

Bağlamsal Çıkarımla Nesne Sezimi

Object Detection with Contextual Inference

Fırat Kalaycılar, Selim Aksoy

Bilgisayar Mühendisliği Bölümü, Bilkent Üniversitesi, Bilkent, 06800, Ankara

{firatk, saksoy}@cs.bilkent.edu.tr

Özetçe

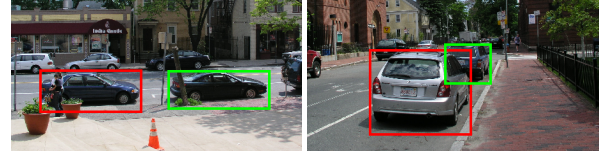
Bu bildiride, sezim başarımını arttırmada tek tek sezilmiş nesneler arasındaki bağlamsal ilişkilerden yararlanan bir nesne sezim sistemi tanıtılmaktadır. Bu çalışmadaki ilk katkı, iki boyutlu görüntü uzayında yapılan ölçümlerden olasılıksal çıkarım yaparak nesneler arası gerçek dünya ilişkilerinin (çevresinde, yakınında, üzerinde vb.) modellenmesidir. Diğer bir katkı ise, bireysel nesne etiketlerine ve nesne ikilileri arasındaki ilişkilere bağlı olan sahne olasılık fonksiyonunun enbüyütülerek, nesnelerin en son etiketlerinin atanmasıdır. En tutarlı sahne düzenlenişini bulmak için bu enbüyütme problemi, doğrusal eniyileme kullanılarak çözülmüştür. Ofis görüntüleri içeren iki farklı veri kümesinde yapılan deneylerde, gerçek dünya uzamsal ilişkileri bağlamsal bilgi olarak kullanıldığında genel sezim başarımının arttığı gözlemlenmiştir.

Abstract

In this paper, an object detection system that utilizes contextual relationships between individually detected objects to improve the overall detection performance is introduced. The first contribution in this work is the modelling of real world object relationships (beside, on, near, etc.) that can be probabilistically inferred using measurements in the 2D image space. The other contribution is the assignment of final labels to the detected objects by maximizing a scene probability function that is defined jointly using both individual object labels and their pairwise spatial relationships. The most consistent scene configuration is obtained by solving the maximization problem using linear optimization. Experiments on two different office data sets showed that incorporation of the real world spatial relationships as contextual information improved the overall detection performance.

1. Giriş

Nesne tanıma araştırmaları yıllar içinde göstermiştir ki sahnedeki nesneleri tek tek bulmaya yarayan yöntemler sosyal çözümler sunamamaktadır. Günümüzde, bir nesnenin tanınmasının sahnede bulunan diğer nesnelere bağımlı olduğu kabul edilmektedir. Oliva ve Torralba, bu bağımlılığın neden yararlı olduğunu özetlemektedir [1]. Örneğin, bir bilgisayar monitörü sezicisi bir takım dikdörtgen yapılar bulmaya çalışır. Ancak, pencereler, fotoğraf çerçeveleri vb. monitöre bu anlamda benzemektedir. Bu yüzden, monitör sezicinin hatayla pencere



Şekil 1: Üç boyutta benzer olan ilişkiler, iki boyutta farklı düzenlenişlere denk gelebilmektedir.

bulması olağandır. Burada monitörle çevrede bulunan diğer nesneler (klavye, fare gibi) arasındaki bağlamsal ilişkiler hesaba katılırsa, bunun gibi hataları önlemek mümkündür.

Torralba bağlam ve nesneler arasındaki ilişkiyi, tüm sahneden ve bireysel nesnelerden çıkarılan alt düzey öznitelikler arasındaki ilintiyi kullanarak modellemiştir [2]. Bir diğer çalışmada, Torralba ve diğerleri [3], genel özniteliklerin sahne tahmininde kullandığı bağlam bazlı bir görme sistemini tanıtmıştır. Tahmin edilen sahneden sağlanan bağlam bilgisi yerel seziciler tarafından önsel bilgi olarak kullanılır.

Alt düzey öznitelikler arasındaki ilintiyi kullanan yaklaşımların yanında, daha sezgisel yöntemler de bulunmaktadır. [4] numaralı çalışmada, sahnedeki bağlam, nesnelerin eşdizimlilik eğilimleri cinsinden modellenir. Görüntü öncelikle kararlı bölgelere bölünür ve her bölge için olasılıklarına göre sınıf etiketleri sıralanır. Kullanılan bağlamsal model sayesinde, etiket belirsizliği giderilir ve en uygun etiket bölgelere atanır. Eşdizimliliğe ek olarak, [5]'te bağıl konum bilgisi (uzamsal ilişkiler) de bağlamsal bilgi olarak kullanılmıştır. Parikh ve diğerleri de [6], bağlamsal etkileşim olarak eşdizimlilik, konum ve bağıl ölçekten yararlanmaktadır. Bu çalışmaların ortak bir özelliği ise bağlamsal bilgiyi koşullu rasgele alan çerçevesi kullanarak sistemlerine katmalarıdır.

Uydu görüntüsü analizinde, nesneler ve bölgeler arası uzamsal ilişkilerin (yukarı, aşağı, sağ, sol, çevreleme vb.) karar sürecine dahil edilmesinin nesne tanıma başarımını iyileştirdiği görülmüştür [7]. Ancak yeryüzünün tepeden çekilmiş iki boyutlu görüntülerinde kullanılan bu modellerin, herhangi bir bakış açısından çekilmiş doğal sahneler için geçerli olmasını beklemek pek doğru değildir. Bu tür uzamsal ilişkiler perspektif izdüşümdeki farklılıklara değişimsiz olmadığından, üç boyutlu gerçek dünyada tanımlı olan uzamsal ilişkiler, görüntü uzayında çok değişik şekillerde yorumlanabilir. Şekil 1 bu duruma bir örnektir. İki görüntüde de, kırmızıyla işaretlenen araba yeşille işaretlenen arabanın arkasına park etmiş durumdadır. Ancak iki boyutta algılanan ilişkilere bakıldığında, arabalar birbirlerine göre farklı konumdaymış gibi görünmektedir.

Bu bildiride, bir bağlamsal nesne sezim yöntemi anlatılmaktadır. Buradaki ana amaç, bağlamdan bağımsız bir şekilde sezilen nesnelerden yararlanarak en iyi sahne düzenlenişini elde etmektir. İki boyutta geçerli olan uzamsal ilişkiler yerine, gerçek dünya ilişkileri üzerine olasılıksal çıkarım yapılmaktadır. Daha sonra bu çıkarımlara ve nesne sınıf etiketlerine bağımlı olan sahne olasılık fonksiyonu enbüyütülerek en iyi sahne düzenlenişine ulaşılmaktadır.

2. Nesne Sezimi

Bağlamsal sahne modelini oluşturmaın ilk adımı sahne elemanlarının tek tek sezilmesidir. Bunun için yöntemimizde herhangi bir nesne sezici kullanılabilir. Bu çalışmada birçok zayıf sezicinin ağırlıklı katışımlı olan boosting bazlı seziciler kullanılmıştır [8]. Her zayıf sezici öznitelik tepkilerini değerlendirir ve nesnenin merkezi için oy verir. Sonuç olarak, bu oyların ağırlıklı katışımlı hesaplanır ve aday nesneleri simgeleyen sınırlayıcı dikkörtgenler skorlarıyla bildirilir. Gauss piramitleri kullanılarak bu sezicilerin çoklu ölçekte çalışan sürümleri elde edilmiştir. Her bir nesne sınıfı için farklı bir sezici eğitilmiştir. Sezicilerin bildirdiği skorları olasılık değerlerine dönüştürmek için her skor söz konusu sınıf için gözlenen en yüksek değer kullanılarak normalize edilmiştir.

Bir görüntü verildiğinde, seziciler birbirlerinden bağımsız bir şekilde çalıştırılır ve her sezici sınırlayıcı dikkörtgen kümesi döndürür. $P(X_i|D_i)$, i nesnesinin sezim olasılığını ifade etmektedir. D_i o nesneyi bulan seziciyi gösterirken, X_i nesnenin sınıf etiketini göstermektedir. Her bir sezici sadece tek türde (örneğin c) nesneleri bulabildiğinden ya da bulamadığından, X_i yalnızca iki olası değer alabilir: c ya da *bilinmeyen*. Bu değerlere sahip olma olasılıkları sırasıyla $P(X_i = c|D_i = c)$ ve $P(X_i = \text{bilinmeyen}|D_i = c)$ 'dir. Bu iki olasılık değeri arasındaki ilişki

$$P(X_i = \text{bilinmeyen}|D_i = c) = 1 - P(X_i = c|D_i = c) \quad (1)$$

şeklinde dir.

3. Uzamsal İlişki Tanımlayıcısı

Nesne i 'nin, referans nesnesi ref 'e göre uzamsal ilişki tanımlayıcısı $S_{i,ref}$ 'dir. Bu nesnelerin sınırlayıcı dikkörtgenleri

$$sd_i = [x_0 \ y_0 \ x_1 \ y_1 \ 1]^T \quad (2)$$

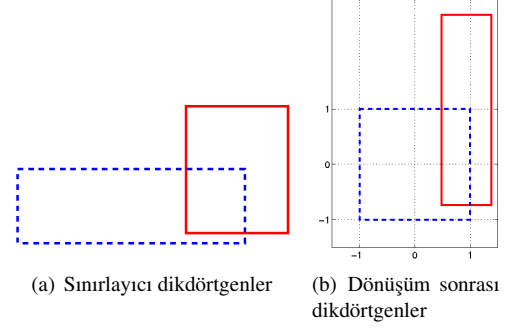
$$sd_{ref} = [x_0^{ref} \ y_0^{ref} \ x_1^{ref} \ y_1^{ref} \ 1]^T \quad (3)$$

şeklinde gösterilmektedir. Burada (x_0, y_0) , (x_0^{ref}, y_0^{ref}) sol alt ve (x_1, y_1) , (x_1^{ref}, y_1^{ref}) de sağ üst köşe koordinatlarına karşılık gelmektedir. Dönüşüm matrisi Γ

$$t_x = -\frac{x_1^{ref} + x_0^{ref}}{2} \quad t_y = -\frac{y_1^{ref} + y_0^{ref}}{2} \quad (4)$$

$$s_x = \frac{2}{x_1^{ref} - x_0^{ref}} \quad s_y = \frac{2}{y_1^{ref} - y_0^{ref}} \quad (5)$$

$$\Gamma = \begin{bmatrix} s_x & 0 & 0 & 0 & s_x t_x \\ 0 & s_y & 0 & 0 & s_y t_y \\ 0 & 0 & s_x & 0 & s_x t_x \\ 0 & 0 & 0 & s_y & s_y t_y \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6)$$



Şekil 2: Uzamsal ilişki tanımlayıcısının görsel gösterimi. Referans dikkörtgeni kesikli mavi çizgilerle çizilmiştir.

şeklinde oluşturulur. Burada s_x , s_y ölçekleme katsayıları iken, t_x , t_y öteleme miktarlarıdır. Bu dönüşüm, sd_{ref} 'i Şekil 2'de görüldüğü gibi normalize eder. Sonuç olarak dönüşüme uğrayan dikkörtgenler

$$sd'_{ref} = \Gamma sd_{ref} = [-1 \ -1 \ 1 \ 1 \ 1]^T \quad (7)$$

$$sd'_i = \Gamma sd_i = [x'_0 \ y'_0 \ x'_1 \ y'_1 \ 1]^T \quad (8)$$

şeklinde bulunur. Bütün bunlar bilindiğinde uzamsal ilişki tanımlayıcısı

$$S_{i,ref} = [x'_0 \ x'_1 \ y'_0 \ y'_1] \quad (9)$$

şeklinde ifade edilmektedir. Anlaşıldığı gibi bu tanımlayıcıda *bağlı ölçek* ve *bağlı konum* bilgisi birleşik bir gösterimde kodlanmıştır.

4. Gerçek Dünya İlişkileri Üzerine Çıkarımlar

Bir sahneyi anlamının yolu onun alt bileşenlerini anlamaktan geçmektedir. Birinci dereceden alt bileşenler sahnedeki nesnelerdir. Günümüzde, iki boyutlu görüntülerden çıkarılan öznitelikler (renk, doku gibi) üzerinden üç boyutlu nesneleri olasılıksal olarak tanımanın mümkün olduğu bilinmektedir. Bu çalışmada, bu yöntemin bir benzeri, bir başka alt bileşen olan, nesne ikilileri arasındaki üç boyutlu tanımlı ilişkilere uygulanmaktadır. Bu durumda amaç, görüntüden çıkarılan iki boyutlu uzamsal ilişki özniteliklerini (denklem (9)'te tanımlanan $S_{i,j}$) kullanarak gerçek dünyada geçerli olan ilişkileri olasılıksal olarak anlamaktır. Bunu yapabilmek için $P(R_{i,j} = r|X_i = u, X_j = v, S_{i,j})$ olasılıklarına ihtiyaç duyulmaktadır. Bu, u cinsinde bir nesnenin v cinsinde bir nesneye göre $S_{i,j}$ tanımlayıcısına sahip olduğunda, aralarındaki gerçek dünya ilişkisinin r olma olasılığıdır. Bu olasılıkları öğrenmek için elle etiketlenmiş bir öğretim kümesi oluşturulmuştur. Bu kümede gözlenen her $(X_i = u, X_j = v, R_{i,j} = r)$ üçlüsü için $P(R_{i,j} = r|X_i = u, X_j = v, S_{i,j})$ olasılık modeli öğrenilmiştir.

5. Bağlamsal Uzlaşmanın Enbüyüklenmesi

Tek tek sezilmiş nesneler ve aralarındaki uzamsal ilişkiler verildiğinde, sistemimiz bağlamsal uzlaşmayı enbüyükleyerek, her nesne için en uygun olan etiketi belirlemektedir. Bu, sahne olasılık fonksiyonunun enbüyüklenmesiyle gerçekleşmektedir.

5.1. Sahne Olasılık Fonksiyonu

Yöntemimizde, üç boyutlu modelin tamamı yerine, alt bileşenleri (nesneler ve ikili ilişkiler) kullanılmaktadır. Buna göre, içinde n adet nesne bulunan bir sahnenin olasılık fonksiyonu **SOF** şu şekilde tanımlanmıştır:

$$\text{SOF} = P(\mathbf{X}, \mathbf{R} | \mathbf{D}, \mathbf{S}), \quad (10)$$

$$\mathbf{X} = \{X_1, X_2, \dots, X_n\}, \quad (11)$$

$$\mathbf{R} = \{R_{1,2}, R_{1,3}, \dots, R_{n-1,n}\}, \quad (12)$$

$$\mathbf{D} = \{D_1, D_2, \dots, D_n\}, \quad (13)$$

$$\mathbf{S} = \{S_{1,2}, S_{1,3}, \dots, S_{n-1,n}\}. \quad (14)$$

Burada $\mathbf{X}$ sınıf etiketlerini ve $\mathbf{R}$ nesne ikilileri arasındaki gerçek dünya ilişki etiketlerini gösteren rastgele ayrık değişken kümeleridir. $\mathbf{D}$ nesneleri bulan sezicilerin kümesi iken, $\mathbf{S}$ nesne ikililerinin birbirlerine göre uzamsal ilişki tanımlayıcılarını içeren kümedir.

(10) numaralı denklem

$$\text{SOF} = P(\mathbf{X} | \mathbf{D}, \mathbf{S}) P(\mathbf{R} | \mathbf{X}, \mathbf{D}, \mathbf{S}) \quad (15)$$

şeklinde yazılabilir. Burada iki varsayım yapılmaktadır:

Varsayım 1: Her nesne ayrı ayrı sezildiği için, bir nesnenin sınıf etiketi sadece onu tanıyan sezicinin türüne bağlıdır.

Varsayım 2: Bir nesne ikilisi arasındaki gerçek dünya ilişkisi sadece o nesnelerin sınıf etiketlerine ve aralarındaki uzamsal ilişki tanımlayıcısına bağlıdır.

Bu varsayımlar doğrultusunda (15) numaralı denklem

$$\begin{aligned} \text{SOF} &= P(\mathbf{X} | \mathbf{D}) P(\mathbf{R} | \mathbf{X}, \mathbf{S}) \\ &= \left(\prod_{i=1}^n P(X_i | D_i) \right) \left(\prod_{i=1}^{n-1} \prod_{j=i+1}^n P(R_{i,j} | X_i, X_j, S_{i,j}) \right) \end{aligned} \quad (16)$$

halini almaktadır.

5.2. Doğrusal Eniyileme ile Çözüm

SOF'u enbüyükleyen etiketlerle **SOF**'un logaritmasını enbüyükleyenler aynıdır. Dolayısıyla, bu enbüyükleme problemi 0-1 programlaması olarak modellenebilmektedir. Bu programın amaç fonksiyonu

$$\begin{aligned} f &= \left(\sum_{i=1}^n \sum_{u \in \mathcal{C}_i} \alpha_{iu} \log P(X_i = u | D_i) \right) \\ &+ \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\substack{u \in \mathcal{C}_i \\ v \in \mathcal{C}_j \\ r \in \mathcal{R}_{u,v}}} \beta_{ijuvr} \log P(R_{i,j} = r | X_i = u, X_j = v, S_{i,j}) \right) \end{aligned} \quad (17)$$

şeklinde. Burada $\mathcal{C}_i$, X_i 'nin olası değerlerinin kümesi iken, $\mathcal{R}_{u,v}$ de u türünde bir nesneyle v türünde bir nesne arasında gözlemlenebilen gerçek dünya ilişkilerinin kümesidir. α_{iu} , i nesnesinin u sınıf etiketine sahip olup olmadığını gösteren değişkendir. β_{ijuvr} de u türündeki i nesnesi ile v türündeki j

nesnesi arasındaki gerçek dünya ilişkisinin r olup olmadığını gösteren değişkendir. Öyleyse problem şu şekilde yazılabilir:

$$\begin{aligned} \max \quad & f \\ \text{s.t.} \quad & \sum_{u \in \mathcal{C}_i} \alpha_{iu} = 1, \quad \sum_{\substack{u \in \mathcal{C}_i \\ v \in \mathcal{C}_j \\ r \in \mathcal{R}_{u,v}}} \beta_{ijuvr} = 1, \\ & \beta_{ijuvr} \leq \alpha_{iu}, \quad \beta_{ijuvr} \leq \alpha_{jv}, \\ & \alpha_{iu} \in \{0, 1\}, \quad \beta_{ijuvr} \in \{0, 1\}. \end{aligned} \quad (18)$$

Bu problemi çözmek NP-zor olduğu için programın gevşeme uygulanmış sürümü kullanılmaktadır. Bunun için $\alpha_{iu} \in \{0, 1\}$ kısıtı, $0 \leq \alpha_{iu} \leq 1$ kısıtıyla ve $\beta_{ijuvr} \in \{0, 1\}$ kısıtı, $0 \leq \beta_{ijuvr} \leq 1$ kısıtıyla değiştirilmiştir. Yeni problem bir doğrusal programlama olduğundan çözüm çokterimli zamanda bulunabilmektedir. Çoğu zaman çözümlerde α_{iu} ve β_{ijuvr} değerlerinin hepsi 0 veya 1'e eşit olmaktadır. Bu durumda en iyi çözüm elde edilmiştir. Ancak bazı amaç fonksiyonları için çözüm ara değerler (0 ile 1 arası) içerebilmektedir. Bu durumda o değerler 0 ve 1'den en yakın olduklarına yuvarlanmaktadır. Sonuç olarak α_{iu} ve β_{ijuvr} 'nın son değerleri nesne etiketlerini ve nesneler arasındaki gerçek dünya ilişkilerini belirlemektedir.

5.3. Model Parametreleri

Nesne sınıflarından *bilinmeyen* sistemimiz tarafından tanınmayan nesnelere denk gelmektedir. Dolayısıyla *bilinmeyen* bir nesnenin içinde geçtiği bir gerçek dünya ilişkisi gözlemek mümkün değildir. $u = \text{bilinmeyen}$ ve/veya $v = \text{bilinmeyen}$ olduğunda $\mathcal{R}_{u,v} = \{\text{tanımsız}\}$ 'dir. O yüzden amaç fonksiyonunun içinde geçen şu olasılık değerlerini tanımlamak gerekmektedir:

- $P(R_{i,j} = \text{tanımsız} | X_i = \text{bilinmeyen}, X_j = v, S_{i,j})$,
- $P(R_{i,j} = \text{tanımsız} | X_i = u, X_j = \text{bilinmeyen}, S_{i,j})$,
- $P(R_{i,j} = \text{tanımsız} | X_i = \text{bilinmeyen}, X_j = \text{bilinmeyen}, S_{i,j})$.

Sistemimiz bu olasılıkları öğrenemeyeceğinden, yerlerine kullanılacak değeri model parametresi olarak almaktadır.

6. Deneyler ve Sonuçlar

Deneyler ofis görüntüleri içeren iki farklı veri kümesinde gerçekleştirilmiştir. İlki 62 kapalı mekan görüntüsü içeren Bilkent Ofis Veri Kümesi'dir. Diğer ise içinde 124 ofis görüntüsü olan LabelMe Veri Küme'sinin bir alt kümesidir. Her iki veri kümesinde bulunan görüntülerde bolca rastlanan *bilgisayar monitörleri*, *klavyeler*, *fareler* ve *kupalar* deneylerde kullanılacak nesne sınıfları olarak seçilmiştir.

Nesne seziciler, LabelMe'nin başka bir alt kümesi kullanılarak eğitilmiştir. Burada 2. bölümde anlatılan seziciler kullanılmıştır. Unutulmamalıdır ki sınırlayıcı diktörtgen ve olasılık bildirebilen herhangi bir sezici sistemimizde kullanılabilir.

4. bölümde açıklanan gerçek dünya ilişki olasılıkları, elle etiketlenen nesneler ve ilişkiler kullanılarak tek sınıflı Gauss dağılımı olarak modellenmiştir. Kullanılan iki veri kümesinde de nesne sınıfı ikilileri için birden fazla baskın gerçek dünya ilişkisi gözlenmemiştir. Dolayısıyla her bir ikili için tek bir

ilişki modeli öğrenilmiştir. Bunlar monitör-klavye, monitör-fare, klavye-fare için *yakınında* ve monitör-monitör, klavye-klavye, fare-fare, fare-kupa, monitör-kupa, klavye-kupa, kupa-kupa için *çevresinde* ilişkileridir. *Çevresinde* ilişkisi *yakınında* ilişkisinin yerine uzaklığa daha az önem verilen durumlarda kullanılmaktadır. Bu bakımdan birbirlerinden farklıdır.

Dört farklı deney yapılmıştır. İlk deneyde bağlamsal model kullanılmamıştır, nesne ve yüzey seziciler LabelMe alt kümesinde çalıştırılmıştır. Hata matrisi Şekil 3(a)'da görülmektedir. İkinci deneyde yine bağlamsal model kullanılmamıştır, nesne ve yüzey seziciler Bilkent veri kümesinde çalıştırılmıştır. Onun da hata matrisi Şekil 3(b)'de verilmiştir. Üçüncü deneyde gerçek dünya ilişki modelleri Bilkent kümesinde eğitilmiş ve bağlamsal model LabelMe'de denenmiştir. Hata matrisi Şekil 3(c)'dedir. Dördüncü deneyde bu kez eğitim LabelMe alt kümesinde gerçekleştirilmiş ve bağlamsal model Bilkent kümesinde kullanılmıştır. Şekil 3(d)'de bu deneyin hata matrisi bulunmaktadır. İlk iki deney karşılaştırma amacıyla yapılmıştır. Son iki deneyde bağlamsal modelde geçen *tanımsız* ilişki olasılık parametresi 0.75 olarak seçilmiştir. Örnek sezimler Şekil 4'te verilmiştir.

Sonuçlar göstermiştir ki bağlam bilgisi kullanıldığında hem yanlış alarm hem de yanlış sezim sayılarında azalma olmuştur. Parametre olarak 0.90 değeri kullanıldığında yapılan hatalar daha da azalmıştır. Ancak bunun yanında doğru sezim sayısı da bir miktar azalmıştır. Parametre 0.50'ye çekildiğinde ise sezim hataları artmış, ama aynı şekilde doğru sezim sayısında da bir artış gözlenmiştir. Bu deneylerden anlaşılmıştır ki parametre ve başarımlar ölçütleri arasında ters orantı vardır. İleriki çalışmalarımızda en iyi parametreyi bulmak için bağımsız geçerlilik sınaması yapılacaktır.

Bir diğer ilginç gözlem ise doğru bir şekilde elenen (sezici tarafından atanmış oldukları sınıftan, bağlamsal model tarafından *bilinmeyen* sınıfına atanan) nesnelerin sayısının, yanlış elemelerden oldukça fazla olmasıdır. Örneğin Şekil 3(a)'dan anlaşıldığı gibi toplam klavye sezimi 1719 iken, bunlardan sadece 78 tanesi doğru sezimlere denk gelmektedir. Şekil 3(c)'de klavye sezim sayısının 621'e ve doğru sezimlerin 76'ya indiği görülmektedir. Bu da $1719 - 621 = 1098$ elemeyen yalnızca $78 - 76 = 2$ tanesinin hatalı olduğunu göstermektedir. Benzer durumlar diğer nesneler için de geçerlidir.

Gelecek çalışmalarda yeni nesne, ilişki sınıfları ve modelleri eklenerek bu yöntemin büyük ölçekli bir bilgisayarla görme sistemine dönüştürülmesi yolunda ilerlenecektir.

7. Kaynakça

- [1] A. Oliva and A. Torralba, "The role of context in object recognition," *Trends in Cognitive Sciences*, vol. 11, no. 12, pp. 520–527, 2007.
- [2] A. Torralba, "Contextual priming for object detection," *International Journal of Computer Vision*, vol. 53, no. 2, pp. 169–191, 2003.
- [3] A. Torralba, K. Murphy, W. Freeman, and M. Rubin, "Context-based vision system for place and object recognition," in *ICCV*, 2003.
- [4] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie, "Objects in context," in *ICCV*, 2007.

	Monitör	Klavye	Fare	Kupa	Boşluk
Monitör	100	3	16	3	51
Klavye	0	78	7	2	57
Fare	1	7	27	14	111
Kupa	0	0	6	18	27
Boşluk	226	1631	4567	2006	0

	Monitör	Klavye	Fare	Kupa	Boşluk
Monitör	71	0	2	1	9
Klavye	0	67	1	0	9
Fare	0	0	6	5	55
Kupa	0	0	11	19	7
Boşluk	95	664	1798	756	0

(a) Birinci deneyin hata matrisi

(b) İkinci deneyin hata matrisi

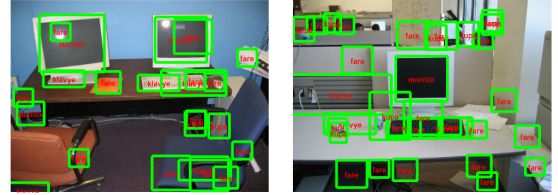
	Monitör	Klavye	Fare	Kupa	Boşluk
Monitör	95	1	11	0	66
Klavye	0	76	4	1	63
Fare	1	4	28	2	125
Kupa	0	0	5	13	33
Boşluk	69	540	1634	424	0

	Monitör	Klavye	Fare	Kupa	Boşluk
Monitör	66	0	0	1	16
Klavye	0	62	0	0	15
Fare	0	0	8	1	57
Kupa	0	0	7	16	14
Boşluk	40	282	846	206	0

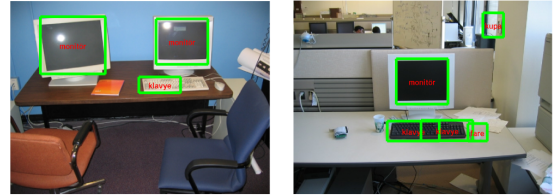
(c) Üçüncü deneyin hata matrisi

(d) Dördüncü deneyin hata matrisi

Şekil 3: Deneylerin hata matrisleri. Satırlar doğru sınıfları, sütunlar ise atanan sınıfları göstermektedir. Boşluk olarak gösterilen sütun sezimleme sonucunda bulunamayan nesnelere, boşluk olarak adlandırılan satır ise gerçekte olmayan fakat hatalı olarak bulunan nesnelere karşılık gelmektedir.



(a) Bağlamsal model olmadan sezimler.



(b) Bağlamsal model kullanıldığında sezimler.

Şekil 4: Nesne sezim örnekleri.

- [5] C. Galleguillos, A. Rabinovich, and S. Belongie, "Object categorization using co-occurrence, location and appearance," in *CVPR*, 2008.
- [6] D. Parikh, L. Zitnick, and T. Chen, "From appearance to context-based recognition: dense labeling in small images," in *CVPR*, 2008.
- [7] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton, "Learning Bayesian classifiers for scene classification with a visual grammar," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 581–589, March 2005.
- [8] A. Torralba, "A simple object detector with boosting," *ICCV Short Courses on Recognizing and Learning Object Categories*, 2005.