

Yüz Bulma Yöntemlerinin Haber Videoları İçin Sistemik Karşılaştırılması

Systematic Evaluation of Face Detection Algorithms on News Videos

Can Acar, Arda Atlas, Koray Çevik, İsa Ölmez, Mustafa Ünlü, Derya Özkan, Pınar Duygulu

Bilgisayar Mühendisliği Bölümü Bilkent Üniversitesi, Bilkent, Ankara

Özetçe

Haber videolarının etkin erişimi için, haberlerdeki en önemli öge olan kişilerin bulunması oldukça önemlidir. Literatürde çok çeşitli yüz bulma yöntemi yer almasına rağmen, kontrollü ortamlarda çok iyi sonuçlar veren bu yöntemler, haber videolarında verilerin çeşitliliği ve görüntülerin gürültülü olması nedeniyle çok iyi çalışmamakta ve oldukça hatalı sonuçlar vermektedir. Ayrıca her bir yöntem farklı özellikte ve sayıdaki yüzleri bulabilmektedir. Bu çalışmada var olan yöntemlerden en iyi verimin alınabilmesi amacıyla, yüz bulma yöntemlerinin sistemik bir değerlendirmesi yapılmıştır. Deneylerde, haber videolarından oluşan TRECVID 2006 veri kümesi kullanılmış ve dört farklı yüz bulma yöntemi değerlendirilmiştir.

Abstract

People are the most important subjects in news videos and for proper retrieval of people 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. In addition to that, with different face detection algorithms, the number and the type of the faces may differ. In this study, in order to get the best performance from existing methods, systematic evaluation of these methods is performed. In the experiments, news videos from TRECVID 2006 data set are used and for evaluation four different face detection methods are chosen.

1. Giriş

Gelişen teknoloji ile birlikte çok büyük resim ve video arşivleri ortaya çıkmış, bu arşivlere etkin ve hızlı bir şekilde erişmek büyük önem kazanmıştır [1]. Haber videolarından oluşan arşivler, sosyal ve kültürel etkisinin yanında, resim, video, ses ve metin gibi farklı birçok veriyi bir arada bulundurmalarından dolayı oldukça önemli bir veri kaynağıdır. Bu önemi dolayısıyla haber videoları, 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, yüzü tanıma problemi, yapay zeka ve bilgisayarla görüntü 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 tanıma 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. Ayrıca, farklı yöntemler farklı sayı ve özellikteki yüzleri bulmakta, hiçbir yöntem tek başına iyi bir sonuç üretememektedir. Sonraki aşamalarda yüz bulma yöntemlerinden en iyi verimin alınabilmesi için, kullanılacak yöntemin çok iyi değerlendirilmesi, eksik yanlarının anlaşılabilmesi gerekmektedir.

Bu çalışmada, yüz bulma yöntemlerinin özelliklerinin anlaşılabilmesi ve ileriki aşamalarda birleştirilebilmesi amacıyla farklı yöntemlerin detaylı bir değerlendirmesi yapılmıştır. Literatürde var olan çok sayıda yöntem arasından seçim yapılırken probleme uygunluk ve daha sık kullanım göz önünde bulundurularak dört farklı yöntem seçilmiştir. Bunlar en sık kullanılan yöntemlerden biri olan ten bulma yöntemi [5], literatürde var olan oldukça başarılı yöntemlerden biri olan Rowley, Baluja ve Kanade tarafından geliştirilmiş olan yöntem [6], haber video ve fotoğrafları üzerinde daha önce başka çalışmalarda da kullanılmış olan bir yöntem [7] ve erişiminin kolaylığı nedeniyle oldukça sık kullanılan OpenCV tarafından sağlanan yöntemdir [8].

Deneyler, TRECVID 2006 haber videoları üzerinde yapılmış ve sonuçlar oldukça geniş bir çerçeveden seçilen 1000 resim üzerinde sunulmuştur.



Şekil 1: Videodan yüz örnekleri

2. Yüz Bulma Problemi

Yakın tarihli bir çalışmada, gerçek hayattaki fotoğrafları daha iyi yansıtmaya amacıyla Yahoo haber kanalı üzerinden toplanan fotoğraflarda var olan yüz tanıma algoritmalarının başarısız olduğu gözlemlenmiştir [9]. 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 değişiklikleri, aydınlatma farklılıkları, ifade değişiklikleri, resimdeki düzeltmeler ve resimlerin çözünürlüğü gösterilebilir. Aynı etmenler yüz bulma yöntemlerinin başarısını da benzer şekilde etkilemektedir.

Videodaki yüz örneklerinin çeşitliliği 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 eşarpla bir kısmı kapatılabilmekte; ırksal 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.

3. Değerlendirilen Yöntemler

Bu çalışmada probleme uygunluk, başarı ve sık kullanım kriterleri göz önünde bulundurularak değerlendirilmek üzere aşağıdaki dört yöntem seçilmiştir.

3.1. Yöntem 1

Ten bulma tabanlı yöntemler yüz bulmada oldukça sık kullanılmaktadır. Başarı oranları diğer yöntemlere göre daha düşük olmasına rağmen, karşılaştırmada bir taban çizgisi oluşturabilmek ve ileriki çalışmalarda diğer yöntemlerle birleştirebilmesi için iyi ve kötü yanlarını görebilmek amacıyla seçilmiştir. Bu çalışmada [10]’da tanımlanan ten rengi tabanlı yüz bulma yöntemi kullanılmıştır. Bu çalışmada RGB renk uzayının ışık parlaklığına olan hassaslığı nedeniyle RGB renk değerleri YCbCr uzayına çevrilmiştir. Yapılan araştırmalarda ve veri arşivinde yapılan incelemeyle ten renginin Y bileşenine bağlı olmadığı, Cb ve Cr değerlerinin ise ten rengi için küçük bir aralıkta kümelendiği görülmüştür. Kullanılan yöntem bu verilerden yola çıkılarak, tek bir Gauss modeli kullanılarak geliştirilmiştir.



Şekil 2: Elle etiketlenmiş doğru yüz bölgeleri.

3.2. Yöntem 2

Yöntem 2 olarak tanımlayacağımız Mikolajczyk [7] tarafından geliştirilmiş yüz bulma yöntemi, yüzlerin pozisyonu hakkında bir kısıtlaması olmayan, karşıdan veya yandan bakan yüzleri iki adet detektörle bulabilen bir yöntemdir. Daha önce haber video ve fotoğrafları üzerindeki benzer başka çalışmalarda da kullanıldığı için tercih edilmiştir.

3.3. Yöntem 3

Yöntem 3 olarak tanımlayacağımız yöntem birçok önemli resim işleme ve bilgisayarlı görme algoritmalarını gerçekleştiren C++ sınıflarının bir koleksiyonu olan OpenCV’den seçilmiştir [8]. Yöntemde nesnelere için istatistiksel bir model (tanımlayıcı) eğitilir ve nesnelere bulmak için kullanılır. Öncelikli olarak yüz tanımlamak için kullanılmış olan bu yöntem, istatistiksel model için basit Haar tipi özellikler ve kademeli hızlandırılmış sınıflandırıcılar kullanılmaktadır.

3.4. Yöntem 4

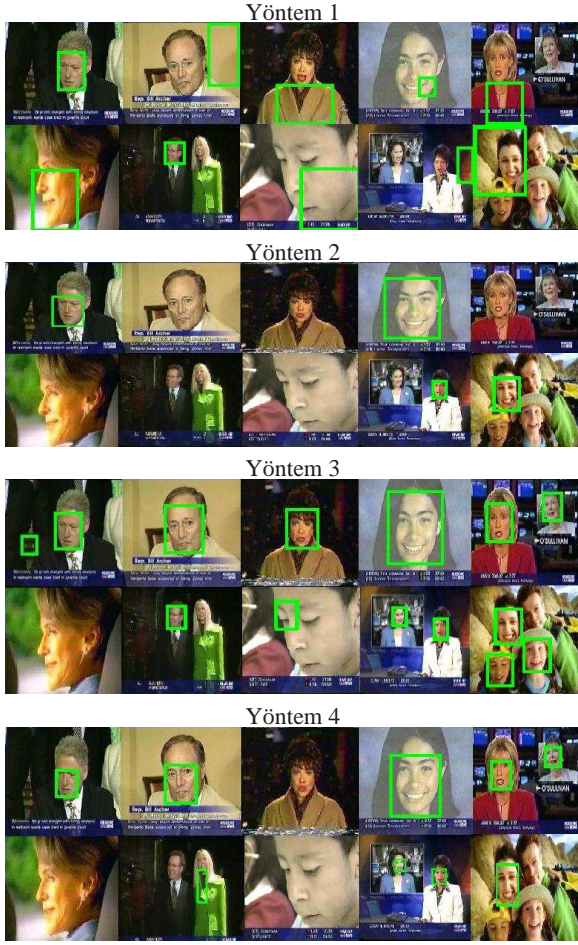
Yöntem 4 olarak tanımlayacağımız Rowley ve ekibi tarafından geliştirilmiş olan yöntem [6] iki aşamalı bir nöral ağ kullanılmaktadır. Literatürdeki başarılı yöntemlerden biri olduğu için seçilmiştir.

4. Değerlendirme

Bu yöntemlerin sistematik bir şekilde değerlendirilebilmesi amacıyla, TRECVID 2006 veri kümesi içerisinde farklı boyut, poz ve özellikte yüzler içeren 1000 video karesi seçilmiş ve yüz bölgeleri elle işaretlenmiştir. Şekil 2’de bu küme içerisinde 10 resim üzerinde elle etiketlenen bölgeler gösterilmektedir. Bu 1000 resim üzerinde dört yöntem ayrı ayrı çalıştırılmış ve her bir yöntem için başarı oranı, bulunan yüz bölgeleri ile doğru yüz bölgeleri karşılaştırılarak yapılmıştır. Ancak, bulunan ve işaretlenen yüz bölgeleri her zaman tam olarak üst üste gelmediği ve farklı büyüklükte olabildiği için değerlendirme yapılırken küçük de olsa bir çakışma varsa doğru sayılmış, değilse sonuç yanlış olarak yorumlanmıştır. Bu varsayım ile alınan sonuçlar Tablo 1’de özetlenmiştir. Sonuçlar yüz olan bir bölgenin doğru olarak işaretlenmesi (Doğru/Doğru), yüz olduğu halde yöntem tarafından bulunmaması (Doğru/Yanlış) ve yüz olmayan bir bölgede yöntem tarafından yüz bulunması (Yanlış/Doğru) şeklinde verilmiştir.

Görüldüğü üzere bu yöntemler sırası ile şu geri çağırma ve kesinlik değerlerini vermektedir: (0.539, 0.672), (0.264, 1), (0.551, 0.876), (0.454, 0.954). Bu sayılar, bu değerlendirme ile Yöntem 3’ün daha iyi olduğunu göstermektedir.

Yukarıda açıklandığı üzere bulunan ve gerçek yüz bölgeleri



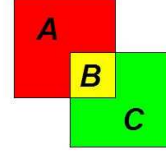
Şekil 3: Her bir yöntem tarafından çıkarılan doğru yüz bölgeleri.

arasındaki küçük bir çakışma dahi doğru olarak değerlendirilmektedir. Oysa çakışma yüzdesi de yöntemin başarısı hakkında oldukça önemli bir etkidir. Bu değerlendirmenin yapılabilmesi amacıyla Şekil 4'te gösterildiği üzere çakışan bölge B, doğru yüz bölgesinde geride kalan kısım A, bulunan yüz bölgesinde geride kalan kısım ise C olarak adlandırılmış ve $B/(A+B)$ ve $B/(B+C)$ oranları hesaplanmıştır. $B/(A+B)$ var olan bir yüzün ne kadar doğrulukla bulunabildiğini, $B/(B+C)$ oranı ise yöntemin ne kadar doğrulukla bir yüzü bulabildiğini göstermektedir. Şekil 5'de yöntemlerin bu oranlara göre başarı yüzdelerinin dağılımını göstermektedir. Bir yöntemin başarısı grafiklerin ne kadar sağa (büyük yüzdelere) yakın olduğuna bağlıdır.

1000'lik veri tabanı üstünde yapılan testlerde, kesişme oranı sadece 0.01 verilse bile, bulunan yüz oranlarının 0.55 in üstüne çıkmadığı görülmüştür. $B/[A+B]$ olarak bakıldığında Yöntem 3'ün iyi bir performans gösterdiğini görmekteyiz. Fakat Yöntem 3 ile bulunan alan, doğru yüz alanının dışına fazla çıktığı için, geri kalan alan oranı ($B/[B+C]$) olarak bakıldığında Yöntem 4 daha doğru sonuçlara ulaşmaktadır. Bunun nedeni Yöntem 4'te bulunan yüz alanının göz üstünden hemen kesilip, dudak sonrasında son bulmasıdır. Yöntem 2'de ise bunlardan daha farklı ama aynı zamanda daha kesin sonuç veren bir algoritma bulunmaktadır, ama bu durum birçok yüzün atlanmasına neden ol-

Tablo 1: Yüz olan bir bölgenin doğru olarak işaretlenmesi (Doğru/Doğru), yüz olduğu halde yöntem tarafından bulunmaması (Doğru/Yanlış) ve yüz olmayan bir bölgede yöntem tarafından yüz bulunması (Yanlış/Doğru)

	Doğru/Doğru	Doğru/Yanlış	Yanlış/Doğru
Yöntem 1	663	567	324
Yöntem 2	325	905	0
Yöntem 3	678	552	96
Yöntem 4	558	672	27



Şekil 4: Kesişim alanları gösterimi.

maktadır. Yöntem 1 ise yüzün duruş açısına bağlı olmadığı için yandan çekilmiş olan yüz resimlerinde daha iyi sonuç vermiştir. Fakat ten rengi bölümlenmesinde iyi bir algoritma kullanılmadığı için çıkan sonuçlar doğrultusunda bu algoritmanın sadece yardımcı amaçlı kullanılması gerektiği sonucuna ulaşılmıştır. Bu bulgular sonucunda ve grafikler doğrultusunda, Tablo-1'de de alınan sonuç ile uyumlu olarak, bu veri kümesi için Yöntem 3'ün en başarılı algoritma olduğunu görmekteyiz.

Şekil 3'de yöntemlerin yukarıda gösterilen 10 resim üzerindeki sonuçları gösterilmektedir. Görüldüğü üzere Yöntem 1, ten renginin yüz dışında başka bölgelerde geçmediği ve ten renginin net görüldüğü resimlerde başarılı olurken, Yöntem 2 yüzün açık bir şekilde görüldüğü ve aynı zamanda ön cepheden yüzü gören resimlerde başarılı olmaktadır. Yöntem 3 ise yine ön cepheden çekilmiş yüzlerde ve diğerlerine oranla daha küçük piksellere sahip yüzlerde başarılı olmaktadır, buna rağmen şekilde görüldüğü üzere bazı resimlerde fazladan yüz bulabilmektedir. Yöntem 4 önden çekilmiş yüz resimlerinde ve yüzlerin fazla yan yana gelmediği resimlerde başarılı olmaktadır.

5. Sonuç ve Değerlendirme

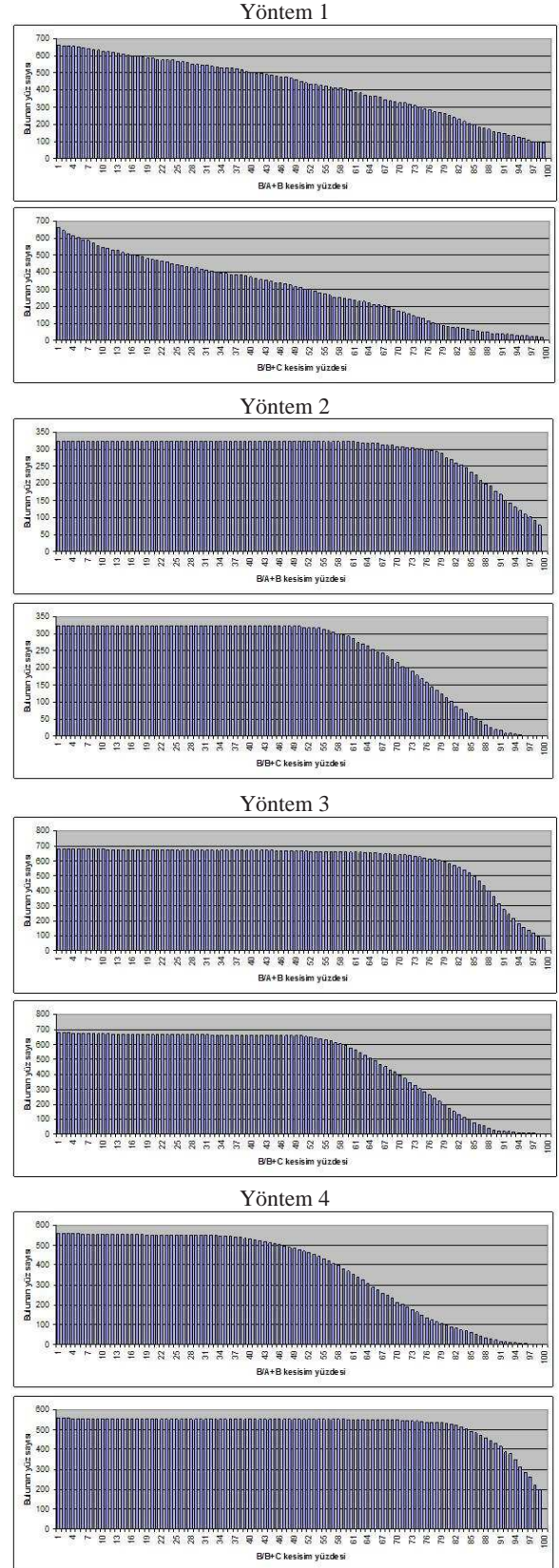
Bu çalışmada yüz bulma yöntemleri oldukça zor bir veri kümesi olan TRECVID 2006 haber videoları üzerinde karşılaştırılmıştır. En çok kullanılan yöntemlerden biri olan ten bulma yöntemi, farklı veri kümelerinde başarısını ispatlamış olan Rowley-Baluja-Kanade yöntemi, benzer çalışmalarda kullanılmış olan Mikolajczyk yöntemi ve ulaşılabilirliğinin ve kullanımının kolaylığı nedeniyle seçtiğimiz OpenCV'den seçilen dört yöntem karşılaştırılmış ve sonuçlar elle etiketlenmiş 1000 resim üzerinde değerlendirilmiştir. Bu çalışmanın amacı, yüz tanıyabilmek amacıyla en doğru yüz bölgelerinin bulunmasıdır. Elde edilen sonuçlar Yöntem 3'ün bu veri kümesinde en doğru sonuçları verdiğini göstermektedir. Ancak yöntemlerin artı ve eksi yönleri göz önüne alındığında yöntemlerin birleştirilmesinin de daha iyi sonuçlar verebilme ihtimali görülmüştür.

6. Teşekkür

Yardımlarından dolayı Muhammet Baştan'a teşekkür ederiz. Bu çalışma TÜBİTAK Kariyer 104E065 ve TÜBİTAK 104E077 no'lu projeler tarafından desteklenmiştir.

7. Kaynakça

- [1] C.G.M. Snoek and M. Worring, "Multimodal video indexing: A review of the state-of-the-art," *Multimedia Tools and Applications*, 2005, Accepted for publication.
- [2] "Trec video retrieval evaluation, <http://www-nlpir.nist.gov/projects/tv2005>," .
- [3] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: A literature survey," 2000.
- [4] Kriegman D. Yang M-H. and Ahuja N., "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.
- [5] "Skin segmentation using color pixel classification: Analysis and comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 148–154, 2005, Member-Son Lam Phung and Sr. Member-Abdesselam Bouzerdoum and Sr. Member-Douglas Chai.
- [6] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, 1998.
- [7] Krystian Mikolajczyk, *Face Detector*, Ph.D. thesis, INRIA Rhone-Alpes.
- [8] "Opencv: Open source computer vision library reference manual," .
- [9] T. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y.-W. Teh, E. Learned-Miller, and D.A. Forsyth, "Faces and names in the news," in *IEEE Conf. on Computer Vision and Pattern Recognition(CVPR)*, 2004.
- [10] Chang-Tsun Li Wen-Hsiang Lai, "Skin colour-based face detection in colour images," 2006, pp. 56–56.



Şekil 5: Her bir yöntem için başarı yüzdesi