

# Haber Videoları için Yüz Bulma Yöntemlerinin İyileştirilmesi<sup>1</sup> Improvement of Face Detection Algorithms for News Videos

Nazlı İkizler, Pınar Duygulu

Bilgisayar Mühendisliği Bölümü, Bilkent Üniversitesi, 06800, Ankara

inazli@cs.bilkent.edu.tr, duyulu@cs.bilkent.edu.tr

## Özetçe

Haber videolarında en önemli öge olan kişiler ile ilgili görüntülerin erişimi için etkin yüz bulma çok önemli bir aşamadır. Ancak, videoda yüz bulma işlemi, verilerin çeşitliliği ve görüntülerin gürültülü olması nedeniyle zor bir problemdir. Bu çalışma, video veri tabanları üzerinde, varolan en iyi yüz bulma yöntemlerinden biri olan Schneiderman-Kanade yöntemi ile ten bulma yöntemlerini birleştirerek, çok daha başarılı bir yüz bulma yöntemi sunmaktadır. Yöntem, TRECVID 2003 verileri üzerinde denenmiş ve oldukça umut verici sonuçlar elde edilmiştir.

## Abstract

People are the most important subjects in news videos and for proper retrieval of person images, face detection is a very crucial step. However, face detection and recognition in news videos is a very challenging task due to the huge irregularities and high noise level in the data. This study presents a method that combines skin detection and Schneiderman-Kanade face detection, for improving the face detection performance in news videos for a better retrieval. This method has been tested on TRECVID 2003 dataset and the results are very promising.

## 1. Giriş

Gelişen teknoloji ile birlikte çok sayıda verinin işlenmesi ve saklanması mümkün hale gelmiş; resim, video, ses, müzik gibi farklı verileri bir arada içeren arşivler büyük önem kazanmaya başlamıştır [1]. Bu arşivler üzerinde etkin bir erişim ve değerlendirme yapabilmenin ana koşulu, anlamsal bilgilerin çıkarılması ve ulaşılabilir hale getirilmesidir.

Haberler, sosyal ve kültürel birikimin ortaya çıkarılması, tarihsel sürecin daha iyi anlaşılabilmesi için oldukça büyük önem taşımaktadır. Bu nedenle haber videolarından oluşan bir arşive hızlı ve doğru bir şekilde erişim gerekmektedir. Bu önemi dolayısıyla haber videoları, son iki yıldır, NIST (Amerikan Standartlar Enstitüsü) tarafından düzenlenen TRECVID video erişimi değerlendirme yarışmasının da konusunu oluşturmaktadır [2]. Bu yarışmada amaç, verilen bazı sorgu cümlelerine denk gelen video bölümlerini kullanıcıya sunmaktır.

Haber videolarında en çok rastlanan sorgular kişiler ile ilgili olanlardır. Yapılan çalışmalar göstermiştir ki, sorgu sadece metin üzerinde yapıp görsel bilgi kullanılmadığı zaman yanlış sonuçlarla karşılaşılabilir. Örneğin, bir kişi ile ilgili bir haber spiker tarafından sunulurken kişinin ismi geçmekte, oysa kişinin görüntüsü ilerleyen dakikalarda, belki o kişi konuşurken verildiğinden, kişi ve isim arasındaki ilişki kaybedilmektedir. Bu nedenle çoğu zaman ses tanıma yöntemlerinin ürettiği metinler üzerinde yapılan sorgular hatalı sonuç üretmekte, istenilen kişinin görüntüsü yerine spikerin görüntüsü kullanıcıya sunulmaktadır.

Sorgulanan kişi ile ilgili doğru haber görüntülerinin bulunabilmesi için görsel verilerin kullanılması ve otomatik olarak yüzlerin tanınması gerekmektedir. Ancak, yüz tanıma problemi, yapay zeka ve bilgisayarla görü alanlarında çok çalışılmış bir konu olmasına rağmen, henüz çözülmüş değildir. Elde edilen başarılı tanıma oranları kontrollü ortamlarda çekilmiş resimler içeren veri tabanlarında sınırlı kalmaktadır [3]. Özellikle video gibi poz ve aydınlatmanın çeşitlilik gösterdiği ve resim çözünürlüğünün düşük olduğu veri arşivlerinde yüz tanımak çok daha zor bir problem haline gelmektedir. Gerçek zamanlı ortamları daha iyi yansıtan bu gibi veri tabanları üzerinde yüz tanıma problemini incelemek, kontrollü ortamlarda elde edilmiş resimleri incelemekten çok daha farklı ve geniş kullanım alanlarına sahiptir.

Yüz tanıma yöntemlerinin en önemli aşaması görüntüde yer alan yüzlerin bulunmasıdır. Bulunan yüzlerin sayısı ve kalitesi yüz tanıma yönteminin başarısını doğrudan etkilemektedir. Literatürde yüz bulma amaçlı birçok yöntem yer almakta [4], ancak çoğu sistem video arşivinin gürültülü yapısı ve düşük çözünürlüğü nedeniyle yetersiz kalmaktadır.

Bu çalışmada, haber videolarındaki yüzlerin doğru olarak bulunabilmesi hedeflenmiştir. Bu amaçla, şu anda literatürdeki en iyi yüz bulma sistemlerinden biri olan Schneiderman ve Kanade'nin yöntemi [5], ten rengi tabanlı bir yöntemle birleştirilmiş; tek başlarına yetersiz kalan iki yöntemin birleştirilmesi sayesinde pek çok hatalı sonucun elenmesi sağlanmış ve başarı oranı artırılmıştır.

Bildirinin kalanı şu şekilde organize edilmiştir: Bölüm 2'de yüz bulma problemi ve bu alandaki zorluklar anlatılmıştır. Bölüm 3'de Schneiderman yüz belirleme algoritması ve Bölüm 4'de bu yöntemin video üzerindeki sonuçlarını geliştirmek için kullandığımız ten belirleme yöntemlerinin detayları verilmiştir. Bölüm 5'te önerilen birleştirme yöntemi açıklanmış ve Bölüm 6'da TRECVID 2003 video arşivi üzerinde elde edilen sonuçlar gösterilmiştir.

## 2. Yüz Bulma Problemi

Yüz tanıma problemindeki aşamaların en önemlisi ve bilgisayarla görü alanının en çok ilgilendiği konulardan biri, resimlerdeki yüz alanının doğru olarak çıkarılmasıdır. Bunun için çok çeşitli algoritmalar geliştirilmiştir [4]. Ancak, bu algoritmaların çoğu sadece kontrollü ortamlarda elde edilmiş, çözünürlüğün oldukça iyi olduğu küçük boyutlu veri arşivleri üzerinde test edilmiş ve bu veri setlerinde başarılı olarak gösterilmiştir.

Yakın tarihli bir çalışmada, gerçek hayattaki fotoğrafları daha iyi yansıtması amacıyla Yahoo haber kanalı üzerinden toplanan fotoğraflarda varolan yüz tanıma algoritmalarının başarısız olduğu gözlenmiştir [6]. Bu çalışmada verilen yüz tanıma oranları %10-17 aralığındadır. Bu sonuçlardan da anlaşılacağı üzere, gerçek veri arşivlerinde yüz tanıma çok daha zor bir problemdir. Buna neden olarak, resimlerdeki poz

<sup>1</sup> Bu çalışma TÜBİTAK Kariyer 104E065 ve TÜBİTAK 104E077 nolu projeler tarafından desteklenmiştir.

değişiklikleri, aydınlatma farklılıkları, ifade değişiklikleri, resimdeki örtmeler ve resimlerin çözünürlüğü gösterilebilir. Aynı etmenler yüz bulma yöntemlerinin başarısını da benzer şekilde etkilemektedir.

Videodaki yüz örneklerinin çeşitliliğini Şekil 1’de gösterilmiştir. Örneklerden de görüldüğü üzere, videodaki resimler çok farklı ortamlarda çekilmiş, farklı boyutlarda yüzleri içermektedir. Poz ve ifadeler çokça değişiklik göstermektedir. Yüzler çok küçük olabilmekte; önden ya da yandan değişik açılarda çekilebilmekte; sakal, bıyık, gözlük, şapka ya da eşarp ile bir kısmı kapatılabilmekte; irksal ten rengi farklılıkları gösterebilmekte; ya da ışığın farklı açılardan vurması nedeniyle gölgelenebilmektedir. Bütün bu etmenler, görüntülerde yer alan yüzlerin bulunmasını zorlaştırmaktadır. Bu nedenle video arşivleri, yüz bulma ve tanıma algoritmaları için çok zor bir uygulama ortamı oluşturmaktadır.

Şekil 1: Videodaki yüzlere örnekler



### 3. Schneiderman-Kanade Yüz Bulma Yöntemi

Literatürdeki en başarılı yüz bulma yöntemlerinden biri Schneiderman ve Kanade tarafından önerilen istatistiksel yüz bulma yöntemidir [5]. Bu sistem, çoklu sınıflandırıcıları bir arada kullanır. Bu sınıflandırıcıların her biri farklı bir yönelim alanını sabit pencere aralıkları kullanarak tarar. Her bir sınıflandırıcı bölgesel parçaların istatistiksel bilgilerini kullanır. Bölgesel parçalar, dalgacık katsayılarının bir sabit sayı alt kümesine dönüşümleri ile hesaplanır. Sınıflandırıcıları oluşturmak için yüz olan ve olmayan resim alanlarından örnekler kullanılarak olasılık dağılımları belirlenir, ve daha sonra sınıflandırıcılar hata oranlarını düşürmek amacıyla Adaboost kullanılarak öğretilir. Bu algoritmanın başarı oranı Kodak test kümesinde %80.4 olarak verilmiştir.

Bu çalışmada, bu yöntemin Mikolajczyk tarafından yapılan uygulaması temel alınmıştır [7]. Ancak, bu uygulama TRECVID2003 veri arşivi üzerinde çalıştırıldığında elde edilen kesinlik oranlarının çok fazla olmadığı ve sistemin çok fazla yanlış alanı yüz olarak tanımladığı gözlenmiştir. Bu nedenle bu yüz bulma metodunun geliştirilmesi amaçlanmış ve yüz bulma başarısını artırmak amacıyla bulunan alandaki ten rengi bilgisinden faydalanmak için ten bulma yöntemi geliştirilmiştir.

### 4. Ten Bulma Yöntemi

Yüz bulma yöntemlerinde yaygın olarak insan yüzü için çok ayırdedici olan ten renginden faydalanılmaktadır. Literatürde ten sınıflandırmak için pek çok yöntem denenmiştir [8]. Bu yöntemlerden Bayes histogram yöntemi en başarılı yöntem olarak belirtilse de, video üzerinde uyguladığımız ilk deneylerde, Gauss olasılık dağılımı yönteminin daha etkin sonuçlar verdiği gözlemlenmiştir. Bu nedenle, bu çalışmada, ten belirlemek için HSV renk düzleminde Gauss olasılık dağılımları kullanılarak alan genişletme tekniği uygulanmıştır. Literatürde HSV renk düzlemi ten bulma konusunda başarılı sonuçlar verdiği belirtilmektedir [8].

Ten alanının belirlenmesinde, önce bütün resimler HSV renk düzleminde çevrilmiş, üç ayrı video üzerinde 30’ar resimden örnek olarak alınan ten pikselleri kullanılarak Gauss dağılımının ortalama ve standart sapma değerleri bulunmuştur.

Daha sonra, piksellerin ten olma olasılıkları bulunan Gauss dağılımı baz alınarak hesaplanmıştır. Bu hesaplamada, pikselin Gauss dağılımına olan Mahalanobis uzaklığı göz önünde bulundurulmuştur. Bu yöntem ile, bulunan yüz alanındaki ten olma olasılığı belli bir eşik değerden yüksek olan pikseller ten olarak nitelendirilmiştir.

### 5. Yüz Bulma Yöntemlerinin Birleştirilmesi

Ten bulma yöntemleri çoğu zaman tek başlarına yetersiz kalmaktadır. Bunun nedeni, yüz olmayan ama ten rengine çok yakın bölgelerin de bu yöntemle yüz olarak çıkarılabilesidir. Örneğin, Şekil 1’de üst solda yer alan resimde hem yüz ve hem de manto ten rengi olarak sınıflanabilmekte, aynı şekilde alt ortadaki resimde duvarı yüzden ayırmak pek mümkün olamamaktadır.

Bu nedenle, Schneiderman-Kanade yöntemi, ten bulma yöntemi ile birleştirilmiş, böylece bu yöntemin yanlış olarak bulunduğu yüzler ten bulma metoduyla elenmiş, tek başına yetersiz kalacak ten bulma metodu diğer yöntemle birleştirilerek güçlendirilmiştir. Bu sayede her iki metodun da eksiklikleri en aza indirgenmiş ve başarı oranı artırılmıştır.

İki sistemi birleştirmek için izlenen yöntem şu şekildedir: Öncelikle görüntülerdeki yüzler Schneiderman-Kanade yöntemi kullanılarak bulunmakta, daha sonra bulunan yüz bölgelerindeki piksellerin ten renginde olma olasılığına bakılmaktadır. Yukarıda açıklandığı gibi, bu olasılık belli bir eşik değerinden yüksekse saklanmakta, değilse elenmektedir. Bu yöntemin detayları deneysel sonuçlar kısmında verilmiştir.

### 6. Deneysel Sonuçlar ve Tartışma

Bu çalışmada TRECVID 2003 veri tabanından 30’ar dakikalık üç video kullanılmıştır. Bu videoların biri CNN, diğer ikisi ABC haber videosudur. Videolar kesitlere (shot) bölünmüş, ve her bir kesitten tek bir görüntü (keyframe) kullanılmıştır. Toplam görüntü sayısı ilk videoda 322, ikinci videoda 346 ve üçüncü videoda 313’tür. İlk video 193 ayrırt edilebilir yüz alanı içerirken, ikinci video 249, üçüncü video ise 242 yüz içermektedir. Schneiderman-Kanade yüz bulma algoritması tarafından çıkarılan yüz alanları belli bir güven katsayısı ile gösterilmektedir. Bu güven katsayısına bağlı olarak, yüz belirleme algoritmasının bu üç video için doğruluk dereceleri Tablo 1’de gösterilmektedir.

Tablo 1: Schneiderman-Kanade yönteminin sonuçları

Video	Güven Katsayısı	Kesinlik Değeri	Geri Getirme Yüzdesi
Video1	-50	0.85	0.31
Video1	-70	0.51	0.40
Video1	-100	0.13	0.60
Video2	-50	0.68	0.30
Video2	-70	0.33	0.37
Video2	-100	0.05	0.44
Video3	-50	0.82	0.40
Video3	-70	0.42	0.45
Video3	-100	0.10	0.57

Tablo 1'deki sonuçlardan görülebileceği üzere, yüz bulma algoritması yüksek güven katsayısı değerlerinde yüksek kesinlik göstermekle birlikte pek çok yüzü tanımlanamamaktadır ve geri getirme yüzdesi düşüktür. Güven katsayısı -100 gibi düşük bir değer olarak alındığında ise sistemin geri getirme yüzdesi yükselse de, sistem çok fazla yanlış alanı yüz olarak adlandırmakta, bu durum da kesinlik değerini çok fazla azaltmaktadır.

Bu nedenle, hem varolan yüzlerin sistem tarafından kaçırılmaması ve bu sayede geri getirme yüzdesinin yüksek olması için, hem de kesinlik değerini makul ölçülerde tutabilmek amacıyla, güven katsayısı değeri -70 olarak alınmış ve yüz bulma algoritmasının bu değer ve üstünde kesinlikte değerlendirdiği yüzler bir sonraki aşamada işlenmek üzere kullanılmıştır. Şekil 2, 3 ve 4'de yüz bulma algoritmasının getirdiği doğru ve yanlış yüzler ile bulunamayan yüzlere örnekler verilmiştir. Görüldüğü üzere değişik yönlerde ve büyüklüklerdeki yüzler çoğu zaman bulunabilmekte, ancak bazen kötü aydınlatma ya da yüzü kapatan nesnelere dolayısıyla, bazense yüz açısının çok fazla olması nedeniyle bazı yüzler kaçırılmaktadır. Bununla beraber, yüz olmayan bazı bölgeler hatalı olarak yüz diye çıkarılabilmekte, ya da yüzün sadece belli bir bölgesi bulunabilmektedir.

Yüz bulma algoritması tarafından yanlış olarak betimlendirilen yüz alanlarını ayıklamak için, daha önce açıklandığı şekilde, bu alanların ten rengi bilgisi kullanılmıştır. Tablo 2'de gösterildiği üzere ten bulma yöntemi tek başına kullanıldığında çok düşük doğruluk değerleri vermektedir. Ancak, diğer yöntemi iyileştirme amaçlı kullanıldığında başarılı sonuçlar üretmiştir. Birleştirmede kullanılan ilk yöntem, yüz olarak tarif edilen alan içindeki ortalama ten rengi değerinin hesaplanmasıdır. Bu ortalama ten rengi değeri özellikle -50 güven katsayısı ile -70 güven katsayısı arasında artmış olan yanlış yüz belirlemelerini indirmek için kullanılmıştır. Bu aralıktaki yüz alanları için ortalama ten rengi belli bir eşik değerden yüksek olması durumunda, alan yüz olarak etiketlenmiştir. Bu çalışmanın sonucunda elenen yüzler Şekil 5'te, elenemeyenler Şekil 6'da gösterilmiştir.

Tablo 2: Ten bulma yönteminin sonuçları

Video	Kesinlik Değeri	Geri Getirme Yüzdesi
Video1	0.13	0.43
Video2	0.12	0.28
Video3	0.16	0.47

Şekil 2 : Yüz bulma algoritmasının bulduğu doğru yüz örnekleri



Şekil 3 : Yüz bulma algoritmasının bulduğu yanlış yüz örnekleri



Şekil 4 : Yüz bulma algoritmasının bulamadığı yüz örnekleri



Ortalama ten olma olasılığı değeri ile eleme yapıldıktan sonra geri kalan yüz örnekleri incelendiğinde, çoğunlukla küçük alanlarda hatalı belirlemeler olduğu göze çarpmıştır. Bu alandaki hataları düzeltmek için, ikinci bir eleme yöntemi olarak, ten olarak sınıflandırılmış piksel sayıları baz alınmıştır. Bu eleme güven katsayısı -20 ile -50 arasında olan yüzler için uygulanmıştır. Güven katsayısı -20'nin üzerinde gösterilen yüz alanları yüz olarak etiketlenirken, geri



