

Görme Engelliler için Akıllı Telefon ile Nesne Arama

Hüsnü Mert Polat, Muhammed Ensar Özer, Kevser Sertel, Sıdıka Tuğçe Yılmaz,
Süleyman Eken, Ahmet Sayar

Bilgisayar Mühendisliği, Kocaeli Üniversitesi,
41380 İzmit, Türkiye
{husnumertpolatt, ensr.6060, kewser.3806,
sdk.tugce.yilmaz}@gmail.com
{suleyman.eken, ahmet.sayar}@kocaeli.edu.tr

Özet. Görme engelliler hayatta birçok zorlukla karşılaşmaktadır. Kamuoyunda birçok kurum ve kuruluş bu engelleri aşmak için çalışmalar yapsa da maalesef bazı sorunlar hala çözüme kavuşturulmamıştır. Bunlardan birisi de görme engellilerin kaybettikleri veya o an yerini tam kestiremedikleri eşyaları bulamamasıdır. Geliştirilen uygulama mobil cihazlar üzerinde çalışarak görme engelli olan kişinin bulmak istediği bir nesneyi, cihazın kamerası aracılığı ile bulmasını sağlamaktadır. Uygulama açıldığında kamera devreye girmekte ve kullanıcı akıllı telefonu arama yapacağı mekan veya zemin üzerinde gezdirmektedir. İlk olarak kameradan alınan görüntüdeki nesneler, Tensorflow nesne saptama API'si ile bulunmaktadır. Daha sonra saptanan nesnelerin görüntü içindeki yeri alan ayrıştırma yoluyla konumlandırıldıktan sonra Google metin seslendirme ile kullanıcıya nesnenin ismi ve görüntü içindeki yeri bildirilmektedir.

Anahtar kelimeler: Nesne arama, akıllı telefon, Tensorflow, nesne konumlandırma, metin seslendirme

Object Search with Smart Phone for Visually Impaired People

Abstract. Visually impaired people face many difficulties in life. Although many institutions and organizations are working to overcome these difficulties, unfortunately some problems have not yet been solved. One of them is that they can not find items that the blind people have lost or that they do not know where they are at the moment. The developed application works on mobile devices to find an object that the visually impaired wants to find through the camera. When the application is turned on, the camera is switched on and the user runs the smartphone on the floor or space to search. Firstly, objects in the image taken from the camera are detected with Tensorflow object detection API. Later, after the detected objects are positioned by the parsing taking place in the image, their names and positions are notified with Google text-to-speech.

Keywords: Object search, smart phone, Tensorflow, object positioning, text to speech

1 Giriş

Günlük hayatta hemen hemen hepimiz eşyalarımızı kaybederiz. Bu kaybettiğimiz eşyalar beraberinde çok yüksek maliyetleri de getirebilir. Pixie'nin kayıp eşyalar ile ilgili ABD'de yaptığı ankete göre Amerikalılar kaybettikleri eşyaların yerine yenilerini almak için bir yılda yaklaşık 2.7 milyar dolar gibi inanılmaz bir meblağ harcamak durumunda kalıyor. Kayıp eşyalar beraberinde getirdiği maddi maliyetin yanı sıra belki de hayattaki en değerli şey olan zamandan da ciddi ölçüde alıyor.

Pixie'nin araştırması kaybedilen eşyaların zaman maliyeti ile ilgili bir takım korkutucu verileri açığa çıkarıyor. Araştırmaya göre insanlar kaybettikleri eşyaları bulabilmek için yılda ortalama 2.5 günlerini harcıyor. Aynı zamanda araştırma, eşyalarını kaybeden insanların zaman ile bağlantılı başka maliyetlere de katlandığını ortaya koyuyor. Araştırma sonuçları, eşyalarını kaybedenlerin %60'ının işe veya okula geç kaldığını, %49'unun önemli toplantı veya iş görüşmelerini kaçırdığını, %22'sinin ise binmesi gereken otobüs veya trene yetişemediğini gösteriyor. Kaybedilen bir eşya genellikle 20 saniye ve 5 dakika arasında bulunabiliyor; fakat bazı eşyaları bulabilmek daha çok zaman alabiliyor. Ortalamaya göre bulunması 15 dakikadan fazla süren eşyalar şöyle: ev, araba anahtarları (%21.3), cüzdanlar (%20.2), şemsiyeler (%19), pasaportlar (%18.9), sürücü ehliyetleri (%18.8) ve banka kartları (%18.7) [1]. Veriler, günlük yaşamda sıklıkla kullandığımız eşyaları sürekli kaybettiğimizi ve onları bulabilmek için ciddi zaman ve enerji harcadığımızı gösteriyor.

Sağlıklı insanlar bile kaybettikleri veya çoğu zaman gözleri önünde olan dikkatsizlik yüzünden göremediği eşyaları ararken bu kadar zaman kaybediyor ve o gün yapması gereken işleri yapamıyorlar.

Bu çalışmada yukarıdaki sonuçları verilen ankette de geçen, gündelik hayatta daha fazla kullanılan nesnelerin bulunması sağlanmıştır. Görme engelli kişinin telefon içerisinde uygulamayı kısa yolu kullanarak açması çok zor hatta imkansızdır. Bundan dolayı uygulama ses açma tuşuna beş saniye boyunca basılı tutunca açılmaktadır. Açıldığının anlaşılması için bir bildirim sesi verilmektedir. Uygulama açıldığında kamera devreye girmekte ve kullanıcı akıllı telefonu arama yapacağı mekan veya zemin üzerinde gezdirmektedir Bu sırada akıllı telefon kamerasında görüntülenen nesneler Tensorflow nesne saptama API'si ile bulunmakta ve nesnenin ismi ve konumu kullanıcıya Google metin seslendirme uygulaması yoluyla sesli olarak bildirilmektedir.

Çalışmanın geri kalan kısmı şu şekilde organize edilmiştir. 2. bölümde literatürdeki çalışmalar verilmiş, 3. bölümde geliştirilen uygulamanın detayları sunulmuştur. 4. bölümde uygulamanın arayüzleri verilmiştir. Son bölümde ise sonuçlar sunulmuş, ne gibi geliştirmeler yapılabilir tartışılmıştır.

2 Literatürdeki Çalışmalar

Bu bölümde literatürde nesne tanıma üzerine yapılmış çalışmalar verilmiştir. Görme tabanlı mobil robot ile farklı renklerde nesnelerin gerçek zamanlı takibiyle ilgili bir çalışma gerçekleştirilmiştir. Gerçekleştirilen robot farklı alanlarda kullanılmak üzere tasarlanmıştır. Bu robotun otonom bir şekilde hareket etmesi

sağlanarak farklı renklerdeki nesneleri takip edebilmesi gerçekleşmiş olup tepki verme süresi 96-106 ms aralığında ölçülmüştür [2].

Gerçek zamanlı endüstriyel nesne tanımadaki kameraların kullanılması konusunda Sedat ve Bayram, kameralı bir sistemin renk algılayıcı bulunduran bir sisteme göre gerçek zamanlı nesne tanıma işlemini daha hızlı gerçekleştirebildiği gözlemlemişlerdir. Renk algılayan sistemin yerine yerleştirilen bir kamera ile sistemin yaklaşık üç kat daha başarılı sonuç verdiği sonucuna ulaşılmıştır [3].

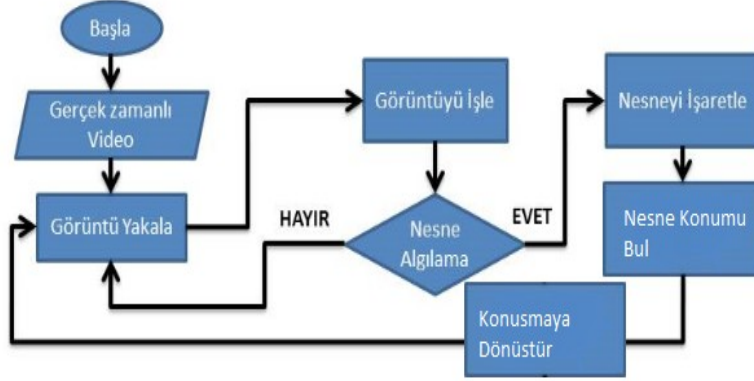
Gerçek zamanlı olarak, Anfis ile renk tabanlı nesne tespit ve motorlu sistem ile takip edilmesi projesinde kamera tarafından alınan görüntü MATLAB programı ile adaptif bir biçimde işlenerek hedef nesnenin merkezi tespit edilmektedir. Tespit edilen merkezin görüntünün merkezine olan uzaklığı bulunarak, seri porta bağlı bulunan mikrodenetleyicili devre aracılığı ile üzerinde kamera bulunan adım motora hareket vermesi sağlanmıştır. Böylece hedef nesnenin kamera tarafından alınan görüntünün merkezine çekilmesi sağlanarak, gerçek zamanlı nesne takip uygulaması gerçekleştirilmiştir [4].

Üç boyutlu sahneler ve nesne tanıma için gürbüz anahtar nokta eşleştirilmesi çalışmasında yakın zamanda düzlemsel nesneler için ikilik betimleyiciler ile anahtar nokta eşleme amacıyla önerilen bir yöntem üç boyutlu nesneler için uyarlanmıştır. Bu yöntemin başarısı yüzden fazla resim içeren bir müze nesne tanıma uygulamasında test edilmiştir. Ayrıca sadece eşleme başarısı yüksek betimleyicilerin kullanılmasının nesne tanıma uygulamasının başarısına etkisi de ölçülmüştür [5].

Günümüzde Konvolüsyonel Sinir Ağları (KSA), sınıflandırma ve algılama uygulamalarının ayrılmaz parçası haline gelmişlerdir. Konvolüsyonel Sinir Ağlarının pratikte kullanılmamasının temel nedenlerinden biri, onların güçlü hesaplamaları yapabilen işlemcilerle ihtiyaç duymalarıydı. Ekran kartlarının gelişmesi ile Krizhevsky ve arkadaşları 2012 yılında gerçekleştirilmiş ILSVRC-2012 yarışmasını KSA yardımı ile sınıflandırıcı bir model geliştirerek kazanmışlardır. İlerleyen zamanlarda R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN ve YOLO olmak üzere kendinden önceki modelin performansını iyileştiren, çeşitli algılama ve sınıflandırma modelleri geliştirilmiştir [6-10]. Bu modeller kullanılarak çeşitli nesne tanıma uygulamaları geliştirilmiştir. Ayrıca farklı alanlarda çeşitli mobil uygulamalar görme engelliler için geliştirilmiştir. Full otonom market uygulamalarında görme engelli kimseler için yeni bir ürün tanıma ve bilgilendirme yapılabilmektedir [11]. TapTapSee, CloudSight Görüntü Tanıma API'si tarafından desteklenen kör ve görme engelli kullanıcılar için özel olarak tasarlanmış bir mobil kamera uygulamasıdır. Uygulama, nesneleri fotoğraflamak ve kullanıcı için yüksek sesle tanımlamak için cihazın kamera ve VoiceOver işlevlerini kullanmaktadır [12].

3 Nesne Rehberim: Nesne Arama ve Konumlandırma

Şekil 1'de geliştirilen sisteme ait akış diyagramı genel hatlarıyla verilmiştir. Alt başlıklarda sistemin detaylandırılması yapılacaktır.



Şekil 1 Geliştirilen sistem akış diyagramı

3.1 Görüntü Yakalama ve Ön İşlemler

Akıllı telefonun kamerasından nesne çıkarımı yapılacak olan yerin görüntüsü alınır. Bu görüntü derin öğrenme modeline girdi olacağından boyutları modelin girdi boyutuna ([224, 224, 3]) göre düzenlenir, piksel değerleri [-1, 1] aralığına map edilir.

3.2 Nesne Saptama

KSA çeşitli katmanlardan oluşur. Her katmanın kendine ait bir sorumluluğu vardır. Genel olarak KSA mimarisinde aşağıdaki üç ana katman kullanılır. Bunlar (i) konvolüsyon, (ii) havuzlama (pooling) ve (iii) tam bağlı (fully connected) katmanları olarak adlandırılmışlardır. Herhangi bir KSA mimarisi oluşturulurken bu üç katman belli bir düzende toplanır. Bahsedilen katmanları kullanarak basit bir KSA mimarisi geliştirilebilir.

Konvolüsyon katmanındaki çıkışın nöron sayısı üç tane hiper parametre ile kontrol edilir. Bunlar derinlik (depth), kaydırma adım uzunluğu (stride) ve sıfır-tamponlama (zero-padding). Derinlik; kullanmak istenilen filtrelerin sayısına karşılık gelir ve her bir filtre giriş görüntüsünde farklı birşeyleri aramayı öğrenir. Filtrelerin kaydırılacağı adım uzunluğu belirtilmelidir. Kaydırma adım uzunluğu büyüdükçe çıkışta elde edilecek çıktının boyutu küçülür. Kaydırma adım uzunluğu ne kadar küçük ise, görüntüden daha fazla bilgi elde edilir. Bazı durumlarda girişin kenarlarını sıfırlar ile doldurmak uygun görülür. Böylece çıkışın boyutunu kontrol edebilecek bir tane daha hiper parametre sıfır-tamponlama yardımı gelir.

Pratikte ardışık konvolüsyonel katmanların arasına periyodik olarak bir pooling katmanı eklemek yaygın bir yöntemdir. Pooling katmanı; ağ mimarisindeki parametrelerin ve hesaplamaların sayısını azaltan ve aşırı öğrenmeye engel olan bir mekanizmadır. KSA'da en yaygın kullanılan pooling çeşidi ise max pooling'dir.

Tam bağlı katman, çıktı katmanında bir softmax aktivasyon fonksiyonu kullanan geleneksel çok katmanlı perceptrondur. Tam bağlı terimi, önceki katmandaki her nöronun sonraki katmandaki her nörona bağlı olduğunu ima eder.

Görüntülerden nesne saptaması yapmak için Tensorflow nesne saptama API'sinden yararlanılmıştır [13]. İlgili API içinde birçok KSA modeli barındırmaktadır. Bu modellerden “ssd_mobilenet_v1_android_export” kullanıldı.

3.3 Nesne Konum Bulma

Bir önceki adımda saptanan nesnelerin görüntü içinde konumlandırılması bu aşamada yapılmaktadır. Bu işlem çerçeve içine alınmış nesnelerin sınırları kullanılarak yapılmaktadır. Konum tespiti yapılması için ekran belirli aralıklara bölünerek atamalar yapılır. Başlangıç, sol bitiş, sağ başlangıç, sağ bitiş ve ön izleme alanıdır. Ekranı bölümlendirdikten sonra ekranda, bulunan nesnelerin koordinatlarına göre nesnenin ekranda kapladığı alan tespit edilir. Daha sonra tespit edilen bu alan bölümlendirilmiş blokların hangisinin içerisinde yer aldığı kontrol edilir ve bu kontrol sonucu nesne ile ilgili lokasyon tespit edilir.

3.4 Seslendirme

İsimleri ve konumları belirlenmiş nesnelerin kullanıcılara sesli olarak okunması Google metin seslendirme API'si yardımıyla yapılmaktadır. Google metin seslendirme, Android işletim sistemi için Google tarafından geliştirilen bir ekran okuyucu uygulamasıdır. Birçok dil desteğiyle ekrandaki metni seslendirmeyi sağlar [14].

4 Kullanıcı Arayüzleri

Görme engelli olan kullanıcı telefon ekranında uygulamayı bulup açma imkanı bulamayabileceği için ses açma tuşuna basılı tuttuğunda uygulama açılır ve açıldığına dair kişiye bildirim sesi gelir Kullanıcı daha sonra arama yapmak istediği bölgeye doğru telefon kamerasını tutarak bulmak istediği nesneyi arar. İsmi tespit edilen nesne veya nesneler için konum bilgilerinin belirlenmesi sağlanır. İsimleri ve konum bilgileri tespit edilen nesneler görme engelli kişilere sözlü olarak aktarılır. Bu işlemler ekranda algılanan nesne sayısı kadar tekrar eder ve ekranda algılanan tüm nesnelerin isimleri ve konumları söylenmiş olur. Örnek bir arayüz Şekil 2'de verilmiştir.



Şekil 2 Örnek bir uygulama arayüzü

5 Sonuç ve Gelecekteki Çalışmalar

Gerçekleştirdiğimiz uygulama ile görme engelli kişilerin yaşamlarını kolaylaştırmayı hedefledik ve uygulamamız ile görme engelli vatandaşlarımıza yardımcı olacak, onlara rehberlik yapabilecek bir uygulama geliştirmiş olduk. Uygulama sayesinde görme engelli kişiler günlük hayatta sık olarak kullandıkları nesneleri bulmada daha rahat edecek ve çeşitli yaralanmalara sebep olan kazaların bir miktarda olsa önüne geçmiş olacaklardır.

Gelecek çalışmalarda daha fazla nesnenin olduğu, daha hassas konumlandırmanın yapıldığı bir mimari geliştirmeyi hedeflemekteyiz. Ayrıca görme engelli kişiler üzerinde uygulamanın kullanılabilirliği ile ilgili deneyler yapılacaktır.

Teşekkür

Bu çalışma, TÜBİTAK tarafından 1919B011703287 nolu proje ile desteklenmektedir. Desteklerinden dolayı TÜBİTAK'a teşekkür ederiz.

Kaynaklar

1. Pixie anketi, <https://getpixie.com/blogs/news/lostfoundsurvey> (Eriřim Tarihi, 14 Ağustos 2018)
2. M. Serter Uzer, N. Yılmaz, M. Bayrak: “Görme Tabanlı Mobil Robot İle Farklı Renklerde Nesnelerin Gerçek Zamanlı Takibi”, Gazi Üniv. Müh. Mim. Fak. Der., 25(4): 759-766, 2010.
3. S. Meltek, B. Çetiřli: “Gerçek Zamanlı Endüstriyel Nesne Tanımda Kameraların Kullanılması”, Süleyman Demirel University Journal of Natural and Applied Sciences, 16(2): 212-217, 2012.
4. Ö. Altınkurt, M. Kahrıman: “Gerçek Zamanlı Olarak, Anfis İle Renk Tabanlı Nesne Tespit Ve Motorlu Sistem İle Takip Edilmesi”, SDU Journal of Technical Sciences, 1(1): 1-5, 2011.
5. A. Köksal, F.E. Uzyıldırım, M. Özuysal: “Üç Boyutlu Sahneler ve Nesne Tanıma için Gürbüz Anahtar Nokta Eřleřtirilmesi”, İzmir İleri teknoloji Enstitüsü, İzmir, 2017
6. R. Girshick, J. Donahue, T. Darrell, and J. Malik: “Rich feature hierarchies for accurate object detection and semantic segmentation”, *arXiv:1311.2524*, 22 Ekim 2014.
7. R. Girshick: “Fast R-CNN”, Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), 1440-1448, 2015.
8. S. Ren, K. He, R. Girshick, and J. Sun: “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149, 2017.
9. K. He, G. Gkioxari, P. Dollár, and R. Girshick: “Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow”, *arXiv:1703.06870*, 24 Ocak 2018.
10. Redmon and A. Farhadi: “YOLO9000: Better, Faster, Stronger”, Proceedings of Conference on Computer Vision and Pattern Recognition, 7263-7271, 2017.
11. <https://www.aipoly.com/> (Eriřim Tarihi, 14 Ağustos 2018)
12. <https://taptapseeapp.com/> (Eriřim Tarihi, 14 Ağustos 2018)

13. Tensorflow Object Detection API,
https://github.com/tensorflow/models/tree/master/research/object_detection
(Erişim Tarihi, 14 Mayıs 2018)
14. Google Text-to-Speech, <https://cloud.google.com/text-to-speech/>
(Erişim Tarihi, 14 Mayıs 2018)