

Nesne tanımda bağlam ve anlambilimsel sınıflandırmanın önemi: Bilgisayarla görme ve insanda görme alanlarındaki çalışmalar

The importance of context and semantic descriptions in object recognition: Studies in computer vision and human vision

Selim Aksoy¹, Hüseyin Boyacı², Didem Gökçay³

1. Bilgisayar Mühendisliği Bölümü

Bilkent Üniversitesi

saksoy@bilkent.edu.tr

2. Psikoloji Bölümü

Bilkent Üniversitesi

hboyacı@bilkent.edu.tr

3. Enformatik Enstitüsü, Sağlık Bilişimi Bölümü

Orta Doğu Teknik Üniversitesi

didem@ii.metu.edu.tr

Özetçe

Sahne sınıflandırması ve nesne tanıma, bilgisayarla görme alanında çok uzun yıllardır üzerinde çalışılan temel problemlerdir. Bilgisayarlara kazandırılmaya çalışılan, sahnelerin ve içerdikleri nesnelerin otomatik olarak tanınması ve etiketlenilmesi yeteneği, yapay zeka konusunda önemli bir adım olma potansiyeli taşımaktadır. Bu yetenek, her geçen gün artan hesaplama kapasitesi ile birleştirildiğinde de henüz çözümü mümkün olmayan birçok uygulamanın da gerçekleştirilmesi mümkün olacaktır. Yazımızda, nesne tanımayı kolaylaştıran anlambilimsel sınıflandırma yaklaşımları, hem bilgisayarla hem de insanda görme alanlarındaki literatür üzerinden özetlenmektedir.

Abstract

. Object recognition and scene classification are among the main interests in computer vision which have been investigated for long. Automatic recognition and classification of objects and scenes is an important skill to be gained by computers, especially in the field of artificial intelligence. Merging this skill with the ever increasing computing power of the computers will help in the development of many applications that are yet to be resolved. In this article, we present a survey on contextual and semantic approaches for object recognition by reviewing both computer vision and human vision literatures

1. Giriş

Bilgisayarla görme açısından bakıldığında, sahne sınıflandırması ve nesne tanıma probleminde iki ana bileşen vardır: içerik modelleme ve anlambilimsel sınıflandırma.

1.1. İçerik Modelleme

Bilgisayarla görme açısından bakıldığında, sahnelerin içeriklerinin modellenmesi konusunda geleneksel olarak izlenilen yöntem, görüntülerin bütününden çıkarılan düşük seviyeli (low-level) özelliklerin (feature) kullanılmasıdır. Bu konuda, renk ve doku histogramları gibi özellikler görüntüleri iç mekan/dış mekan, şehir/doğa gibi sınıflara ayırtmak için kullanılmıştır. Örneğin, Vailaya ve ekibi [1] renk ve ayırt yönelim histogramları kullanarak Bayesçi sınıflandırıcılar ile görüntüleri bu tür ikili gruplara ayırmıştır.

Bu tür yöntemlerle küçük ve sınırlı veri kümelerinde başarılı sonuçlar elde edilmesine karşın, görüntülerdeki nesnelerin çeşitliliği ve karmaşık arkaplanlar görüntülerin sınıflandırılması için bölge temelli yerel özelliklerin analizini gerektirmektedir. Bu konuda son zamanlardaki en popüler yöntemlerden biri, görüntülerde ilgi noktalarının (interest points) [2] bulunmasına, bu noktaların çevrelerinden çıkarılan özelliklerin nicemlenmesi ile bir kod tablosu oluşturulmasına, ve görüntülerin bu kod tablosundaki görsel kelimeler (visual words) kullanılarak modellenmesine dayanmaktadır. Örneğin, Fei-Fei ve Perona [3], bu tür görsel kelimeleri ve oluşturdukları grupları öğreticisiz bir şekilde öğrenen yöntemleri sahne sınıflandırması için kullanmışlardır. Quelhas ve ekibi [4] de benzer modelleri görüntü bölgelerini

insan yapısı ya da doğal yapı şeklinde sınıflandırmak için kullanılmışlardır.

İlgi noktalarıyla oluşturulan görsel kelimelerin histogramlarını (bag-of-words) kullanan yöntemler nesne tanıma problemine de çözüm olarak sunulmuştur [5]. Histogram kullanan bu yöntemlerin uzamsal bilgiyi (spatial information) gözardı etmesi nedeniyle oluşan problemlerin çözümü için, Lazebnik ve ekibi [6] görüntüleri ızgaralara bölmüş, ve histogramları her bir ızgara bölgesi için ayrı ayrı hesaplamıştır. Bu yöntem, görüntünün bütününden hesaplanan görsel kelime histogramları karşısında daha başarılı olsa da sabit bir ızgara kullanılması nedeniyle uygulama alanı çok kısıtlıdır. Ayrıca, bu tür yöntemler, Caltech veri kümesi [7] gibi sadece tek bir nesne içeren görüntülerden oluşan verilerde yüksek başarı oranı verebilir de, birden fazla nesne içeren daha karmaşık görüntülerin analizine ve genel nesne tanıma probleminin çözümüne katkı yapabilmekten henüz uzaktır.

1.2. Anlambilimsel sınıflandırma

Bilgisayarla görme literatüründe sahne sınıflandırması ve nesne tanıma için önerilen yöntemler, bu iki önemli problemi genel olarak birbirinden bağımsız olarak incelemiştir. Oysa, görüntülerdeki nesnelerin çeşitliliği ve karmaşık arkaplanlar hem görüntünün bütününden elde edilen bilgilerin hem de yerel özelliklerin kullanımını gerektirmektedir. Görüntüdeki bağlam bilgisinin hem insan algılamasındaki hem de bilgisayarla görmedeki önemi bilinmektedir. Yerel özniteliklerin, görüntünün anlamsal açıdan ilgisiz bölgelerinde de bulunabiliyor oluşu, sahne sınıflandırma ve nesne tanımda bağlam bilgisinin kullanılmadığı durumlarda sadece yerel bilgileri kullanan bu yöntemleri de başarısız kılabilir. Üstelik, her ne kadar görüntüdeki bölgeler doğru olarak sınıflandırılabilir olsa dahi aynı bölgeler, farklı sahnelerde, içeriğe bağlı olarak farklı anlamlar kazanabilmektedir.

Son yıllarda ortaya çıkan çalışmalar, hem sahne sınıflandırmada hem de nesne tanımda bağlam bilgisini kullanarak düşük seviyedeki renk, doku, şekil gibi özniteliklerle görüntülerin algılanan anlamları arasındaki anlambilimsel uzaklığı (semantic gap) giderek kapamayı amaçlamaktadır. Bu çalışmalar, görüntülerdeki bağlam bilgisini nesnelerin birbirleriyle ve bulundukları sahneyle olan ilişkilerini değişik seviyelerde modelleyerek çıkarmaktadır. Nesneler ve sahnelerdeki kurguları arasındaki ilişkiler aşağıdaki açılardan gruplanabilir [8]:

- i. Destek :** Nesneler ve üzerinde bulundukları yüzeyler,
- ii. Olasılık:** Bazı nesnelerin sadece bazı ortamlarda bulunması,
- iii. Konum:** Nesnelerin sadece bazı konumlarda bulunması,
- iv. Göreli büyüklük:** Beraber bulunabilecek nesnelerin sa-hip olabilecekleri göreli büyüklüklerin sınırlı olması.

Sahnelerin, içerdikleri nesnelerin ve bu nesnelerin birbirleriyle olan ilişkilerinin modellenmesinde sahne ve nesne arasında karşılıklı etkileşim mevcuttur. Bir sahnenin anlambilimsel olarak sınıflandırılması, içerdikleri nesnelerin tanınmasıyla mümkün olabilmektedir. Bunun yanında, bir nesnenin birbirine çok benzeyen alternatifler arasından hangisi olduğu bulunduğunda sahneye bakılarak daha kolay kararlaştırılabilir. Örneğin, renk, şekil ve büyüklük özellikleri itibarıyla sarı,

yuvarlak ve küçük bir nesne, hem bir limona hem de bir tenis topuna benzemekten, görüntünün bir tenis karşılaşması içerdiği bilindiğinde, bu nesnenin tenis topu olabileceğine kolayca karar verilebilecektir. Benzer şekilde, bir görüntüdeki bölgeler sadece piksel değerlerine bakılarak tek başlarına anlamlandırılmasa bile, bağlam bilgisi yardımıyla belirsizliklerin çözümü kolaylaşabilmektedir.

2. İnsanda görme alanında örnek çalışmalar

Nesne algılama konusu, insanda görme açısından ele alındığında, insanların doğal ortamlarda pek çok görsel uyarana maruz kaldığı ve bu uyaranların görme sistemimize tek başlarına değil, çevresel bir mızanda sunulduğu düşünülmelidir. Günümüzde insanda görme alanında tek bir uyaran içeren değerli çalışmalar olmakla birlikte, görülen odur ki gerçek hayattaki karmaşık bağlamlarda görme sisteminden elde edilen davranışsal ve nöral tepkiler, tekil uyaranlara gösterilen davranışsal ve nöral tepkilerin birleştirilmesi ile elde edilememektedir. Bu durumda, insanın görsel algıdaki başarısının anlaşılabilmesi için çevresel faktörlerin gözardı edilemeyeceği anlaşılmaktadır.

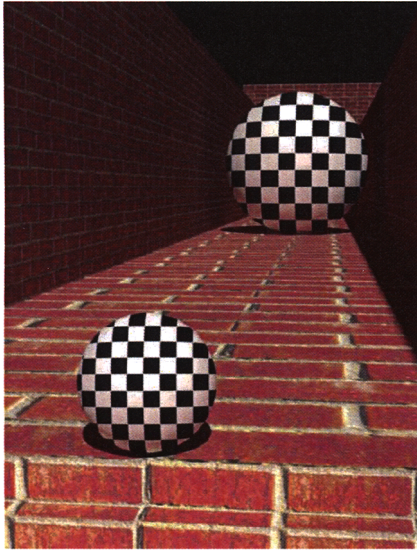
Geçtiğimiz yüzyılda insanın görsel sisteminin özellikle alt-düzye özniteliklere gösterdiği tepkiler ayrıntısıyla çalışılmıştır.

En önemli alt-düzye öznitelikler, her bir nöronun fiziksel olarak tek bir uyarana bağlı olarak aktivite göstermesi ile ortaya çıkar. Örneğin, birincil görme korteksinde (V1) (primary visual korteks) her bir nöron, görsel alanımızdaki ufak bir bölgedeki ışığa karşı duyarlıdır [25]. Bu duyarlılık fovea'ya düşen görüntüde sadece 1 derecelik bir alanı kapsar. Noktasal duyarlılığı olan nöronların birleşmesi ile doğrusal öznitelikler, doğrusal özniteliklerin birleşmesi ile de şekilsel öznitelikler oluşturulur. Doğal manzaraların bağımsız bileşenler analizi (ICA) ile ayrıştırılması sonrasında elde edilen görsel bileşenlerin, birincil görme korteksindeki kortikal kolonların temsil ettiği doğrusal öznitelikler ile bire bir örtüşmesi, V1'deki nöron popülasyonları arasındaki ileri-doğru (feed-forward) bağlantıların varlığının göstergesidir [29].

Yakın zamana kadar geçerli olan görüş, alt-düzyede gerçekleşen ve temel olarak özniteliklerden oluşan görsel algıların, nöronların oluşturduğu ileri-doğru (feed-forward) bağlantılar aracılığıyla üst düzeylere aktarıldığı, ve üst düzey bölgelerde nesnel yada renk bilgisi içeren karmaşık algılara dönüştüğü yönündeydi. Bu görüşe göre üst düzeyde bakıldığında insan görsel sisteminde 'ne' ve 'nerede' bilgisini kodlayan iki ayrı yolak (pathway) mevcuttur. 'Ne' yolağı ('what' pathway), nesneleri algılamamıza ve sınıflandırmamıza, 'Nerede' yolağı ise hareketi algılamamıza yardım eder. Geçtiğimiz onyılda, 'Ne' yolağını içeren pek çok işlevsel beyin görüntüleme çalışması yapılmıştır. Bu çalışmalardan öğrendiklerimize göre nesneler, üst düzey semantik özelliklerine göre adeta bir grid yapısı oluşturularak beyindeki özel bölgelerde depolanmaktadır. Tanıdığımız canlı yada cansız objeler, aletler, hayvan resimleri, insan yüzleri [30,35], semantik özelliklerinin ortaklığı [31], yada işlevlerine göre [32], genelden özele [32], hatta kişiden kişiye farklılık göstererek, örneğin kişinin uzmanlık alanına göre sınıflandırılarak [33] tanımlanmaktadır. Üst düzeydeki semantik özniteliklere bağlı bu karmaşık yapının insan

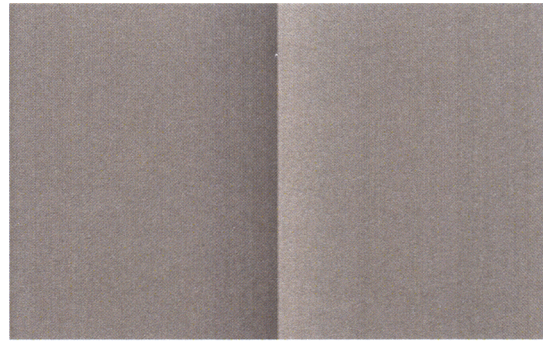
beyninde nasıl oluşturulageldiği, hala çözümlenememiş bir konudur. Burada anlatılanları görsel alanımızdan türetilen öznitelikler ile ilişkilendirecek olursak, alt-düzeylerde 1^0 gibi görsel alanın çok ufak bir kısmına karşılık gelen bölgeye tepki veren nöronlar, bilginin birleştirilerek aktarımı sayesinde, üst düzeylere gelindiğinde, 50^0 gibi görsel alanın çok büyük bir kısmını kaplayan bölgelere tepki verebilir hale gelirler [21].

Diğer yandan, görsel olarak nesne tanımının sadece ileri-doğru bağlantılar ile gerçekleştiği savı insan görsel sisteminde oluşan bazı illüzyonları açıklayamamaktadır. Örneğin 'ay yanılsaması'nda, aynı ebatlardaki ay, şehir silüetinin üzerine oturtulduğunda olduğundan daha büyük, gökyüzünde tepedeyken tek başına gözleendiğinde ise daha küçük olarak algılanır (moon illusion). Oysa görsel sistemimizde sadece ileri-doğru bağlantılar olsaydı, ayın ebatları iki mizansende de tıpa tıp aynı olduğu için alt-düzey öznitelikler saptanırken aynı miktarda nöron işe koşulacağından, iki mizansende de üst-düzeylere aynı bilgi iletilecek, ve ayın algılanması, çevresel faktörden bağımsız olarak, iki mizansende de eşit olarak gerçekleşecekti. Şehir silüeti olan durumda ayın daha büyük algılanmasının bir nedeni, üst-düzeyde tanınan çevresel faktörlerin (contextual influence) geri-besleme yolu ile alt-düzeye iletilmesi ve burada yer alan erken görme (early vision) bölgelerindeki öznitelikleri manipüle etmesi olabilir. Bu varsayımı denemek için yapılan bir çalışmada, şekil 1 de gösterildiği gibi, küresel bir objenin uzaktayken daha büyük algılandığı, dolayısıyla birincil görme korteksinde daha çok miktarda nöronun tepkisine neden olduğu, aynı obje yakında olduğunda ise daha küçük algılandığı ve görme korteksinde daha az miktarda nöronun bu objeye tepki verdiği gösterilmiştir [26]. Buradan anlaşıldığı gibi, önceki varsayımların tersine, görme alanımızdaki noktalar bire bir olarak alt-düzeyde görme korteksindeki nöronlar tarafından temsil edilmemekte, görmenin oluşturduğu üst-düzey uzamsal algılar, örneğin bir objenin algılanan ebatları, birincil görme korteksindeki nöronların aktivitesinde belirleyici olmaktadır.



Şekil 1: Ebatları aynı olan iki küreden öndekinin küçük, arkadakinin büyük algılanışı [26]

Görsel sistemdeki üst-düzey ve alt-düzey yapılar arasındaki geri beslemenin varlığı, ışık yeğinliğinin (intensity) algılanması ile ilgili çalışmalarda da gösterilmiştir. Açıklık (lightness), bir yüzeyin ışık yansıtıcılığının kişiye bağlı algılanış durumudur. Açıklık, yüzeyin yansıtıcılığı ve ortamın aydınlığına bağlı olduğu kadar, objenin içinde bulunduğu bağlam ile de yakından ilgilidir [20,24]. Şekil 2 de verilen CO (Craik-O'Brien [27]) uyarını, açıklığın bağlama dayalı kısmını araştırmakta kullanılan etkili bir görseldir. Boyacı ve diğerleri tarafından yapılan çalışmalarda [22,23], Şekil 2'deki her iki yüzeyin aydınlanması aynı olduğu halde, ortadaki ayraç nedeniyle, iki yüzeyin açıklığının farklı olarak algılandığı, ve birincil görme korteksindeki nöronların aktivitesinin, aydınlanmadan çok algılanan açıklık ile orantılı olduğu gösterilmiştir. Bu çalışmalarda, bağlam bilgisinin açıklığı etkilediği, bağlam değiştikten 500-1000ms sonra açıklık algısının da değiştiğinin gösterilmesi sayesinde anlaşılmıştır.



Şekil 2: CO uyarısında eşit aydınlanmış iki yüzeyde, ortadaki gradyan nedeniyle sol yüzey koyu, sağ yüzey açık olarak algılanır [27]

Aslında uzun yıllar önce Felleman ve Van Essen tarafından yapılan, maymunlardaki görme sisteminin bağlantılarını inceleyen bir çalışma [34], görme sistemimizin karmaşıklığını gözönüne sermektedir. Bu çalışmaya göre, sadece maymun beynindeki görme sisteminde, farklı işlevler üstlenen 37 civarında alan hiyerarşik bir yapıda sıralanmıştır. Bu alanlar arasında, direkt/indirekt, ileri/geri ve aynı seviyedeki bölgelerde kendi aralarında (lateral, recurrent) çok karmaşık bağlantılar mevcuttur. Görsel sistemimizdeki pek çok kademe arasındaki karmaşık bağlantıların arasındaki ilişkileri araştırmak için her kademenin teker teker uyarılacağı deneyler düzenlenmektedir. Örneğin, bir kademedeki aktivitenin diğerini etkilemesini engelleyebilmek için deneğe gösterilen görsel uyarandan ilgiyi uzaklaştırıcı, dikkat dağıtıcı başka uyarılar karıştırılması bir yöntemdir. Flaş baskılama tekniği adı verilen bir başka yöntemde ise, bir göze temel görsel uyarı verilirken, diğer göze bu uyarının bilinç düzeyindeki algısını -ve dolayısıyla bağlam bilgisini-baskılayıcı yüksek kontrastlı gürültü gösterilir [28]. Günümüzde insanda görme çalışmalarında temel hedef, görsel sistemde farklı işlevler üstlenen modüller alanların ve bunlar arasındaki karmaşık ilişkilerin çözümlenebilmesidir.

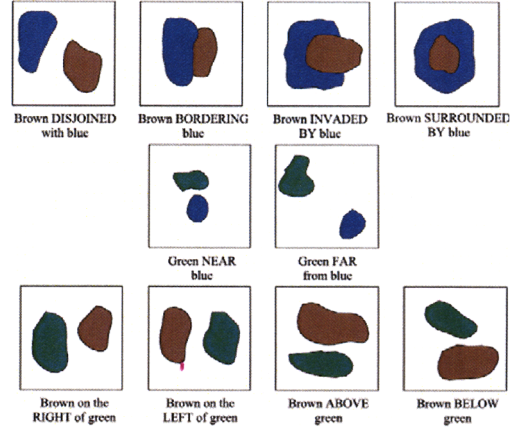
3. Bilgisayarla görme alanında örnek çalışmalar

Bilgisayarla görme alanına döndüğümüzde, nesne tanımının başarısını artırmak için nesnelerin yer aldığı sahnelerin modellenmesinin sonuçları iyileştirdiği gözükmemektedir. Hoiem ve ekibi [10], bir görüntüdeki düzlemleri kestirerek üç boyutlu geometrik yapıyı modellemeyi amaçlamıştır. Bulunmaya çalışılan düzlemler, yer düzlemi, yere dik düzlem ve gökyüzü olarak üç gruba ayrılmıştır. Bunun için, görüntüler bir bölütleme algoritmasıyla küçük parçalara ayrılmış, bu parçalardan renk, doku, konum, şekil ve doğru parçası istatistikleri gibi özneliteler çıkarılmış, Adaboost yöntemi kullanılarak sınıflandırma yapılmıştır. Hoiem ve ekibi sonraki çalışmalarında [9] kestirilen üç boyutlu geometrik bilgiyi yine görüntüden kestirilen bakış açısı bilgisi ile birleştirerek nesne tanıma algoritmalarının çıktılarını düzeltmişlerdir. Kestirilen bilgiler, nesne tanıma sonucunda ortaya çıkan bölgelerin büyüklüklerini ve sahneye göre olan göreceli konumlarını da kullanarak sonuçları iyileştirmiştir.

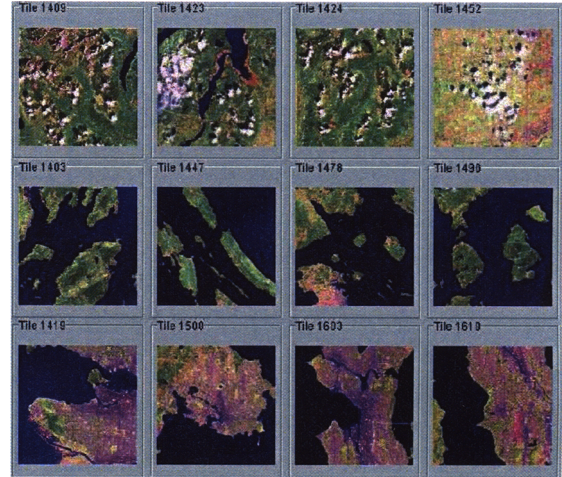
Rabinovich ve ekibi [8] ise bölütleme sonucu elde edilen nesne adaylarını hem bölgelerin tek tek içeriklerini hem de komşu oldukları bölgelerin içeriklerini kullanarak sınıflandırmış, böylece nesne tanımda oluşan belirsizlikleri en aza indirmeye çalışmıştır. Komşu bölgeler arasındaki etkileşimleri modellemek için koşullu rasgele alan (conditional random field) kullanılmış, bu modellerin öğrenilmesi aşamasında da öğrenme veri kümesindeki görüntülerdeki etiketlenmiş nesnelerin beraber bulunma istatistiklerinden yararlanılmıştır.

Parikh ve Chen [11] de benzer bir varsayımla nesneleri ve beraber bulunabilecekleri diğer nesneleri sıradüzensel bir şekilde modellemiştir. Bu modelde, ilgi noktaları sıradüzensel ve göreceli konum temelli tanımlar önerilmiştir. Bu ilişki bulma yöntemlerinin çoğu, noktalar arası açı hesaplamalarına dayanmaktadır [12]. Genellikle, nesnelerin kütle merkezleri arasındaki açı veya nokta ikililerinin arasındaki açılardan oluşturulan açı histogramları göreceli konumu yaklaşık olarak hesaplamakta kullanılmaktadır. Ayrıca, kuvvet histogramı, projeksiyon alma ve biçimbilime dayanan yöntemler de bulunmaktadır [12]. Bölütleme sonucunda elde edilen bölgeler arasındaki topolojik, uzaklığa dayalı ve göreceli konuma dayalı ilişkilerin (Şekil 3) otomatik yöntemlerle çıkarılması ve bulanık (fuzzy) modellerle gösterimi için algoritmalar geliştirilmiş ve bu modellerin görüntü sınıflandırma ve görüntü erişimi uygulamalarında geleneksel piksel veya bölge temelli yöntemlerin uygulanamayacağı durumlardaki kullanımı gösterilmiştir. Şekil 4, uydu görüntülerindeki çeşitli yapıların bölütleme sonucu elde

edilen bölgelerin uzamsal ilişkileri kullanılarak modellenmesi ile ilgili örnekler içermektedir.

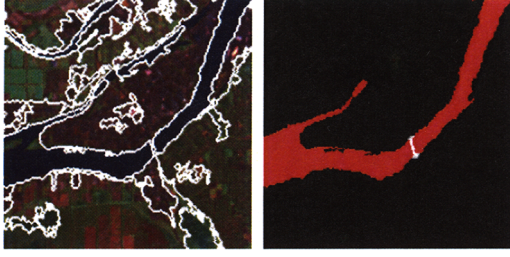


Şekil 3: Bölge çiftleri arasındaki topolojik, uzaklığa dayalı ve göreceli konuma dayalı ilişkiler [13].



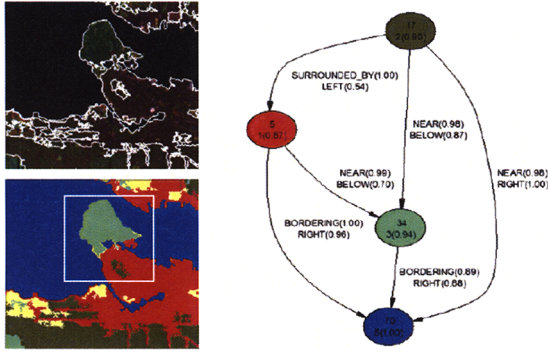
Şekil 4: Bölge ilişkileri kullanılarak bulutların, adaların ve deniz kıyısındaki şehirlerin bulunması [13].

Daha sonraki çalışmalarda [14] ise bir nesnenin diğer iki nesne arasında kalması durumunu modelleyen “arada” ilişkisi için nesnelerin şekillerini (örneğin, içbükeylik), uzamsal duruş ve uzaklık bilgisini de hesaba katan matematiksel biçimbilime dayalı yöntemler önerilmiştir. Şekil 5’te bir uydu görüntüsünde asfalt olarak sınıflandırılmış bir bölgenin su olarak sınıflandırılmış iki ayrı bölge arasında kaldığında köprü olarak tanımlanabileceği gösterilmiştir. Ayrıca, tek tek sınıflandırılmış bölgelerin ikili gruplar halinde üzerinde- altında ilişkileri de kullanıldığında sahne sınıflandırma başarısının arttığı da ortaya konmuştur [15].



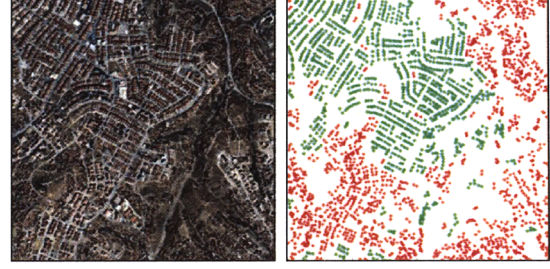
Şekil 5: Arada ilişkisi kullanılarak köprülerin bulunması [14].

Literatürde, ikili ve üçlü ilişkilerin yanında daha karmaşık ilişkileri modellemek için çizge (graph) bazlı yöntemler de kullanılmıştır. Örneğin, Boutell ve ekibi [16] bir görüntüdeki bölgeleri gökyüzü, bulut, çimen, su, kum gibi sınıflara ayırdıktan sonra bu bölgeler arasındaki ilişkileri çizgeler kullanarak modellemiştir. Bölge sayıları ve ilişki çeşitlerinin artması durumunda bu çizgelerin karmaşıklığı da çok arttığı için sahne sınıflandırması aşamasında faktör çizgeleri (factor graphs) kullanmıştır. Benzer şekilde, uydu görüntülerindeki karmaşık yapıların çizgeler ile modellendiği durumlarda (Şekil 6) hem bu çizgelerin oluşturulması aşaması için hem de çizgeleri kullanarak görüntü arşivlerinde etkili ve hızlı erişim için geliştirilmiş yöntemler de mevcuttur [17].

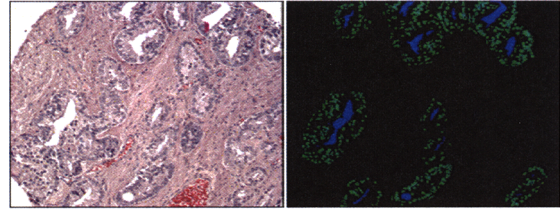


Şekil 6: Bir uydu görüntüsü, otomatik olarak sınıflandırılmış bölgeler ve beyaz kare ile gösterilen alan için oluşturulmuş çizge [17].

Bu tür istatistiksel ve yapısal modeller, bölütleme ve basit sınıflandırma algoritmalarıyla elde edilemeyecek karmaşık yapıların bulunması konusunda oldukça umut vericidir. Örneğin, Şekil 7'de bina gruplarının düzenli yerleşim (planlı şehirleşme) ve düzensiz yerleşim (gecekondu) şeklinde ayrıştırılması [18], Şekil 8'de kanser araştırmalarında hücre gruplarının tanınması [19] konularında yeni gelişmeler gösterilmiştir.



Şekil 7 :Bina gruplarının yerleşim yapılarına göre ayrıştırılması: düzenli (yeşil) ve düzensiz (kırmızı) yerleşimler [18].



Şekil 8: Mikroskoptan alınmış bir görüntüde hücre gruplarının bulunması [19].

4. Sonuç

Son zamanlarda yayınlanan ve sadece yüz, insan ya da araba gibi belirli nesneleri tanımayı amaçlayan algoritmaların başarısında önemli gelişmeler gözlenmektedir. Özellikle, sınırlanmış ortamlarda çekilmiş görüntülerde çok güzel sonuçlar elde edilmiştir. Bununla birlikte, herhangi bir şekilde kısıtlanmamış ve çok sayıda değişik nesneleri içeren görüntülerde başarılı sonuçlar veren soysal (generic) nesne tanıma algoritmaları henüz mevcut değildir. Bu tür görüntülerde bölütleme algoritmaları da henüz istenilen başarı seviyelerine ulaşamamıştır. Bilgisayarla görme alanında en önemli problemler arasında yer alan bölütleme ve nesne tanıma, insanda görmenin nasıl gerçekleştiği incelenerek ve sahne sınıflandırması ile bağlam modelleme algoritmalarına bu doğrultuda katkılar yapılarak geliştirilebilir.

5. Kaynakça

- [1] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, H.-J. Zhang, "Image Classification for Content-Based Indexing", *IEEE Trans. on Image Processing*, 10(1):117-130, January 2001.
- [2] D. G. Lowe, "Distinctive Image Features From Scale-Invariant Keypoints", *International Journal of Computer Vision*, 60(2):91-110, November 2004.
- [3] L. Fei-Fei, P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories", *IEEE Conference on Computer Vision and Pattern Recognition*, 2:524-531, June 20-25, 2005.
- [4] P. Quelhas, F. Monay, J.-M. Odobez, D. Gatica-Perez, T. Tuytelaars, "A Thousand Words in a Scene", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(9):1575-1589, September 2007.

- [5] L. Fei-Fei, R. Fergus, P. Perona, "One-shot learning of object categories", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(4):594-611, April 2006.
- [6] S. Lazebnik, C. Schmid, J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories", *IEEE Conference on Computer Vision and Pattern Recognition*, 2:2169-2178, June 2006.
- [7] http://www.vision.caltech.edu/Image_Datasets/Caltech101
- [8] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, S. Belongie, *IEEE International Conference on Computer Vision*, October 14-21, 2007.
- [9] D. Hoiem, A. A. Efros, M. Hebert, "Putting Objects in Perspective", *IEEE Conference on Computer Vision and Pattern Recognition*, 2:2137-2144, 2006.
- [10] D. Hoiem, A. A. Efros, M. Hebert, "Geometric Context From a Single Image", *IEEE International Conference on Computer Vision*, 1:654-661, October 2005.
- [11] D. Parikh, T. Chen, "Hierarchical Semantics of Objects", *IEEE International Conference on Computer Vision*, October 14-21, 2007.
- [12] I. Bloch and A. Ralescu, "Directional relative position between objects in image processing: A comparison between fuzzy approaches", *Pattern Recognition*, 36(7):1563-1582, July 2003.
- [13] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton, "Learning Bayesian classifiers for scene classification with a visual grammar," *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):581-589, March 2005.
- [14] R. G. Cinbis, S. Aksoy, "Relative Position-Based Spatial Relationships Using Mathematical Morphology," *IEEE International Conference on Image Processing*, 11:97-100, San Antonio, Texas, September 16-19, 2007.
- [15] D. Gokalp, S. Aksoy, "Scene Classification Using Bag-of-Regions Representations," *IEEE International Conference on Computer Vision and Pattern Recognition, Beyond Patches Workshop*, Minneapolis, Minnesota, June 23, 2007.
- [16] M. R. Boutell, J. Luo, C. M. Brown, "Scene Parsing Using Region-Based Generative Models", *IEEE Transactions on Multimedia*, 9(1):136-146, January 2007.
- [17] S. Aksoy, "Modeling of Remote Sensing Image Content using Attributed Relational Graphs", *IAPR International Workshop on Structural and Syntactic Pattern Recognition*, 475-483, Hong Kong, China, August 17-19, 2006.
- [18] E. Dogrusoz, S. Aksoy, "Modeling Urban Structures Using Graph-Based Spatial Patterns", *IEEE International Geoscience and Remote Sensing Symposium*, 4826-4829, Barcelona, Spain, July 23-27, 2007.
- [19] S. Aksoy, "A Bayesian Textual and Multimedia Information Retrieval Engine", Technical Report, Insightful Corporation, Seattle, WA, USA, 2003
- [20] E.H. Adelson, "Perceptual organization and judgment of brightness," *Science*, 262: 2042-2044, 1993
- [21] D. Boussaoud, R. Desimone, L.G. Ungerleider, "Visual topography of area TEO in the macaque," *The Journal of Comparative Neurology* 306: 554-575, 1991.
- [22] H. Boyaci, F. Fang, S.O. Murray, D. Kersten, "Responses to Lightness Variations in Early Human Visual Cortex," *Current Biology* 17: 989-993, 2007.
- [23] H. Boyaci, F. Fang, S.O. Murray, G. Albenese, D. Kersten, "Time course of cortical responses to illusory and real lightness changes," *The Annual Meeting of the Vision Science Society*, Naples, Florida, May 2008
- [24] A. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, E. Economou, "An anchoring theory of lightness perception," *Psychological Review* 106: 795-834, 1999
- [25] D.H. Hubel, T.N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology* 160: 106-154, 1962
- [26] S.O. Murray, H. Boyaci, D. Kersten, "The representation of perceived angular size in human primary visual cortex," *Nature Neuroscience*, 9: 429-434, 2006
- [27] V. O'Brien, "Contour perception, illusion and reality," *J. Opt. Soc. Am.*, 48: 112-119, 1958
- [28] N. Tsuchiya, and C. Koch, "Continuous flash suppression reduces negative afterimages," *Nature Neuroscience*, 8: 1096-1101, 2005
- [29] J.H. Van Hateren, A. Van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex", *Proceedings of Royal Society London B*, 265: 359-366, 1998
- [30] L. Reddy, N. Kanwisher, "Coding of visual objects in the ventral stream", *Current Opinions in Neurobiology*, 16(4): 408-414, 2006
- [31] B. Crosson, M.A. Cato, J.R. Sadek, L. Lu, "Organization of semantic knowledge in the human brain: toward a resolution in the new millenium", *Brain and Cognition*, 42(1): 146-148
- [32] A. Martin, L.L. Chao, "Semantic memory and the brain: structure and processes", *Current Opinion in Neurobiology*, 11:194-201, 2001
- [33] C.M. Bukach, I. Gauthier, M.J. Tarr, "Beyond faces and modularity: the power of an expertise framework", *Trends in Cognitive Science*, 10(4): 159-66, 2006
- [34] D.J. Felleman, D.C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex", *Cerebral Cortex*, 1:1-47, 1991
- [35] R.F. Schwarzllose, C.I. Baker, N. Kanwisher, "Separate face and body selectivity on the fusiform gyrus", *Journal of Neuroscience*, 25(47):11055-9, 2005