Mask R-CNN ile Çoklu Barkod Tespiti
Multiple Barcode Detection with Mask R-CNN

Enes Polat, Hussein Mahmood Abdo Mohammed, Asli Nur Omeroglu, Nida Kumbasar, I.Yucel Ozbek, Emin Argun Oral
Atatürk Üniversitesi, Erzurum, Türkiye
enes.polat@turktelekom.com.tr, halameri080@gmail.com, asli.omeroglu@atauni.edu.tr, nidakumbasar@gmail.com , iozbek@atauni.edu.tr, eminoral@atauni.edu.tr

***Özetçe*— Günümüzde kullanılan birçok ürünün kendileriyle ilişkili benzersiz kodu veya kimliği vardır. Bu özel kimlik barkod olarak adlandırılır. Barkodlar çeşitli endüstriyel ortamlarda otomasyona olan yüksek talep nedeniyle son yıllarda kapsamlı araştırmalara konu olmuştur. Birçok ticari uygulamada kullanılan ürün ile ilgili tüm detayların öğrenilebildiği barkodların hızlı ve doğru okuması oldukça önemlidir. Bu çalışmada 1B barkodların görüntüdeki bölgelerini tespit etmek amacıyla, Mask R-CNN algoritması kullanılmıştır. Mask R-CNN ile görüntüdeki barkodlar tespit edilmiş, her barkodun sınırlayıcı kutu konumunun yanı sıra, sınırlayıcı kutudaki bu sınıfa karşılık gelen piksel bilgileri de görselleştirilmiştir. Farklı ortam ışıklarında ve farklı açılarla çekilmiş çeşitli ürünler üzerindeki renkli barkodlar toplanarak 1114 görüntüden oluşan yeni bir veri seti hazırlanmıştır. Bu veri seti kullanılarak Mask R-CNN ile başlangıç çalışması olarak %74.41 doğruluk oranı elde edilmiştir.**

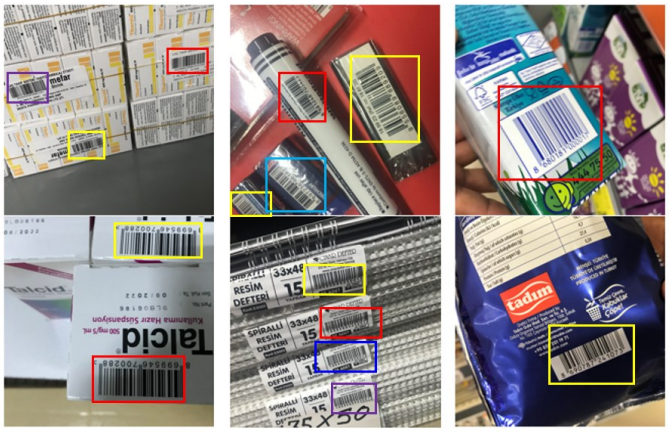
Anahtar Kelimeler — Barkod tespiti; Mask RCNN.

***Abstract*— Almost all products on the market today have a unique code or ID associated with them. This special identification is called a barcode. Barcodes have been the subject of extensive research in recent years due to the high demand for automation in various industrial environments. Fast and accurate reading of barcodes, where all details about the product used in many commercial applications can be learned, is very important. In this study, Mask R-CNN algorithm was used to determine the regions of the 1B barcodes in the image. In the Mask R-CNN, barcodes in the image have been detected, as well as the bounding box position of each barcode, as well as the pixel information corresponding to this class in the bounding box. Colored barcodes on various products taken at different ambient lights and at different angles were collected and a data set of 1114 images was prepared. Using this dataset, 74.41 % accuracy was achieved with Mask R-CNN.**

Keywords — Barcode detection; Mask R-CNN

I. GİRİŞ

Son yıllarda, sanal alışveriş sürekli büyüme göstermiştir. Bu nedenle akıllı telefonlarını veya tabletlerini kullanarak ürün satın alan müşteri sayısı da gün geçtikçe artmaktadır. Birçok araştırmacı ve şirket genellikle yüksek kaliteli kameraları içeren bu cihazlardan çekilen görüntülerde, görsel özelliklerine dayanarak gösterilen ürünleri belirleme sorununu çözmeye odaklanmıştır [1]. Barkodlar günümüzde hayatın vazgeçilmez bir parçasıdır ve yerel süpermarketlerden reklamcılık kullanımına kadar birçok farklı alanda kullanılmaktadır.



Şekil. 1. Veri Setinden Örnek Görüntüler

Adını İngilizce'de bar (çubuk) ve code (şifreleme) kelimelerinin birleşmesinden alan barkodlar üzerinde bulunduğu nesne hakkında bilgiler taşıyan küçük etiketlerdir ve bilgisayarlar tarafından hızlı bir şekilde tanımlanabilirler. İlk olarak 1952 yılında Norman Joseph Woodland ve Bernard Silver tarafından, ürünlerin otomatik olarak tanınması ihtiyacını karşılamak için önerilmiştir. 1B barkodların tespiti üretim, perakende, lojistik ve nakliye gibi sektörlerde çeşitli uygulamalara sahiptir. Geleneksel insan eliyle çalışan lazer barkod okuyucular yerine otomasyon sistemlerinin yaygınlaşması nedeniyle son yıllarda barkod tespitinin cihazlar (robotlar) tarafından yapılması ihtiyaç haline gelmiştir. Bu amaç doğrultusunda barkodların ürün üzerindeki yerlerinin kamera yardımıyla otomatik olarak tespit edilmesi ve sonrasında çözümlenmesi gerekmektedir. Otomatik barkod algılama işlemi temel olarak barkod bölgesinin ve varsa dönme açısının belirlenmesini amaçlamaktadır. Bununla birlikte zorlu görsel ortamlarda barkodların yüksek doğruluk ve hız ile tespit edilmesi hala üzerinde çalışılmaktadır.

Bu çalışmada, görüntüde bulunan tüm barkodların bölütleme tabanlı aynı anda tespiti amaçlanmıştır. Bu çalışmanın ayrıntıları: II. Bölüm barkod tespiti üzerine literatürde yapılan çalışmaları sunar. Bölüm III'te, tespit problemi için kullanılan veri seti, öğrenme algoritması tanıtılır.

Sonuç ve gelecek çalışmalar Bölüm IV'te tartışılırken, elde edilen deneysel sonuçlar Bölüm V'te verilmiştir.

978-1-7281-7206-4/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: ULAKBIM UASL - Ataturk University. Downloaded on June 18,2021 at 07:52:38 UTC from IEEE Xplore. Restrictions apply.

II. İLGİLİ ÇALIŞMALAR

Son yıllarda, bilgisayarlı görmede barkod algılamaya odaklanan yöntemler hızla gelişmiştir. Önceki yaklaşımlar, kullanıcıların daha iyi bir performans elde etmesi için tespitten önce kameranın el ile barkoda hizalanmasını gerektirmekteydi. Verimliliği artırmak için son çalışmalarda daha az etkileşimli bir yaklaşım olarak barkod algılamanın tam otomatik olarak gerçekleştirilmesi hedeflenmiştir.

Literatürde barkodların tespitine ilişkin birçok farklı çözüm önerilmiştir. Barkodları bulmaya çalışan ilk çalışmalardan biri Muniz ve arkadaşları [1] tarafından geliştirilen İspanyol tıbbi reçetesindeki barkodları otomatik olarak işleyen bir uygulamadır. Ancak teknoloji yıllar içinde geliştikçe, daha fazla yöntem önerilmiş, kameralı cep telefonlarının gösterdiği gelişim ile birlikte mobil cihazlarla barkodların ürün numaralarını ve bulundukları bölgeleri tanımlamak için çeşitli yöntemler geliştirilmiştir [2, 3].

Öte yandan sinir ağları bilgisayarla görme dahil birçok alanda çok umut verici sonuçlar vermekte ve barkod tespiti dahil farklı alanlarda kullanılmaktadır. Yapılan ilk çalışmalarda sinir ağı muhtemel barkodları bulmak için Hough uzayı analizini kullanmıştır [4]. Yapay sinir ağı kullanılarak barkod lokalizasyon algoritması için evrişimli bir sinir ağının yapısı geliştirilmiştir [5]. Barkodlarını tespit etmek için daha yeni ve umut verici sonuçlar YOLO (Sadece Bir Kez Bak) dedektörü kullanılarak elde edilmiş sonrasında tespit edilen barkodun dönme açısını belirlemek için başka bir sinir ağı kullanılmıştır [6]. Bir diğer çalışmada kamera ve video görüntüleri ile elde edilen barkodların bulunduğu bölgeyi tespit etmek için SSD (Single Shot Detector) algoritması kullanılmıştır [7]. Barkod tespitinde Derin Sinir Ağları, semantik bölütleme ve farklı geometrik yaklaşımların kullanılmasıyla doğruluk oranı ve verimlilik arttırılmıştır [8-10].

III. YÖNTEM

A. Mask R-CNN

He ve arkadaşları tarafından sunulan Mask R-CNN [11] aynı anda her örnek için yüksek kaliteli bir segmentasyon maskesi oluştururken görüntüdeki nesneleri algılar. Bu mimari Şekil 2 ile gösterilmiştir. Mask R-CNN geliştirilmiş Faster R-CNN [12] tabanlı bir nesne tespiti ağı olup, aşağıdaki işlem adımlarını gerçekleştirmektedir:

- İşlenecek görüntü, öznitelik haritaları elde edilmek üzere önceden eğitilmiş bir ResNet-101 ve Öznitelik Piramit Ağı (Feature Pyramid Network, FPN) [13] modeline uygulanır.
- Elde edilen öznitelik haritası, RPN (Region Proposal Network) den geçirilerek çok sayıda aday çerçeve (Region of Interest, RoI) elde edilir ve bütün lokal bölgelerin nesne bulunabilirlik skorları ve sınırlayıcı kutu koordinatları, sırasıyla ikili sınıflandırma ve regresyon ile belirlenir.
- Öznitelik haritası ve elde edilen RoI bölgeleri RoI Align katmanına uygulanarak sabit boyutlu RoI/öznitelik eşlemesi oluşturulur. RoI Align'daki en önemli adım bilinear enterpolasyon kullanımıdır [14].
- Son adımda, ağ iki ayrı kola ayrılır. Biri nesnenin ait olduğu sınıfı tespit etmek ve sınırlayıcı kutu regresyonu için Tam Bağlantılı Katmana (FC Layer) diğeri ise piksel bölütleme için Tam Evrişim Ağına (FCN) uygulanır.

B. Bölge Öneri Ağı (Region Proposal Network, RPN)

RPN evrişimli sinir ağının yapısına dayanan kayan pencere tabanlı bir nesne algılayıcıdır Bu ağ, belirli bir bölgede başka işlemlerin yapılması gerekip gerekmediğini değerlendirmek için görüntünün her bölgesini hızlı ve verimli bir şekilde tarar. Bunu, her bir bölgede nesne olasılığını temsil eden bölge öneri çıktıları ile yapar. Kayan çerçeve taraması referans amaçlı ankraj kutuları (anchor box) üretip farklı nesne türlerini barındırmak için farklı en-boy oranlarına ve ölçeklere sahip olacak şekilde seçilirler. Önerilen bir alan farklı boyut ve en-boy oranlarında çok sayıda ankraj oluşturup, mümkün olduğu kadar çok görüntüyü kapsayacak şekilde üst üste gelir. Önerilen alanın büyüklüğü ve istenen alan çakışması (IoU) sınıflandırma etkisini doğrudan etkiler. Birkaç ankraj çok fazla örtüşürse, en yüksek kesişimin birleşime oranı (Intersection over Union, IoU) değerine sahip olanı seçilir.

C. Mimari Yapısı

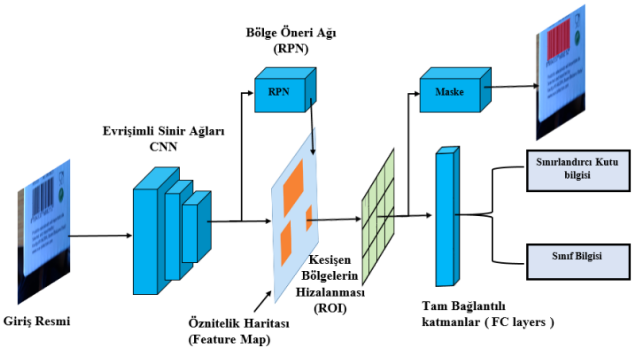
Görüntülerde yer alan barkod bölgelerinin boyutu farklı olduğundan, tek bir evrişimli sinir ağı tüm görüntü özelliklerini belirleyemez. Bu nedenle, bu çalışmada ResNet-101 omurga yapısı ve bir FPN kullanılmıştır. Evrişim katmanı, RPN ağına girdiğinde bazı basit öznitelikler çıkarılsa da küçük hedefler gizli katmalarda gürültü olarak filtrelenebilir. FPN küçük hedef özelliklerinin filtrelenmesini önleyerek CNN tarafından elde edilen öznitelik haritasını RPN ağına uygulamaktır. Sığ özellikler basit büyük hedefleri belirlerken derin özellikler küçük hedefleri ayırt edebilir. FPN tarafından üretilen farklı boyut öznitelik haritaları RPN'e uygulanır [15]. Daha sonra RPN hedef nesnenin boyutuna göre öznitelik piramidinin farklı seviyelerinden RoI özelliklerini çıkarabilir. Böylece basit ağ yapısı hesaplama yükünü çok fazla arttırmadan, küçük nesnelerin algılanma performansını geliştirip, sonuç olarak doğruluk ve hızda iyileşme sağlar.

IV. DENEYSEL ÇALIŞMALAR

A. Veri Seti

Önerilen barkod algılama ve tanıma yöntemi için günlük yaşamda yaygın olarak kullanılan çeşitli ürünleri içeren eczane, kırtasiye ve süpermarketlerden toplanan yeni bir veri kümesi hazırlanmıştır. Veri setinde yer alan görüntüler 12 Megapiksel çözünürlüğe sahip bir iPhone7 cep telefonu kamerası kullanılarak elde edilmiştir. Veri seti gerçek dünyadaki durumu en iyi şekilde taklit etmek için farklı döndürme ve barkod bölgesinin çözünürlüklerine sahip 1114 görüntüden oluşmaktadır. Veri setindeki resimlerde farklı sayılarda barkod bulunmakta olup 100 tanesi bir adet, 200 tanesi iki adet, 814 tanesi üç ve daha fazla sayıda barkod bölgesini içerirler.

Authorized licensed use limited to: ULAKBIM UASL - Ataturk University. Downloaded on June 18,2021 at 07:52:38 UTC from IEEE Xplore. Restrictions apply.



Şekil.2 Mask R-CNN genel mimarisi

Aşırı parlama, ışık azlığı ve çoklu barkod örneklerinin yer aldığı veri seti görüntülerinin boyutları 3024x3024 olup, veri setinden bazı örnekler Şekil 1'de gösterilmiştir. Tablo.1'de ise barkod tespiti için literatürde kullanılan bazı veri setleri ve bunlarda yer alan barkod sayıları verilmiştir. Veri setinde yer alan barkod görüntülerinin etiket dosyaları ise VGG Image Annotator (via 2.0.8) ile manuel olarak oluşturulmuştur [16].

Çalışmada her bir veri seti görüntüsü 1280x1280 boyutuna yeniden şekillendirilmiş ve 814 eğitim, 150 doğrulama ile 150 test verisi yer alacak şekilde veri seti üç kısma ayrılmıştır.

Tablo.1 Farklı barkod veri setlerinin karşılaştırılması

Kaynak	Barkod Veri Seti Adı	Barkod Sayısı
[4]	ArTe-Lab Medium Barcode	365
[18]	EGO-Raw	1060
[19]	WWU Muenster Barcode (Muenster)	1055
[20]	-	100
Önerilen Dataset	-	**1114**

B. Ağın Eğitimi

Modelin eğitimi üç adet NVIDIA GTX 2080 Ti ekran kartı bulunan bir bilgisayar ile gerçekleştirilmiştir. Toplam 160 epok (epoch) ile 814 eğitim verisi yaklaşık 3 saatte eğitilerek model oluşturulmuştur. Referans model olarak Matterport tarafından TensorFlow ve Keras kütüphaneleri ile gerçekleştirdiği Mask R-CNN ağ yapısı kullanılmış [16] ve modelin hiper-parametreleri değiştirilerek barkod tespiti için en iyi model oluşturulmaya çalışılmıştır. Eğitimde ilk 40 epokta Mask R-CNN mimarisinin baş katmanları (head layers) eğitilmiş, daha sonraki 80 epokta mimarinin alt kısımları da dahil edilerek modelin ağırlıkları güncellenmiştir. Son olarak ağ yapısı uçtan uca 40 epok boyunca eğitilerek, öğrenme modeli elde edilmiştir. Mimarinin eğitiminde öğrenme oranı (learning rate) 0.001 olarak seçilmiştir. ResNet-101 mimarisinde önceden eğitilmiş MS COCO [17] veri seti ağırlıkları başlangıç ağırlık değerleri olarak kullanılmıştır. Ankraj boyutları en-boy oranı olarak 0.5, 1, 2 ve pencere boyutları olarak 8, 16, 32, 64, 128 değerleri kullanılmıştır.

C. Performans Ölçütü

Nesne tespiti problemlerinde kullanılan semantik bölütleme performansını değerlendirmek için kullanılan yöntemlerden birisi görüntüdeki doğru sınıflandırılmış piksellerin yüzdesini belirlemektir. Bu amaçla kullanılan Jaccard indisi semantik bölütlemede en yaygın olarak kullanılan metriklerden biridir. Son derece etkili olan bu metrik iki kümenin ortak özelliklerinin tüm özelliklerine oranı biçiminde ifade edilir. Denklem 1 ile her örnek veri için etiketli segmentlerin kesişimin birleşime oranının sınıflar üzerinden ortalaması olarak hesaplanan Jaccard indisi verilmiştir. Öte yandan, piksel doğruluğu, her sınıf için ayrı ayrı ve genel olarak hesaplanır. Sınıf başına piksel doğruluğu hesaplanırken; biri verilen sınıfa ait olup, doğru şekilde tahmin edilen piksel (Gerçek Pozitif), diğeri ise verilen sınıfa ait olmayıp, yine bu şekilde doğru olarak tahmin edilen piksel (Gerçek Negatif) olmak üzere iki ayrı piksel kümesi (maske) belirlenir. Sonrasında her sınıf için etiketli segmentlerin doğruluğu Jaccard indisi ile belirlenir.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \qquad (1)$$

D. Deneysel Sonuçlar

Çoklu barkodu algılamak için kullanılan Mask R-CNN algoritmasının etkinliğini değerlendirmek için ayrıntılı deneyler yapılmıştır. Bu deneyler neticesinde Resnet-101 mimarisinin ResNet-50'ye göre daha yüksek algılama başarısına sahip olduğu gözlemlenmiştir. Karmaşık ortamlarda algılama doğruluğu ölçülmeye çalışılan algoritma, farklı görüntü boyutları, epok sayıları, öğrenme oranları ve ankraj boyutlarında eğitilmiştir. Ayrıca, bu algılama yönteminin bazı bozulmalara karşı dayanımını gözlemleyebilmek için test veri setine farklı görüntü niteliklerine sahip barkod görüntüleri seçilmiştir. Test seti kullanılarak doğru algılanan nesneler için piksel doğruluğu ölçülmüş, %74.41 doğruluk oranı elde edilmiştir. Şekil 4'de 150 adet görüntüden oluşan test kümesi örnek resimleri ve model test çıktıları gösterilmiştir. Model test çıktıları sırasıyla beklenen çıktıları ve olması gerekenden daha fazla tespit edilen barkod bölgeleri şeklindeki hataları içermektedir.

V. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada Mask R-CNN ile karmaşık ortamlarda barkod algılama ve tanıma için, günlük yaşamda yaygın olarak kullanılan çeşitli ürünlerden toplanan yeni bir veri kümesi önerilmektedir. Literatürde 80 farklı nesnenin tespiti için eğitilmiş COCO ağırlıkları kullanılarak Mask R-CNN'in farklı ürünler üzerindeki barkodları tespit edebilmesi sağlanmıştır. Başlangıç çalışması olarak %74.41 doğruluk oranı elde edilmiştir. Gelecekteki çalışmalarda, hazırlanan veri setinin farklı barkod tipleri ile zenginleştirilmesi, farklı nesne tespit mimarileri ile bölgesel tespit edilme sonuçlarının iyileştirilmesi ve çözümlenmesi planlanmaktadır

Authorized licensed use limited to: ULAKBIM UASL - Ataturk University. Downloaded on June 18,2021 at 07:52:38 UTC from IEEE Xplore. Restrictions apply.



Şekil.4. Modelin test veri setindeki örnek resimlerde uygulanması (Hatalı bulunan barkodlar sarı, tespit edilemeyen barkodlar kırmızı halka ile gösterilmiştir.)

KAYNAKLAR

[1] Muniz, Ruben, Luis Junco, and Adolfo Otero. "A robust software barcode reader using the Hough transform." *Proceedings 1999 International Conference on Information Intelligence and Systems*. IEEE, 1999.

[2] J. *C. Rocholl, S. Klenk and G. Heidemann, "Robust 1B Barcode Recognition on Mobile Devices," 2010 20th International Conference on Pattern Recognition, Istanbul, 2010, pp. 2712-2715*

[3] *W. P. Fernandcz, Y. Xian and Y. Tian, "Image-Based Barcode Detection and Recognition to Assist Visually Impaired Persons," 2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Honolulu, HI, 2017, pp. 1241-1245.*

[4] *A. Zamberletti, I. Gallo, and S. Albertini, "Robust angle invariant 1B barcode detection," in 2013 2nd IAPR Asian Conference on Pattern Recognition, Nov 2013, pp. 160–164.*

[5] *N. N. Ventsov and L. A. Podkolzina, "Localization of Barcodes Using Artificial Neural Network,"* 2018 IEEE East-West Design & Test Symposium (EWDTS)*, Kazan, 2018, pp. 1-6.*

[6] *D. Kold Hansen, K. Nasrollahi, C. B. Rasmusen, and T. Moeslund, "Real-time barcode detection and classification using deep learning," 01 2017, pp. 321–327.*

[7] *N. N. Ventsov and L. A. Podkolzina, "Studying the effect of paralleling settings on the functioning of a barcode recognition app,"* 2017 International Conference on Industrial Engineering, Applications and *Manufacturing (ICIEAM), St. Petersburg, 2017, pp. 1-5.*

[8] *Xiao, Yunzhe, and Zhong Ming. "1B Barcode Detection via Integrated Deep-Learning and Geometric Approach." Applied Sciences 9.16 (2019): 3268.*

[9] Zharkov, Andrey, and Ivan Zagaynov. "Universal Barcode Detector via Semantic Segmentation." arXiv preprint arXiv:1906.06281 (2019).

[10] Li, Jianjun, et al. "Using deep convnet for robust 1B barcode detection." International Conference on Intelligent and Interactive Systems and Applications. Springer, Cham, 2017.

[11] K. He, G. Gkioxari, P. Dollr, R. Girshick, "Mask R-CNN", IEEE International Conference on Computer Vsion (ICCV), pp. 2980-2988, 2017

[12] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.T.-Y. Lin, P. Dollar, R. Girshick, K. He, B.

[13] Hariharan, and S. Belongie, ''Feature pyramid networks for object detection,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 320–329. X

[14] S WANG, K J YANG, "An Image Scaling Algorithm Based on Bilinear Interpolation with VC++", Techniques of Automation and Applications, no. 07, pp. 44-45+35, 2008.

[15] X. Zhang, J. Zou, X. Ming, K. He, and J. Sun, ''Efficient and accurate approximations of nonlinear convolutional networks,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1984–1992

[16] A. Dutta, A. Gupta, A. Zisserman, Vgg image annotator (via), 2016, [online], Available: http://www.robots.ox.ac.uk/~vgg/software/via

[17] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, "Microsoft coco: Common objects in context", European conference on computer vision, pp. 740-755, 2014

[18] Creusot, Clement, and Asim Munawar. "Low-computation egocentric barcode detector for the blind." 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016.

[19] S. Wachenfeld, S. Terlunen, X. Jiang, "Robust recognition of 1-D barcodes using camera phones", *Proc. 19th Int. Conf. Pattern Recognit. (ICPR)*, pp. 1-4, 2008.

[20] M. Katona, L. Nyúl, "Efficient 1D and 2D barcode detection using mathematical morphology", *Proc. Math. Morphol. Appl. Signal Image Process.*, pp. 464-475, 2013.

Authorized licensed use limited to: ULAKBIM UASL - Ataturk University. Downloaded on June 18,2021 at 07:52:38 UTC from IEEE Xplore. Restrictions apply.