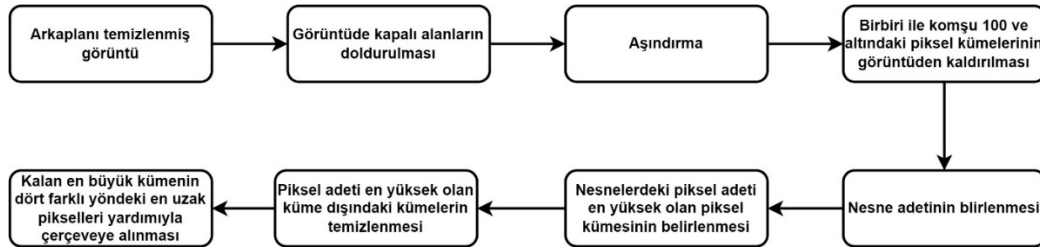
Görüntü işleme teknikleri kullanılarak multimedya görüntülerinde alt yazı bölgelerinin tespit edilmesi



Şekil 4. Köşe noktaları dışındaki piksellerin temizlenmesi

2.4. Metin içermeyen uzak noktaların kaldırılması

Arka plan piksellerinin ortadan kaldırılması neticesinde metin olamayacak noktaların görüntüden temizlenmesi gerekmektedir. Yapı itibari ile metin ifadelerinde daha fazla köşe noktası bulunmaktadır. Dolayısı ile köşe noktaları, içinde metin barındıran bir bölgede daha sık ve birbirlerine daha yakın olmaktadır. Başka bir ifadeyle, yapı itibariyle metin karakterleri birbirine yakındır ve köşe noktaları da birbirine yakın olmaktadır. İçinde altyazı barındıran bir görüntü de altyazı bölgesine uzak olan bölgelerin ortadan kaldırılması gerekmektedir. Uzak bölgedeki piksellerin ortadan kaldırılması için görüntü morfoloji teknikleri kullanılmaktadır. Bu süreçte gerçekleştirilen işlem adımlarına ait blok diyagram Şekil 5’ de verilmektedir.



Şekil 5. Metin içeren bölgenin morfolojik görüntü işleme ile yerinin tespit edilmesi aşamaları

Görüntü morfolojisi aşamalarının uygulanması neticesinde elde edilen örnek bir sonuç görüntüsü Şekil 6’ da verilmektedir.



Şekil 6. Metin olan bölgenin elde edilmesi örneđi

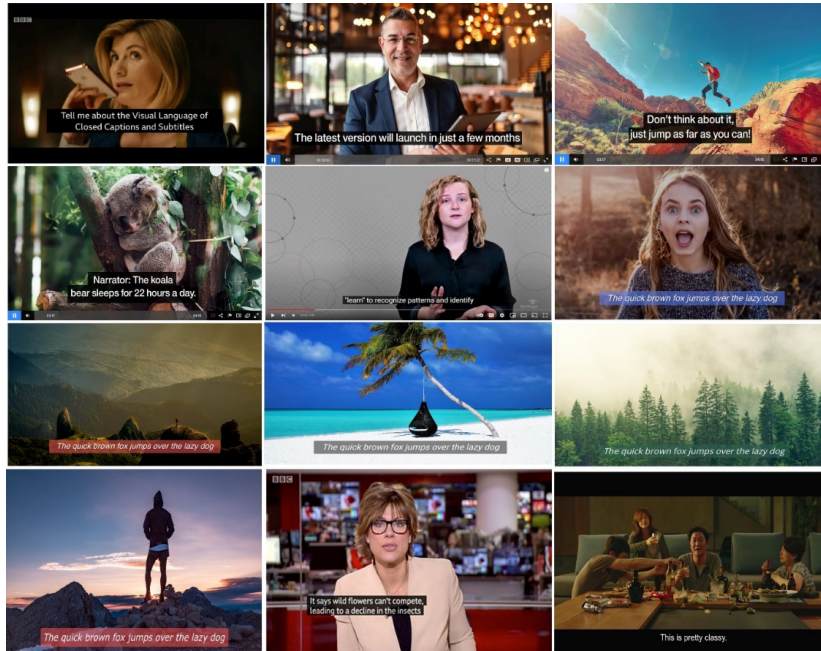
Son aşamada konumu belirlenen metin bölgesine ait dört farklı yöndeki dış kenar koordinatları bulunarak orijinal görüntü üzerinde işaretlenmektedir. İstenildiđi takdirde bu koordinatlar yardımıyla bölge pikselleri elde edilerek OCR yazılımı yardımıyla metne dönüştürülebilmektedir. İşaretlenmiş metin bölgesinin tespit edilmiş bir örneđi Şekil 7’ de verilmektedir.



Şekil 7. Uygulamadan elde edilen örnek sonuç görüntüsü

3. Deneysel sonuçlar

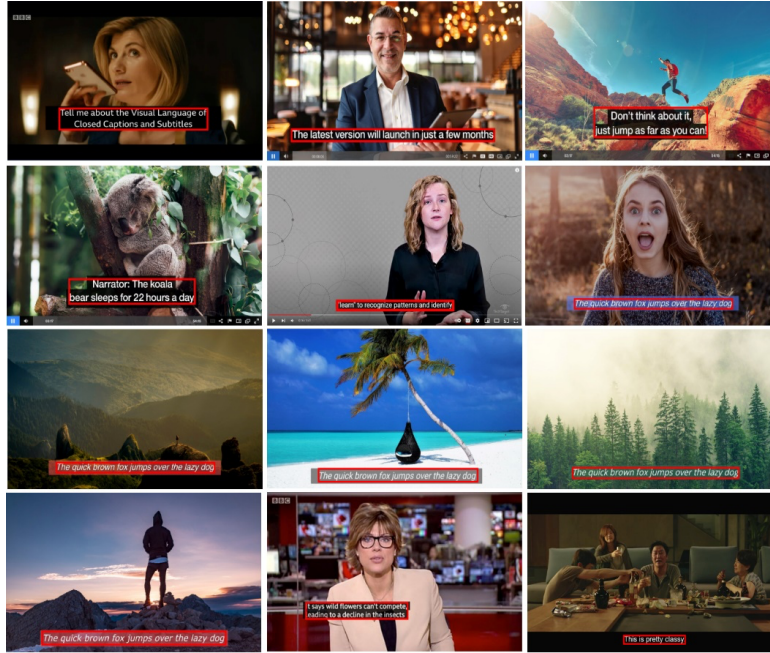
Önerilen yöntem, birbirinden farklı ölçülerde yüz adet görüntü üzerinde denenmiştir. Intel Core i7-4700 CPU 2.4 GHz 4 çekirdek, 16 GB RAM ve MS Windows 10 Pro 64 bit işletim sistemi üzerinde test edilmiştir. Algoritmanın geliştirilmesinde MATLAB 2020b sürümü kullanılmıştır. Çalışmada kullanılan bazı örnek görüntüler Şekil 8’ de verilmiştir.



Görüntü işleme teknikleri kullanılarak multimedya görüntülerinde alt yazı bölgelerinin tespit edilmesi

Şekil 8. Çalışmada kullanılan örnek bazı test görüntüleri

Önerilen yöntemin test görüntüleri ile gerçekleştirilmesi neticesinde elde edilen ve yazı olan bölgeleri tespit edilmiş bazı sonuç görüntüleri Şekil 9’ da verilmiştir.



Şekil 9. Metin bölgeleri tespit edilmiş örnek görüntüler

Çalışmanın test edilmesi neticesinde altyazı bölgelerinin doğru tespit edildiği ve başarılı sonuçlar verdiği görülmüştür. Çalışmanın gerçekleştirilmesi amacıyla farklı ölçülere ve özelliklere sahip 100 adet görüntüden oluşan bir veri seti oluşturulmuştur. Görüntüler internet taramaları ve sabit altyazılara sahip multimedya görüntülerinden elde edilmiştir. Görüntü veri setinin en küçüğü 900×500 ve en büyüğü 1920×1080 çözünürlükte görüntülerden oluşmaktadır. Çalışmanın gerçekçi sonuçlar vermesi bakımından görüntü veri setinin tamamen rastgele olmasına dikkat edilmiştir.

Tablo 2. Veri seti üzerinde gerçekleştirilen gözlem sonuçları

Veri seti görüntü sayısı	100
Doğru tespit edilen görüntü sayısı	94
Yanlış tespit edilen görüntü sayısı	6

Elde edilen sonuç görüntülerinde değerlendirme işlemi bölgesel bazda olduğundan doğru sonuç elde edilip edilmediği yapılan gözlemler neticesinde belirlenmiştir. Yapılan gözlemlerde yüz adet rastgele görüntü üzerinde yapılan testlerde %94 oranında bir başarı elde edilmiştir. Altı adet görüntüde

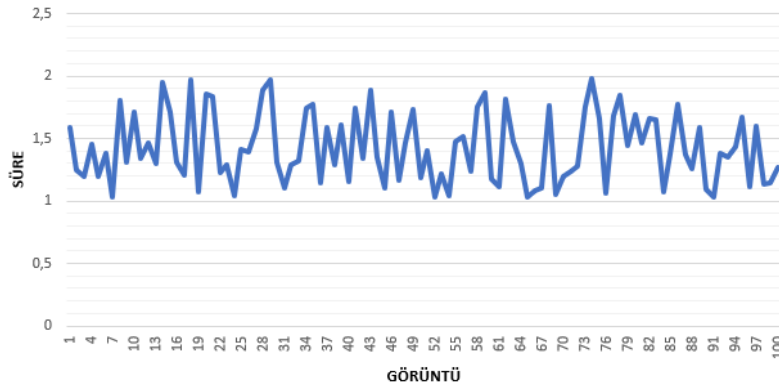
metin bölgesi tespit edilmiş ancak metnin tamamının tespit edilemediđi anlaşılmıştır. Bu tespitin yapılamamasının en temel nedeni metin ifadeleri ile metnin arka planında bulunan zemin renk değerlerinin metin karakterleri ile aynı veya birbirine çok yakın olmasından kaynaklandığı görülmüştür. Dolayısı ile bu bölgelerde yeteri miktarda köşe noktası tespit elde edilememiş ve bu durumun da metnin yerinin doğru olarak tespit edilememesine neden olmuştur. Bu ve benzeri durumların düzeltilmesi ve metin bölgelerinin tespit edilebilmesi morfolojik görüntü işleme tekniklerine el yordamı ile müdahale ile ortadan kaldırılabilir. Ancak böyle bir durumun çalışmanın amacı dışında olduğu düşünöldüğünde herhangi bir müdahalede bulunulmamıştır. Genel olarak değerlendirildiğinde yöntemin başarılı olduğu söylenebilmektedir.

Tablo 3. Çalışma sürelerinin karşılaştırılması

Yöntem	Süre (s)	Metot	Veri seti
Önerilen Yöntem	1,56	Görüntü İşleme	Rastgele görüntüler
[1]	0,366	Derin Öğrenme	FiViD
[37]	0,688	Görüntü İşleme	ICDAR'13
[38]	0,810	Görüntü İşleme	ICDAR'13
[39]	0,830	Görüntü İşleme	ICDAR'13
[40]	0,938	Görüntü İşleme	ICDAR'13
[41]	0,892	Görüntü İşleme	ICDAR'13

Yöntem çalışma süresi bakımından da test edilmiştir. Tüm sürecin ortalama olarak 1.56 sn' de tamamlandığı görülmüştür. Tablo 3' e bakıldığında çalışma süresinin diğer yöntemlerde yüksek olduğu, bazı yöntemlere ise çok yakın olduğu görülmektedir. Bunun en temel sebebi önerilen yöntemde görüntü veri setinin en küçük 900×500 çözünürlükte görüntülerden oluşmasıdır. Literatürde var olan yöntemlerde ise kullanılan ICDAR'13 veri setinde yer alan görüntüler 100×100 boyutundadır [42]. Dolayısı ile önerilen yöntemde kullanılan görüntülerin boyutlarının büyük olması sonucu doğrudan etkilemektedir. En küçük görüntü çözünürlükleri dikkate alınarak süre karşılaştırması yapıldığında yaklaşık 45 kat büyük olan görüntülerde oldukça başarılı bir sonuç elde edildiđi söylenebilmektedir. Çalışmada kullanılan her bir görüntünün süre ölçümü yapılmış ve elde edilen süre grafiđi Şekil 10' da verilmiştir.

Görüntü işleme teknikleri kullanılarak multimedya görüntülerinde alt yazı bölgelerinin tespit edilmesi



Şekil 10. Test görüntülerine ait süre ölçüm grafiği

Önerilen yöntemin test sonuçları dikkate alındığında süre bakımından oldukça hızlı çalıştığı görülmektedir. Testlerde kullanılan görüntüler standart ölçülerde olmamakla birlikte rastgele seçilmiş ve farklı ölçülerden oluşan görüntülerden oluşmaktadır. Dolayısı ile boyut olarak büyük görüntülerde işlem süresi artmaktadır. İşlem süresinin farklı çözünürlükte görüntüler için yüksek olmasına rağmen önerilen yöntemin oldukça başarılı olduğu söylenebilmektedir.

4. Sonuç

Bu çalışmada, multimedya görüntüleri içerisinde yer alan sabit alt yazıların bulunduğu bölgelerin tespit edilmesi için bir yöntem sunulmuştur. Günlük hayatımızda multimedya uygulamaları yaygın olarak kullanılmaktadır. Bununla birlikte multimedya görüntülerinde yer alan altyazıların elde edilmesi, metinlerin seslendirilmesi v.b. farklı kullanım alanlarına sahiptir. Bu alanda çeşitli çalışmalar yapılmaktadır. Bu çalışmalarda en temel sorun altyazının yerinin görüntüde doğru tespit edilememesidir. Derin öğrenme gibi tekniklerin kullanılmasında ise görüntülerin her birinin eğitim aşaması öncesinde etiketlenmesi gerekmektedir. Bu durum uzun zaman alabilmektedir. Bu çalışmada önerilen yöntem ile görüntü içerisinde yer alan alt yazı bölgeleri Harris köşe saptama algoritması, Gauss filtresi ve morfolojik görüntü işleme teknikleri kullanılarak belirlenmiştir. Farklı çözünürlükte rastgele görüntüler üzerinde testler yapılmış, literatürdeki yöntemler ile kıyaslanmış ve başarılı sonuçlar elde edilmiştir. Literatürdeki yöntemlerin çalışma süresi bakımından kıyaslanmasında önerilen yöntemde kullanılan görüntü boyutlarına oranla yaklaşık 45 kat küçük görüntüler kullanıldığı görülmüştür. Dolayısı ile süre bakımından yapılan kıyaslamalarda etkili sonuçlara ulaşılmıştır. Test sonuçları dikkate alındığında önerilen bu yöntemin altyazı tespiti için derin öğrenme tekniklerinde etiketleme süresini kısaltabileceği düşünülmektedir. Benzer bir biçimde alt yazıların yer aldığı farklı amaçlar için de kullanılabilenliği öngörülmektedir.

Araştırma ve yayın etiği beyanı: Bu çalışmada bilimsel araştırma ve yayın etiğine uygun davranılmış ve yazar tarafından kabul edilmiştir.

Etik kurul onayı: Bu çalışmada herhangi bir etik kurul onayına gerek yoktur.

Yazarların katkısı: Bu çalışma tek yazarlıdır ve tüm katkı yazar tarafından sağlanmıştır.

Çıkar çatışması beyanı: Herhangi bir kurum ya da kişiler ile çıkar çatışması bulunmamaktadır.

Kaynakça

- [1] Elshahaby, H., Rashwan, M. (2022). An end to end system for subtitle text extraction from movie videos. *Journal of Ambient Intelligence Humanized Computing*, 13, 1853-1865.
- [2] Wang, Y., Wu, M., Zheng, B., Zhu, S. (2022). HuZhouSpeech: A huzhou dialect speech recognition corpus. 5th International Conference on Information Communication and Signal Processing (ICICSP), 153-157, Shenzhen, China.
- [3] Wang, D. (2018). The experimental implementation of grabcut for hardcode subtitle extraction. 17th International Conference on Computer and Information Science (ICIS), 1-4, Singapore.
- [4] Ye, Q., Doermann, D. (2014). Text detection and recognition in imagery: A survey. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 37 (7), 1480-1500.
- [5] Liu, X. (2008). A camera phone based currency reader for the visually impaired. *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, 305-306, Canada.
- [6] Huang, M. et. al., (2022). Swintextspotter: Scene text spotting via better synergy between text detection and text recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4593-4603, USA.
- [7] Naiemi, F., Ghods, V., Khalesi, H. (2022). Scene text detection and recognition: a survey. *Multimedia Tools Applications*, 81 (14), 20255-20290.
- [8] Kim, G., Yokoo, S., Seo, S., Osanai, A., Okamoto, Y., Baek, Y. (2023). On text localization in end-to-end OCR-Free document understanding transformer without text localization supervision. *International Conference on Document Analysis and Recognition*, 215-232, USA.
- [9] Chaitra, Y., Dinesh, R. (2022). An impact of radon transforms and filtering techniques for text localization in natural scene text images. *ICT with Intelligent Applications: Proceedings of ICTIS 2021*, 563-573, India.
- [10] Goud, D. S., Vigneshwari, M., Aparna, P., Vijayasekaran, G., Yadav, A. S., Kumar, A. (2022). Text localization and recognition from natural scene images using AI. *International Conference on Automation, Computing and Renewable Systems (ICACRS)*, 1153-1158, India.
- [11] Jayanth, R. M., Kapanaiyah, M. (2022). Dominating set based arbitrary oriented bilingual scene text localization. *International Journal of Electrical Computer Engineering*, 12 (4), 3730-3738.
- [12] Champa, H. (2023). Text localization and recognition in video and images. *Journal of Data Acquisition Processing*, 38 (2), 3512.
- [13] Wang, P., Da, C., Yao, C. (2022). Multi-granularity prediction for scene text recognition. *European Conference on Computer Vision*, 339-355, Israel.
- [14] Favorskaya, M.N., Zotin, A.G., Damov, M.V. (2010). Intelligent inpainting system for texture reconstruction in videos with text removal. *International Congress on Ultra Modern Telecommunications and Control Systems*, 867-874, Moscow, Russia.

Görüntü işleme teknikleri kullanılarak multimedya görüntülerinde alt yazı bölgelerinin tespit edilmesi

- [15] Khodadadi, M., Behrad, A. (2012). Text localization, extraction and inpainting in color images. 20th Iranian Conference on Electrical Engineering (ICEE2012), 1035-1040, Tehran, Iran.
- [16] Neumann, L., Matas, J. (2015). Real-time lexicon-free scene text localization and recognition. IEEE Transactions on Pattern Analysis Machine Intelligence, 38 (9), 1872-1885.
- [17] Koo, H.I., Kim, D.H. (2013). Scene text detection via connected component clustering and nontext filtering. IEEE Transactions on Image Processing, 22 (6), 2296-2305.
- [18] Neumann, L., Matas, J. (2011). A method for text localization and recognition in real-world images. Computer Vision–ACCV 2010: 10th Asian Conference on Computer Vision, 770-783, Queenstown, New Zealand.
- [19] Neumann, L., Matas, J. (2012). Real-time scene text localization and recognition. IEEE Conference on Computer Vision and Pattern Recognition, 3538-3545, USA.
- [20] Zulkeflee, A. N., Yussof, W.N.J.H.W., Umar, R., Ahmad, N., Mohamad, F. S., Man, M., Awalludin, E. A. (2022). Detection of a new crescent moon using the Maximally Stable Extremal Regions (MSER) technique. Astronomy Computing, 41, 100651.
- [21] Tian, Z., Huang, W., He, T., He, P., Qiao, Y. (2016). Detecting text in natural image with connectionist text proposal network. Computer Vision–ECCV 2016: 14th European Conference, 56-72, Netherland.
- [22] He, P., Huang, W., Qiao, Y., Loy, C., Tang, X. (2016). Reading scene text in deep convolutional sequences. Proceedings of the AAAI conference on artificial intelligence, USA.
- [23] Shi, B., Bai, X., Belongie, S. (2017). Detecting oriented text in natural images by linking segments. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2550-2558, USA.
- [24] Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., Liang, J. (2017). East: an efficient and accurate scene text detector. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5551-5560, USA.
- [25] Kazmi, W., Nabney, I., Vogiatzis, G., Rose, P., Codd, A., (2020). An efficient industrial system for vehicle tyre (tire) detection and text recognition using deep learning. IEEE Transactions on Intelligent Transportation Systems, 22 (2), 1264-1275.
- [26] Hassan, H., El-Mahdy, A., Hussein, M. E. (2021). Arabic scene text recognition in the deep learning era: Analysis on a novel dataset. IEEE Access, 9, 107046-107058.
- [27] Wang, M., Niu, S., Gao, Z. (2019). A novel scene text recognition method based on deep learning. Computers, Materials Continua, 60 (2), 781-794.
- [28] Long, S., He, X., Yao, C. (2021). Scene text detection and recognition: The deep learning era. International Journal of Computer Vision, 129, 161-184.
- [29] Wang, X.-F., He, Z.-H., Wang, K., Wang, Y.-F., Zou, L., Wu, Z.-Z. (2023). A survey of text detection and recognition algorithms based on deep learning technology. Neurocomputing, 556, 126702.

- [30] Wang, X., Jiang, Y., Luo, Z., Liu, C.-L., Choi, H., Kim, S. (2019). Arbitrary shape scene text detection with adaptive text region representation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6449-6458, USA.
- [31] Zhang, C., Liang, B., Huang, Z., En, M., Han, J., Ding, E., Ding, X. (2019). Look more than once: An accurate detector for text of arbitrary shapes. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10552-10561, USA.
- [32] Liu, Y., He, T., Chen, H., Wang, X., Luo, C., Zhang, S., ... Jin, L. (2021). Exploring the capacity of an orderless box discretization network for multi-orientation scene text detection. International Journal of Computer Vision, 129, 1972-1992.
- [33] Baek, Y., Lee, B., Han, D., Yun, S., Lee, H. (2019). Character region awareness for text detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9365-9374, USA.
- [34] Wang, W., Xie, E., Song, X., Zang, Y., Wang, W., Lu, T., ... Shen, C. (2019). Efficient and accurate arbitrary-shaped text detection with pixel aggregation network. Proceedings of the IEEE/CVF International Conference on Computer Vision, 8440-8449, Korea.
- [35] Wang, W., Xie, E., Li, X., Hou, W., Lu, T., Yu, G., Shao, S. (2019). Shape robust text detection with progressive scale expansion network. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9336-9345, USA.
- [36] Guiming, S. Jidong, S. (2018). Multi-scale Harris corner detection algorithm based on canny edge-detection. IEEE International Conference on Computer and Communication Engineering Technology (CCET), 305-309, China.
- [37] Yao, C., Bai, X., Liu, W., Ma, Y., Tu, Z. (2012). Detecting texts of arbitrary orientations in natural images. 2012 IEEE Conference on Computer Vision and Pattern Recognition, 1083-1090, Providence, RI, USA.
- [38] Shivakumara, P., Phan, T. Q., Tan, C. L. (2010). A laplacian approach to multi-oriented text detection in video. IEEE Transactions on Pattern Analysis Machine Intelligence, 33(2), 412-419.
- [39] Koo, H. I., Kim, D. H. (2013). Scene text detection via connected component clustering and nontext filtering. IEEE Transactions on Image Processing, 22 (6), 2296-2305.
- [40] Yin, X.-C., Yin, X., Huang, K., Hao, H.-W. (2013). Robust text detection in natural scene images. IEEE Transactions on Pattern Analysis Machine Intelligence, 36 (5), 970-983.
- [41] Ye, Q., Doermann, D. (2014). Scene text detection via integrated discrimination of component appearance and consensus. Camera-Based Document Analysis and Recognition: 5th International Workshop, CBDAR 2013, 47-59, Washington, DC, USA.
- [42] Karatzas, D., Shafait, F., Uchida, S., Iwamura, M., i Bigorda, L. G., Mestre, S. R., ... & De Las Heras, L. P. (2013). ICDAR 2013 robust reading competition. 12th International Conference on Document Analysis and Recognition, 1484-1493, Washington, DC, USA.