



EVRIŞİMLİ SİNİR AĞI TABANLI DERİNLİK HARİTASI İYİLEŞTİRME

Alper EMLEK^{1,*} , Murat PEKER² 

^{1,2} Niğde Ömer Halisdemir Üniversitesi, Elektrik-Elektronik Mühendisliği Bölümü, 51240, Niğde Türkiye

ÖZET

Bir ortamın derinlik bilgisi üç boyutlu yeniden oluşturma, otonom sistemler gibi derinlik bilgisine ihtiyaç duyulan birçok uygulamada yoğun olarak kullanılmaktadır. Stereo görüntü çiftleri kullanılarak derinlik bilgisi elde edilmesi, uzun yıllardır öne çıkan çalışma konuları arasında yer almaktadır. Son yıllarda derin öğrenme alanındaki gelişmeler, stereo görüntüler kullanılarak derinlik elde edilmesi alanında yapılan çalışmaları da etkilemiştir. Son dönemde, derin öğrenme tabanlı stereo eşleştirme alanındaki çalışmalar daha çok derinlik (eşitsizlik) haritalarının elde edilmesi ve düzeltilmesi alanında yoğunlaşmaktadır. Derin öğrenme tabanlı derinlik haritası düzeltme işlemlerinde eşleşme maliyetleri üzerinden yapılan düzeltme işlemleri daha başarılı sonuçlar üretilmesini sağlamıştır. Fakat eşleşme maliyetlerine doğrudan ulaşmak her zaman mümkün olmamaktadır. ZED ve Intel RealSense gibi hazır sistemler tarafından üretilen derinlik haritaları son kullanıcıya sunulmaktadır. Çalışmada sadece sol görüntü ve derinlik haritası aracılığıyla düzeltme işlemi yapan bir ağ yapısı önerilmiştir. Önerilen ağ, KITTI 2012 ve KITTI 2015 veri kümeleri üzerinde test edilmiştir. Deneysel çalışmalar sonucunda olarak giriş derinlik haritalarının doğruluğunun artırılması için önerilen ağ yapısının kullanılabilceği nicel ve nitel sonuçlar ile gösterilmiştir.

Anahtar kelimeler: Stereo eşleşme, Derinlik haritası, Derinlik haritası düzeltme, Evrişimli sinir ağları

IMPROVING DEPTH MAP BASED ON CONVOLUTIONAL NEURAL NETWORK

ABSTRACT

Depth information of an environment is used extensively in many applications that require depth information such as three-dimensional (3D) reconstruction and autonomous systems. Obtaining depth information using stereo image pairs is one of the prominent study topics for many years. In recent years, developments in the field of deep learning have also affected the studies in the field of obtaining depth using stereo images. These studies mostly focus on obtaining and correcting depth (disparity) maps. In deep learning based depth map correction, correction operations over matching costs have enabled more successful results. However, it is not always possible to get matching costs directly. Disparity maps produced by some of the existing systems are provided to the end-user. In this study, a network structure that corrects only one image and disparity map are proposed. The proposed network has been tested on KITTI 2012 and KITTI 2015 datasets. As a result, it is shown with quantitative and qualitative results that the proposed network structure can be used to increase the accuracy of the entrance depth maps.

Keywords: Stereo matching, Depth map, Depth map correction, Convolutional neural networks

1. GİRİŞ

Stereo eşleştirme, bilgisayarla görme alanında uzun zamandır çalışılan önemli konular arasında yer almaktadır. Stereo eşleştirme, aynı yatay düzlem üzerinde hizalanmış kamera çiftinden alınan görüntüler arasında eşleşen piksellerin tespit edilmesi problemi olarak ifade edilebilir. Yatay düzlemde yer alan kameralardan alınan görüntüler sol ve sağ görüntü olarak ifade edilmektedir. Sol görüntüde yer alan her bir piksele karşılık gelen ve sağ görüntüde yer alan bir piksel olduğu kabul edilmektedir. Eşleşen her piksel çifti için derinlik bilgisi, eşleşen piksellerin yer aldıkları görüntülerdeki konumları arası yatay uzaklık kullanılarak hesaplanmaktadır. Elde edilen yatay uzaklık, derinlik (eşitsizlik) değeri olarak isimlendirilir ve d ile gösterilir. Böylece, sol görüntüde (x, y) konumunda yer alan pikselle eşleştirilen pikselin sağ görüntüdeki konumu $(x - d, y)$ ifadesi ile hesaplanabilir. Burada x ve y sırasıyla görüntünün yatay ve dikey eksenini üzerinde yer alan piksel konumlarını belirtmektedir. Eşleştirilen pikselin üç boyutlu (3B) düzlemdeki derinlik değeri (z) ise Eşitlik 1 kullanılarak hesaplanmaktadır.

* Sorumlu yazar / Corresponding author, e-posta / e-mail: alperemlek@ohu.edu.tr

Geliş / Recieved: 25.03.2020 Kabul / Accepted: 08.07.2020 doi: 10.28948/ngumuh.709159

A. Emlek, M. Peker

$$z = f\left(\frac{B}{d}\right) \quad (1)$$

Eşitlik 1'deki terimlerden z ilgili pikselin 3B düzlemdeki derinlik değerini, f sağ ve sol kameralarının odak uzaklığını ve B terimi ise iki kamera merkezi arasındaki mesafeyi temsil etmektedir.

Bir sahnenin derinlik bilgisinin elde edilmesi, sahnenin yeniden modellenmesi, nesnelerin üç boyutlu yeniden oluşturması, robotik ve otonom sürüş gibi birçok gerçek dünya uygulaması için hayati önem taşımaktadır. Tek bir görüntüden derinlik bilgisi tahmin etme veya aktif algılayıcılar kullanarak derinlik hesaplama konusundaki ilerlemelere rağmen, stereo görüntü eşleştirme kullanılması verimlilik ve geniş uygulanabilirlik gibi avantajları nedeniyle derinlik tahmini için hala en etkili stratejilerden biri olmaya devam etmektedir. Fakat yapılan onlarca yıllık araştırmaya rağmen, görüntü çifti üzerinden derinlik tahmin etmek için ortaya atılan yöntemler hala bazı sorunlar içermektedir. Stereo eşleştirme işleminin içerdiği sorunlar arasında en önemlileri; nesnelere örtüşme (Occlusion), geniş doku içermeyen alanlar ve görüntüde tekrarlayan desenlerdir.

Son yıllarda derin evrişimli sinir ağları (Deep Convolutional Neural Networks - DCNN) makine öğrenmesi ve makine görmesi alanında oldukça popüler bir çalışma konusu haline gelmiştir. Derin evrişimli sinir ağları güçlü temsil kabiliyetleri nedeniyle, stereo eşleştirme probleminde başarıyı artırmak için kullanılmaktadır ve geleneksel yöntemler ile karşılaştırıldığında önemli ölçüde bir iyileştirme sağlamaktadır. Stereo eşleştirme alanındaki evrişimli sinir ağları (Convolutional Neural Networks - CNN) tabanlı geliştirilen ilk çalışmada Zbontar ve LeCun [1], stereo görüntülerdeki piksellerin benzerlikleri bir siyam CNN kullanılarak karşılaştırılmış ve pikseller arası benzerlik puanı elde edilmiştir. Ardından, elde edilen benzerlik puanları ile eşleşme maliyetleri hesaplanmıştır. Stereo eşleştirme alanındaki evrişimli sinir ağı tabanlı çalışmalarda sadece eşleşme maliyetlerinin hesaplanması [1,2] için değil, ayrıca, maliyetlerin toplamı, derinlik haritası hesaplama ve derinlik haritası düzeltme gibi diğer amaçlar doğrultusunda da CNN yapıları sorunsuz bir şekilde entegre edilmiştir[3-5]. Ayrıca, geliştirilen yöntemlerin yanı sıra uçtan uca öğrenme ile doğrudan derinlik haritası elde etmeye yönelik ve hesaplama verimliliğini veya eşleşme doğruluğunu artırmak için çeşitli yöntemler ve ağ yapıları da önerilmiştir [13,16,17].

Çalışmada, literatürdeki çalışmalardan farklı olarak sadece sol görüntü ve elde edilmiş bir derinlik haritasının kullanılması ile derinlik haritasında yer alan bilgilerin iyileştirilmesi gerçekleştirilmiştir. Bu kapsamda CNN tabanlı bir ağ yapısı önerilmiştir. Önerilen ağ yapısı stereo kameralardan elde edilen görüntülerden sol kameraya ait olan görüntü ile literatürde yapılmış bir çalışma olan Content-CNN [2] ağından elde edilen derinlik haritasını giriş olarak kabul etmektedir. Önerilen yapı ile sadece iki giriş bilgisi kullanılarak derinlik haritasında yer alan hatalı bölgelerde düzeltmeler ve derinlik bilgisinde iyileştirmeler elde edilmektedir. Önerilen ağ yapısı, derinlik bilgisinin elde edilebildiği ve derinlik haritasının iyileştirilmeye ihtiyaç duyulduğu durumlarda kullanılabilir şekilde basit ve uygulanabilir bir yapıdadır.

Bölüm 2'de literatürde önde gelen çalışmalar özetlenmiştir. Bölüm 3'te derinlik haritası düzeltmek için önerilen ağ yapısının detayları verilmiştir. Bölüm 4'te deneysel sonuçlar ardından Bölüm 5'te ise elde edilen sonuçlar ve gelecek çalışmalar hakkında bilgiler verilmiştir.

2. İLGİLİ ÇALIŞMALAR

Stereo görüntüden derinlik bilgisi elde edilmesi için kullanılan yöntemler geleneksel ve derin öğrenme tabanlı yaklaşımlar olarak iki grup altında incelenmektedir.

2.1. Geleneksel Yöntemler

Geleneksel stereo eşleştirme algoritmaları tipik olarak, eşleşme maliyeti hesabı, maliyetlerin toplamı, derinlik haritasının hesaplanması/eniyilenmesi ve derinlik haritası düzeltme işlemi olmak üzere dört adımdan oluşmaktadır [9]. Aşamaların ilkinde tüm piksellerin eşleşme maliyetleri tüm olası derinlik değerleri için hesaplanmaktadır. Eşleştirme maliyetleri hesaplanırken mutlak farklılıkların toplamı, kare farkların toplamı, normalleştirilmiş çapraz ilinti, Census dönüşümü ve bu yöntemlerin birleştirilmiş halleri kullanılmaktadır. İkinci adım olarak, maliyetlerin toplamı adımında bir destek penceresi içindeki komşu piksellerin eşleştirme maliyetleri toplanmaktadır. Destek penceresinin oluşturulması ile ilgili birçok yöntem vardır. Destek penceresinin oluşturulması ile ilgili yöntemlerden bazıları: sabit pencereler [9], kenar bilgilerinden elde edilen uyarlamalı pencereler [10] ve ağırlıklı pencerelerdir [11]. Üçüncü adımda, yerel yöntemler minimum eşleştirme maliyetiyle eşitsizliği seçmek için kazanan hepsini alır (Winner-Take-All, WTA) yöntemini kullanır. Küresel yöntemler ise en uygun derinlik haritasını elde etmek için veri ve yumuşatma terimlerinden oluşan bir enerji fonksiyonunu eniyilemeyi amaçlamaktadırlar. Eniyileme işlemi için genellikle inanç yayılımı [12] ve grafik kesimi (GraphCut) [13] gibi algoritmalar kullanılmaktadır. En son adımda ise derinlik haritalarındaki yanlış noktaların tespit edilmesi ve yanlış eşleşen piksellerin doğru derinlik bilgileri ile doldurulması amaçlanmaktadır. Derinlik haritasının düzeltilme adımında piksellerin komşu piksellerle olan ilişkileri [11] veya süper-piksel/bölütleme gibi pikselleri gruplayan yöntemler vasıtasıyla elde edilen grup bilgilere bakarak düzeltme [13] işlemi yapılmaktadır.

EVRIŞİMLİ SİNİR AĞI TABANLI DERİNLİK HARİTASI İYİLEŞTİRME

Geleneksel yöntemlerin avantajları hızlı çalışmaları, eğitime ihtiyaç duymamaları ve kolay uyarlanabilir olmalarıdır. Fakat geleneksel yöntemler daha önce de bahsedilen nesnelere arası örtüşme, geniş dokusuz alanlar ve görüntüde tekrarlayan desenler gibi stereo görüntülerde oluşabilecek hataların önüne geçmekte yetersiz kalmaktadır. Dolayısıyla, derin öğrenme tabanlı yöntemler, geleneksel yöntemlerin yetersiz kaldığı durumlar için daha çok kullanılmaya başlanmıştır.

2.2. Derin Öğrenme Tabanlı Yöntemler

Son zamanlarda derin öğrenme kavramı görüntü işleme alanında yaygın olarak çalışılan ve kullanılan bir problem çözme aracı olarak ön plana çıkmaktadır. Literatüre stereo eşleştirme yöntemlerine CNN tabanlı birçok farklı yaklaşım önerilmiştir. Önerilen yöntemler genellikle üç ana başlık altında toplanmaktadır; eşleştirme maliyetlerinin hesaplanması, aykırı noktaların tespiti ile derinlik haritasının düzeltilmesi ve uçtan-uca öğrenme ile derinlik haritası elde etmedir.

Derin öğrenme tabanlı geliştirilen ilk çalışmada Zbontar ve LeCun [1], stereo görüntülerdeki piksellerin benzerliklerini bir siyam CNN kullanılarak karşılaştırmış ve pikseller arası benzerlik puanı elde edilmiştir. Ardından, elde edilen benzerlik puanı ile eşleşme maliyetleri hesaplanmıştır. [1] çalışmasında, karşılaştırılacak giriş görüntüleri 9×9 veya 11×11 gibi küçük görüntü parçaları olarak ağına girişine uygulanmaktadır. Luo vd. [2] tarafından önerilen Content-CNN ağ yapısında ise girişlerin birine sol görüntüden alınan parça verilirken diğer giriş ise sağ görüntüden alınan arama uzayı boyunca bütün muhtemel olasılıkları içeren görüntü kümesi giriş olarak verilmiştir. Çıkışta tek bir eşleştirme için benzerlik sonucu elde etmek yerine bütün muhtemel sonuçlar için eşleşme maliyeti değeri elde edilmiştir. Content-CNN ağ mimarisinde giriş görüntülerini sıralı evrişim işleminden geçirdikten sonra benzerlikleri hesaplama için ilinti işleminden faydalanmıştır. Brandao vd. [14] ise çalışmada, Luo vd. [2] tarafından geliştirilen çalışmaya çok benzer bir ağ mimarisi önermişlerdir. Fakat çalışmalarında iki büyük farkı vardır. Bunlardan ilki, ilinti işleminden önce uygulanan havuzlama işlemi sırasında kaybolan detayların ters- evrişim işlemi ile telafi edilmesidir. İkinci olarak ise ilinti öğrenme problemini büyük ölçüde basitleştiren bir öznetelik birleştirme yöntemi ile çıkarılan özelliklerin kalitesinin daha iyi değerlendirilmesine olanak sağlanmıştır. Eşleştirme maliyetlerinin hesaplanması amacıyla önerilen ağ mimarileri sadece maliyet hesabı yaptıkları için derinlik haritası elde etme işleminde geleneksel stereo eşleştirme algoritmalarından faydalanmıştır.

Aykırı değer tespiti ve fark haritası düzeltme işlemleri için kullanılacak güven değeri olarak adlandırılan bir ölçüt Poggi ve Mattocchia [3] tarafından elde edilmiştir. [3] çalışmasında ağına giriş olarak derinlik haritasından alınan 9×9 boyutundaki parçaları kullanmıştır. Giriş görüntüsüne uygulanan evrişim işlemi sonunda tam bağlı (fully connected - FC) katmanların çıkışları her bir derinlik değeri için güven değerini belirtmektedir. Ayrıca, Poggi ve Mattocchia [4] çalışmasında, [3] çalışmasından farklı olarak evrişim işlemlerinin sonunda yer alan tam bağlı katmanlar yerine daha fazla evrişim işlemi eklenmiştir. Tam bağlı katman yerine eklenen evrişim katmanları ile daha yüksek doğrulukta güven değerleri elde edilmiştir.

Seki ve Pollefeys [5] çalışmasında da derinlik haritası düzeltmek için güven değerleri hesaplanmıştır. [5] çalışmasında önerilen ağına girişlerinde hem sol hem de sağ kameradan elde edilen derinlik değerleri kullanılmıştır. Ayrıca, derinlik haritalarını doğrudan kullanmak yerine bir ön işleme tabii tuttukten sonra ağa giriş olarak vermişlerdir. Bu çalışmada elde edilen güven değerleri aykırı değer tespiti ve derinlik haritalarının düzeltilmesi işleminde kullanılmıştır. Güven değerleri ile tespit edilen yanlış eşleştirilen pikseller sonraki adımlarda doğru piksel değerleri ile düzeltilmeye çalışılmaktadır.

Cheng vd. [15], doğrudan derinlik haritası veya stereo giriş görüntülerini kullanarak düzeltme işlemi yapmak yerine eşleşme maliyeti yığınlarını da dikkate almışlardır. Bu çalışmada, eşleşme maliyeti yığınları ve başlangıç derinlik haritasının elde edilmesi için Mayer vd. [6] tarafından önerilen DispNet ağından faydalanılmıştır. Çalışmada, güven değeri ve eğitim haritası üreten iki farklı ağ yapısı önerilmiştir. Elde edilen güven değerleri ve eğitim haritası ile birlikte maliyet yığınlarından elde edilen derinlik haritası kullanılarak küresel eniyileme yöntemi ile derinlik haritası elde edilmiştir.

Kim vd. [16] çalışmasında ise güven değeri elde etmek için doğrudan derinlik haritası kullanılması yerine ham eşleştirme maliyetleri kullanılmıştır. Ham eşleştirme maliyeti MC-CNN ağından [1] elde edilmiştir. LAF-Net [17] ağı ise eşleşme maliyetlerini, derinlik haritasını ve referans sol görüntüyü giriş olarak kabul etmiştir ve eşleşme maliyetlerini ve derinlik haritasını MC-CNN'den [1] elde edilmiştir.

Mayer vd. [6] çalışmasında, DispNet olarak adlandırılan ilk uçtan-uca öğrenme yapısı önerilmiştir. Sadece ağ girişine uygulanan sağ ve sol görüntü kullanılarak ağ çıkışında doğrudan derinlik haritası üretilmektedir. Bu çalışmada, kodlayıcı-kod çözücü (coder-encoder) tabanlı bir yaklaşım kullanılmıştır.

Pang vd. [7] çalışmasında, DispNet [6] çalışması geliştirilerek iki aşamadan oluşan Cascade Residual Learning (CRL) ağ mimarisi önerilmiştir. CRL ağı iki aşamadan oluşmaktadır ve ilk aşamada DispNet ağ mimarisi kullanılmıştır. İkinci aşamada ise DispResNet ağı bulunmaktadır ve birinci aşamada üretilen derinlik haritasının düzeltilmesi için kum saati (hourglass) ağ yapısı ile çok ölçekli bir düzeltme işlemi yapılmaktadır.

Chang ve Chen [8] çalışmasında, içerik bilgisini daha iyi aktarabilmek amacıyla uzamsal piramit havuzu kullanılmıştır. Uzamsal piramit havuzu sayesinde farklı boyutlardan küresel içerik bilgisi çıkarılarak maliyet yığınları oluşturmuştur. Maliyet yığınları, yığınlanmış kum saati (stacked hourglass) ağ yapısından geçirilerek derinlik haritası elde edilmiştir.

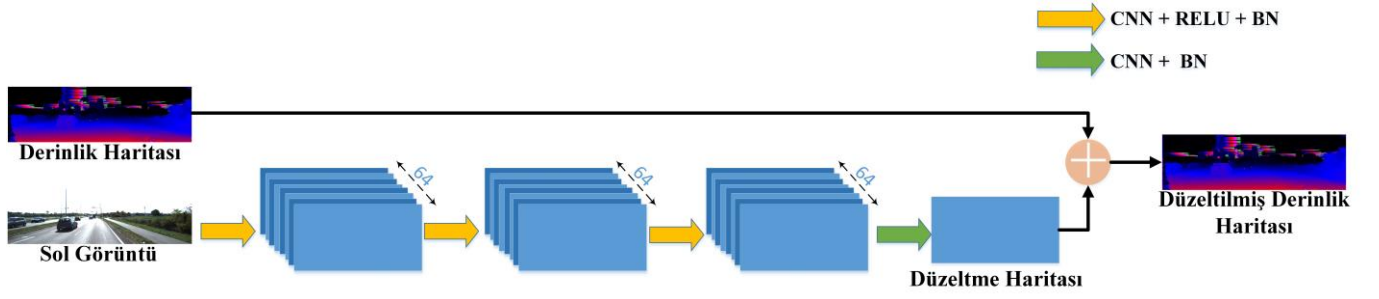
Son zamanlarda yapılan çalışmalar ağırlıklı olarak derin öğrenme tabanlı yöntemlere yönelmektedir. Derin öğrenme tabanlı yöntemlerde de geleneksel yöntemlerle ile iç içe olan ağ yapıları mevcuttur. Fakat genel yönelim stereo eşleşme işlemlerinde yer

A. Emlek, M. Peker

alan bütün adımların derin öğrenme ile gerçekleştirilmesi yönündedir. Derin öğrenme ile gerçekleştirilen ağ yapılarında da başarıyı artırmak için ağ yapıları büyüyerek devam etmektedir. Bu bağlamda, yapılan çalışma ile literatürde yer alan çalışmaların aksine daha az işlem yükü gerektiren ve başarıyı artırmak için mevcut yöntemler ile birlikte kullanılması da mümkün olan bir ağ yapısı önerilmiştir.

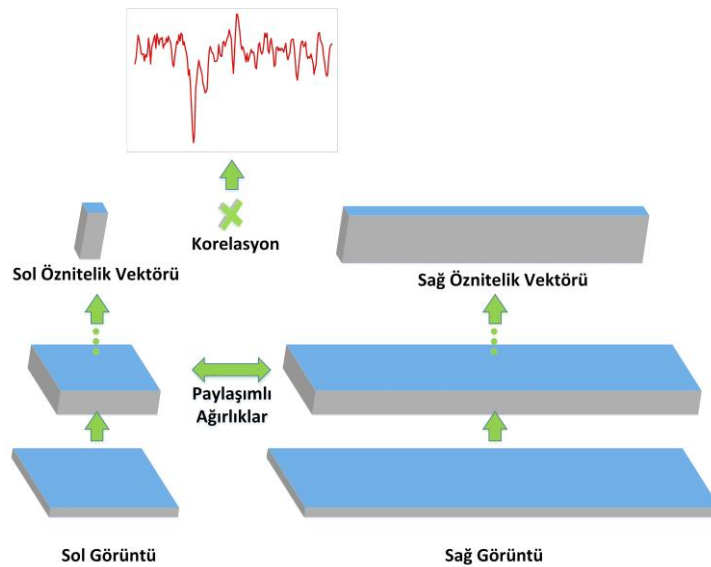
3. YÖNTEM

Önerilen ağ yapısında derinlik haritalarının iyileştirilmesi amacıyla, stereo görüntüler ve derinlik haritalarının giriş olarak kullanıldığı çok katmanlı CNN ağ yapısı önerilmiştir. Literatürdeki çalışmalar, derinlik haritalarının düzeltilmesi için eşleşme maliyetlerinin kullanılmasına dayalı yöntemlerin daha başarılı sonuçlar elde ettiğini göstermektedir. Sadece stereo görüntü ve derinlik haritası kullanmak, eşleşme maliyeti hacimlerinin hesaplayan yöntemler ile karşılaştırıldığında işlem yükü bakımından oldukça düşük seviyede kalmaktadır. Fakat eşleşme maliyetlerine ulaşmak her zaman mümkün olmamakla birlikte bazı hazır sistemler doğrudan stereo görüntüleri ve derinlik haritasını son kullanıcıya sunmaktadırlar. Fakat hazır derinlik haritalarında da hatalar mevcuttur. Çalışma kapsamında önerilen ağ yapısı ile hazır sağlanan derinlik haritası ve sol kameraya ait görüntü kullanılarak giriş derinlik haritasının düzeltilmesi amaçlanmıştır. Önerilen ağ yapısı Şekil 1’de gösterilmiştir.



Şekil 1. Önerilen ağ yapısı

Ağın derinlik haritası girişinde, Content-CNN [2] ağından alınan ham eşleşme maliyetlerine kazanan-hepsini-alır (Winner-Take-All, WTA) yöntemi uygulanarak elde edilen derinlik haritaları kullanılmaktadır. Content-CNN ağı giriş olarak aldığı stereo görüntülerden sol görüntü üzerindeki her bir piksel için eşleşme maliyeti üretmektedir. Eşleşme maliyetleri siyam ağ yapısı kullanılarak elde edilmektedir. Siyam ağlardan ilkinde sol görüntüden alınan 9×9 'luk parça verilirken diğer ağ girişine ise sağ görüntüden alınan arama uzayı boyunca bütün muhtemel olasılıkları içeren görüntü kümesi giriş olarak verilmiştir. Siyam ağın



Şekil 2. Content-CNN ağ yapısı

EVİRİŞİMLİ SİNİR AĞI TABANLI DERİNLİK HARİTASI İYİLEŞTİRME

çıkışında elde edilen sol ve sağ öznitelik vektörlerine ilinti işlemi uygulanarak, verilen giriş için arama uzayı boyunca eşleşme maliyetleri elde edilmiş olur. Content-CNN ağ yapısı ise Şekil 2’de verilmiştir.

Elde edilen eşleşme maliyetleri üzerinden Eşitlik 2’de verilen kazanan-hepsini-alır yöntemi uygulanarak ağıma giriş olarak vereceğimiz derinlik haritası elde edilmektedir.

$$d_p = \operatorname{argmin} C_{\text{Content_CNN}}(p, d) \quad d \in [0, d_{\max}] \quad (2)$$

Eşitlik 2’de yer alan $C_{\text{Content_CNN}}$, Content-CNN ağından sol görüntüdeki her p pikseli için 0 ile maksimum derinlik (d_{\max}) aralığı için elde edilen eşleşme maliyetlerini, d ise derinlik değerini temsil etmektedir.

Önerilen ağ yapısına giriş olarak uygulanan sol görüntü üzerinde sıralı evrişim işlemleri uygulanmıştır. Her bir katmanın sonunda yığın normalizasyonu uygulanarak ağı daha kolay ve iyi eğitilebilmesi amaçlanmıştır. Ayrıca son katman hariç her bir evrişim katmanında aktivasyon fonksiyonu olarak ReLU kullanılmıştır. evrişim işlemlerinde filtre boyutu 5×5 , öznitelik haritası boyutu ise 64 seçilmiştir. Son katmanda öznitelik haritasının boyutu bir olarak belirlenmiştir ve katmanın sonunda düzeltme haritası elde edilmiştir. En son adımda ise sıralı evrişim katmanları sonucundan elde edilen düzeltme haritası ve başlangıç derinlik haritası birleştirilerek nihai derinlik haritası elde edilmiştir. Sonuç olarak, önerilen ağı mimarisi uygulanabilirlik açısından oldukça basit ve sığdır (shallow).

Bu çalışmada, CNN katman sayısının başarımlar üzerindeki etkisi farklı boyutlarda derinlik değeri kullanılarak incelenmiştir. Bu amaç doğrultusunda üç farklı CNN katman sayısı için ağ eğitimi gerçekleştirilmiştir. CNN katmanların sayısı üç, dört ve beş olarak değiştirilerek en uygun ağ yapısı belirlenmiştir. Parametreler belirlenirken, CNN Katman sayısı ve öznitelik harita boyutları küçük tutularak gelecekteki çalışmalarda yapının uçtan-uca (end-to-end) bir ağ yapısında kullanılması ve gerçek zamanlı çalışabilmesi durumunda daha uygulanabilir olmasını sağlamaktır.

4. DENEYSEL SONUÇLAR

Bu çalışma kapsamında önerilen ağ yapısı PyTorch çerçevesi (framework) kullanılarak geliştirilmiştir. Eğitim ve test işlemleri için Nvidia 2080Ti ekran kartı kullanılmıştır. Eğitim ve test işlemleri için KITTI 2012 [18] ve KITTI 2015 [19] veri kümeleri kullanılmıştır. KITTI 2012 ve KITTI 2015 veri kümeleri otonom sürüş uygulamalarında kullanılabilecek düzeltilmiş dış mekân görüntülerinden oluşmaktadır. Veri kümelerinde sırayla 194 ve 200 adet düzeltilmiş görüntü bulunmaktadır. Veri kümelerinde yer alan her bir stereo görüntü için lazer tarayıcılarla toplanmış doğru kabul edilen derinlik bilgileri verilmiştir. Çalışmada, veri setindeki görüntülerin %10’u doğrulama, %10’u test için ayrılmıştır. Geriye kalan %80’i ise ağların eğitiminde kullanılmıştır.

Eğitim aşamasında ağa giriş olarak sol görüntü ve derinlik haritası olmak üzere iki görüntü verilmektedir. Sol görüntü veri kümesinden derinlik değeri düzeltilen görüntüyü temsil etmektedir. Diğer giriş olan derinlik haritası ise Content-CNN ağından elde edilen düzelecek görüntüyü temsil etmektedir. Eğitim parametrelerinin başlangıç değerlerinin belirlenmesinde Xavier yöntemi kullanılmıştır. Yitim fonksiyonu (loss function) olarak ise Eşitlik 3’de verilen $L1$ fonksiyonu kullanılmıştır. Yitim fonksiyonunu en aza indirmek için ADAM yöntemi kullanılmıştır. Eğitim parametreleri sırayla; öğrenme oranı 0.001, beta değerleri 0.9, 0.999 olarak seçilmiştir. Ağı eğitimi için 100 çevrim gerçekleştirilmiştir.

$$\text{hata} = \frac{1}{N} \sum_{n=1}^N (|Y_{\text{gerçek}} - Y_{\text{tahmin}}|) \quad (3)$$

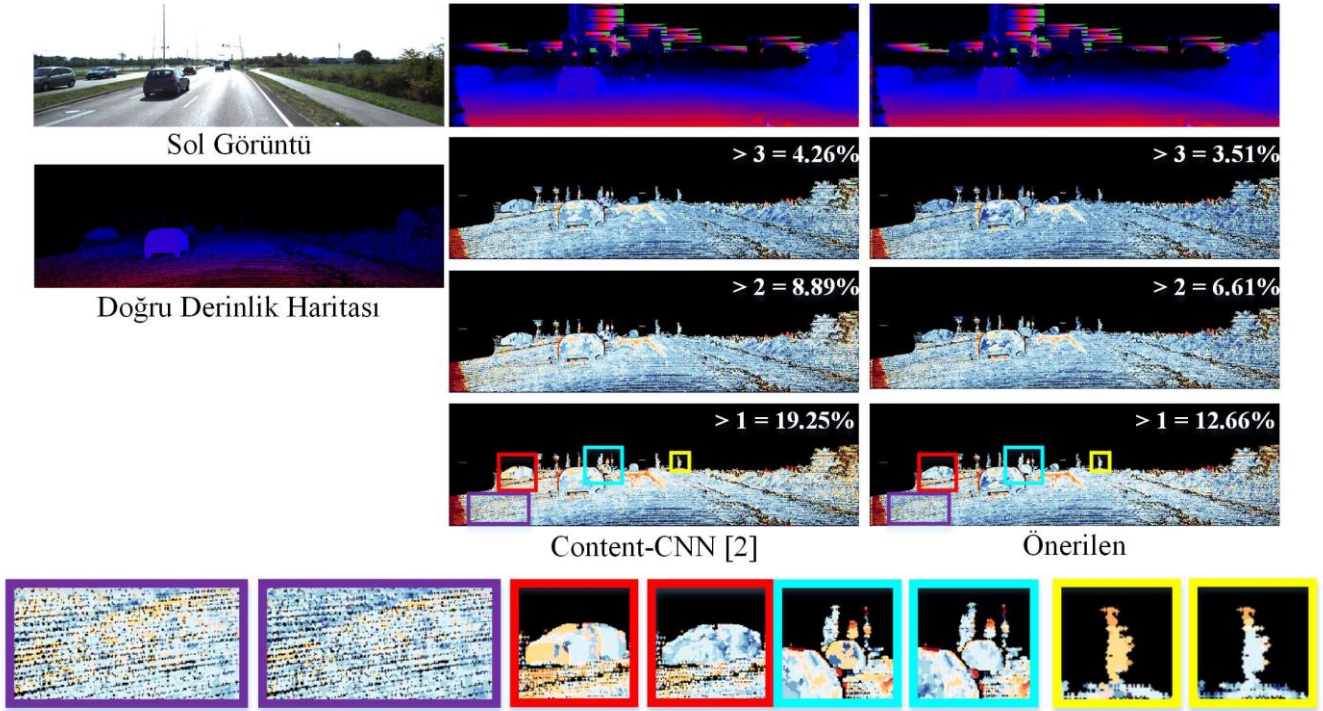
Önerilen ağ yapısında, CNN katmanlarının derinliğine göre başarımların etkisini test etmek için farklı sayıda CNN katmanı eğitilerek test edilmiştir. CNN katman sayıları değiştirilerek elde edilen 3 piksel hata oranları Tablo 1’de verilmiştir.

Tablo 1’de yer alan ifadelerden Content-CNN [2] kullandığımız referans derinlik haritası elde etme yöntemini temsil etmektedir. Önerilen olarak verilen yöntemler çalışma kapsamında oluşturulan ağ yapılarını ifade etmektedir. Parantez içerisindeki rakamlar ise CNN katman sayısını göstermektedir. Veri kümeleri ayrı ayrı (KITTI 2012 ve KITTI 2015) ve birleştirilerek (KITTI 2012 + KITTI 2015) kullanılmıştır. Yöntemler tarafından üretilen derinlik haritaları referans derinlik haritaları ile karşılaştırılmış ve üç piksele kadar olan farklar doğru eşleştirme olarak kabul edilmiştir. Tabloda verilen değerler doğru eşleştirmelerin toplam piksel sayısına oranını belirtmektedir. Tablo 1’den de kolaylıkla görüldüğü gibi nitel değerlendirmeler sonucunda her bir CNN katmanı sayısında referans derinlik haritasının iyileştirildiği gösterilmektedir. CNN katman sayısındaki artış genellikle elde edilen derinlik haritasındaki hatanın düşürülmesine katkı sağlamaktadır. Önerilen yapı özellikle KITTI 2015 veri kümesinde Content-CNN tarafından elde edilen sonuçları iyileştirmede başarılı olmuştur. Ayrıca Tablo 1’de çalışma süresi analizleri de verilmiştir. Önerilen ağ yapılarının çalışma sürelerinin Content-CNN ile kıyaslandığı zaman oldukça düşük seviyelerde kaldığı gözlemlenmiştir. Buradan da görülebileceği gibi önerilen ağı sığ yapısı sayesinde diğer yöntemler ile birlikte kullanılması da mümkün olacaktır.

Tablo 1. Önerilen ağ yapısının farklı CNN katman sayıları için 3 piksel hata oranları

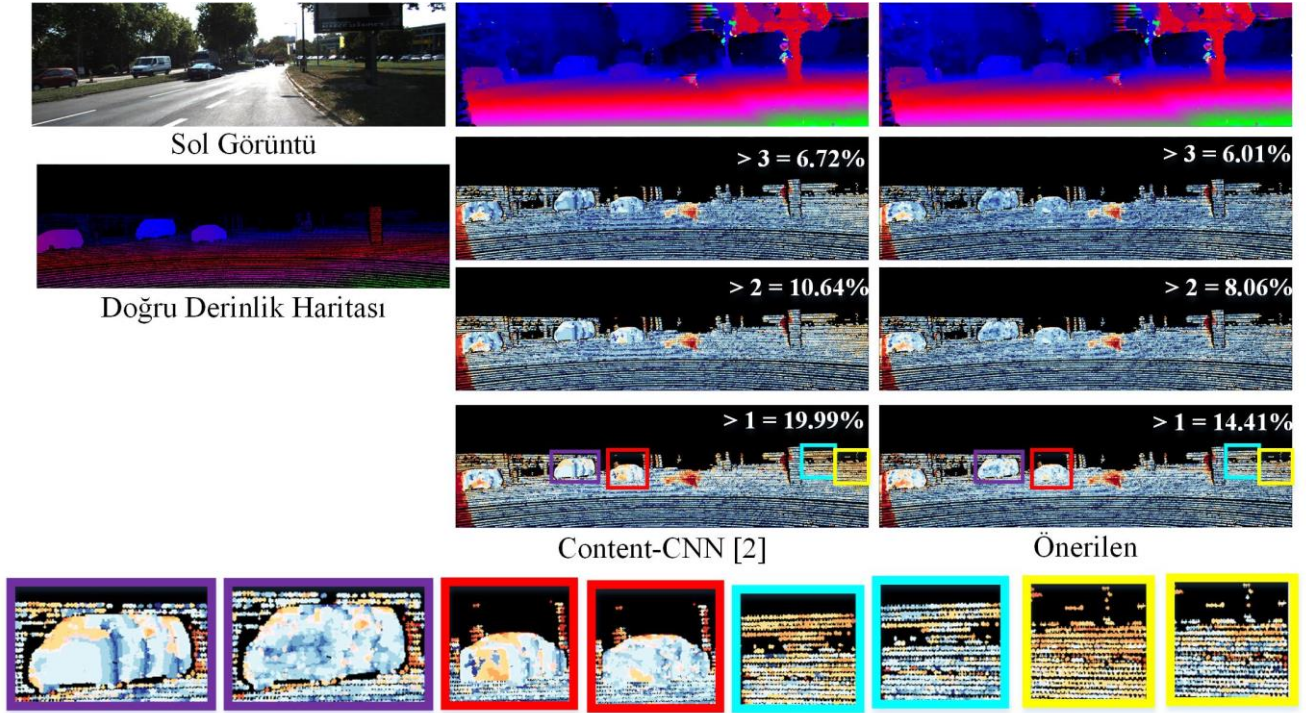
Yöntem	Veri Kümelerinde Oluşan Hatalar (%)			Çalışma Süresi (s)
	KITTI 2012	KITTI 2015	KITTI 2012 + KITTI 2015	
Content-CNN [2]	8,93	6,23	7,73	0.14
Önerilen (3)	8,88	5.94	7,57	1.39×10^{-3}
Önerilen (4)	8,83	5,89	7,54	2.08×10^{-3}
Önerilen (5)	8,83	5.87	7,52	2.73×10^{-3}

Nitel başarımın yanında, görüntü üzerinden yapılan nicel karşılaştırmalar da uygulama esnasında karşılaşılabilecek durumları analiz edebilmek için önem arz etmektedir. Şekil 3 ve Şekil 4'te, önerilen ve referans yöntemden elde edilen derinlik haritaları ile 1, 2 ve 3 piksel hataları incelenmiştir. Görsellerde ilk sütunda sol giriş görüntüsü ve doğru derinlik haritası verilmiştir. İkinci ve üçüncü sütünün ilk satırlarında ise Content-CNN ağı ile ve Önerilen (4) ile elde edilmiş derinlik haritaları verilmiştir. İkinci, üçüncü ve dördüncü satırlarda verilen görseller ise bir, iki ve üç piksel hata haritalarını temsil etmektedir. Örnek olarak bir piksel için elde edilen derinlik değerinin doğru kabul edilebilmesi için, referans derinlik değeri ile yöntemlerden elde edilen değer arasındaki farkın bire eşit veya az olması gerekmektedir. Şekil 3'te piksel hataları görselleştirilirken kullanılan renkler hatanın büyüklüğünü ifade etmektedir. Soğuk renkler hatanın az olduğunu (Koyu mavi en az hata), sıcak renkler (koyu kırmızı en çok hata) ise hatanın çok olduğunu ifade etmektedir.



Şekil 3. Yöntemlerden elde edilen derinlik haritaları (ikinci ve üçüncü kolonda yer alan ilk satır), ve piksel hataları (ikinci ve üçüncü kolonda yer alan iki, üç ve dördüncü satırlar). En alt satırda piksel hatalarının büyütülmüş görüntüleri (sol taraftaki görüntüler Content-CNN [2] ve sağ taraftaki görüntüler ise önerilen yönteme ait piksel hatalarıdır).

EVRIŞİMLİ SİNİR AĞI TABANLI DERİNLİK HARİTASI İYİLEŞTİRME



Şekil 4. Yöntemlerden elde edilen derinlik haritaları (ikinci ve üçüncü kolonda yer alan ilk satır), ve piksel hataları (ikinci ve üçüncü kolonda yer alan iki, üç ve dördüncü satırlar). En altta satırda piksel hatalarının büyütülmüş görüntüleri (sol taraftaki görüntüler Content-CNN [2] ve sağ taraftaki görüntüler ise önerilen yönteme ait piksel hatalarıdır.)

Şekil 3 ve Şekil 4'te, önerilen yöntem ile elde edilen derinlik haritalarında yer alan 3-piksel ve 2-piksel hataları incelendiğinde, derinlik haritalarında çok belirgin bir biçimde iyileştirme sağlandığı görülmektedir. En altta verilen 1-piksel hataları da benzer şekilde nesne köşeleri ve desen içermeyen bölgelerde iyileştirmeler elde edildiğini net bir şekilde göstermektedir. Şekil 3 ve Şekil 4'te görüldüğü üzere önerilen ağ yapısı ağaç, araba gibi nesnelerin hem kenar hem de yüzey alanlarında gözle görülür bir düzeltme sağlamaktadır. Bu düzeltmeler özellikle 3B bölütleme ve 3B nesne tanımlama gibi süreklilik ve kenar bilgilerinin hassasiyetine ihtiyaç duyan uygulamalar için büyük bir avantaj sağlayacaktır.

5. SONUÇLAR

Bu çalışmada, derinlik haritalarının iyileştirilmesi amacıyla sadece sol kamera görüntüsü ve derinlik haritasının giriş olarak kullanıldığı bir ağ yapısı önerilmiştir. Önerilen ağ yapısında, sol görüntüden elde edilen öznelikler derinlik haritasının iyileştirilmesinde kullanılmaktadır. Nitel ve nicel sonuçlar incelendiğinde önerilen ağ ile çok ağır bir işlem yükü eklenmeden literatürde yer alan bir yöntemin derinlik haritasının iyileştirildiği gösterilmiştir. İşlem yükünün az olmasının temel sebebi önerilen ağ yapısının sığ (shallow) ve kolay uygulanabilir yapıda olmasıdır. Bu sayede, literatürde derinlik haritasının elde edilebildiği yöntemlere yardımcı olarak kullanılabilir. Ayrıca, mevcut stereo görüntü tabanlı derinlik bilgisini kullanan sistemler içerisinde performansı arttırmak için doğrudan kullanılabilir bir yöntem önerilmiştir. Böylece, stereo görüntülerin kullanıldığı ve derinlik bilgisine ihtiyaç duyulan uygulamaların da başarımı artırılmış olacaktır.

TEŞEKKÜR

Bu çalışma, Niğde Ömer Halisdemir Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi "MMT 2019/7-BAGEP" nolu proje ile desteklenmektedir.

KAYNAKLAR

- [1] J. Žbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1592-1599.
- [2] W. Luo, A. G. Schwing, and R. Urtasun, "Efficient deep learning for stereo matching", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5695-5703.
- [3] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *International Journal of Computer Vision*, vol 47, pp. 7-42, 2002.
- [4] A. Emlek, M. Peker and K. F. Dilaver, "Variable window size for stereo image matching based on edge information," In Proc. International Artificial Intelligence and Data Processing Symposium, 2017, pp. 1-4.
- [5] A. Emlek, M. Peker and M. K. Yalçın, "Improving the cost-volume based local stereo matching algorithm," In Proc. Signal Processing and Communications Applications Conference, 2018, pp. 1-4.
- [6] J. Sun, N. N. Zheng and H. Y. Shum, "Stereo matching using belief propagation", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 787-800, 2003.
- [7] L. Hong and G. Chen, "Segment-based stereo matching using graph cuts", In Proc. Conference on Computer Vision and Pattern Recognition, 2004, pp. 74-81.
- [8] P. Brandao, E. Mazomenos and D. Stoyanov, "Widening siamese architectures for stereo matching", *Pattern Recognition Letters*, vol. 120, pp. 75-81, 2018.
- [9] M. Poggi and S. Mattoccia, "Learning from scratch a confidence measure", In Proc. British Machine Vision Conference, 2016, pp. 46.1-46.13.
- [10] M. Poggi and S. Mattoccia, "Learning to predict stereo reliability enforcing local consistency of confidence maps", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4541-4550.
- [11] A. Seki and M. Pollefeys, "Patch based confidence prediction for dense disparity map", In Proc. British Machine Vision Conference, 2016, pp. 23.1-23.13.
- [12] F. Cheng, X. He, and H. Zhang, "Learning to refine depth for robust stereo estimation", *Pattern Recognition*, vol. 74, pp. 122-133, 2018.
- [13] N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovitskiy and T. Brox, "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation", In Proc. Computer Vision and Pattern Recognition, 2016, pp. 4040-4048.
- [14] S. Kim, D. Min, S. Kim and K. Sohn, "Unified confidence estimation networks for robust stereo matching", *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1299-1313, 2019.
- [15] S. Kim, S. Kim, D. Min and K. Sohn, "LAF-Net: Locally Adaptive Fusion Networks for stereo confidence estimation", In Proc. Computer Vision and Pattern Recognition, 2019, pp. 205-214.
- [16] J. Pang, W. Sun, J. S. Ren, C. Yang and Q. Yan, "Cascade residual learning: a two-stage convolutional neural network for stereo matching", In Proc. International Conference on Computer Vision Workshops, 2017, pp. 878-886.
- [17] J. Chang and Y. Chen, "Pyramid stereo matching network", In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5410-5418.
- [18] A. Geiger, P. Lenz and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3354-3361.
- [19] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3061-3070.

