

# Görü-tabanlı Gerçek-zamanlı Duygu Tanıma

## Vision-based Real-time Emotion Recognition

İsmail Arı, F. Onur Alsaran, Lale Akarun

Bilgisayar Mühendisliği Bölümü  
Boğaziçi Üniversitesi, İstanbul, Türkiye  
{ismailar,onur.alsaran,akarun}@boun.edu.tr

### Özetçe

Duygusal ifadelerin otomatik olarak tanınması, insan-bilgisayar etkileşimi, sözlü-sözsüz iletişim, veriye-dayalı yüz animasyonu, otistik çocukların yüzlerinin incelenmesi gibi birçok alanda çalışan araştırmacıların ilgisini çeken zorlu bir problemdir. Bu çalışmada duygu tanıma problemi irdelenmekte, video üstünde gerçek-zamanlı olarak çalışabilen ve şaşkınlık, kızgınlık, mutluluk, üzülmeye, korku, iğrenme ve yalın ifadeyi tanıyabilen yeni bir yöntem önerilmektedir. Önerilen yöntemde çok-çözünürlüklü aktif şekil modeline dayalı takipçi ile nirengi noktaları takip edilmekte ve ardından nokta konumlarından ve yüzün belli bölgelerindeki değişimlerden çıkarılan üst seviye öznelik vektörlerinin duygusal ifadelerle ait öznelik vektörlerine olan uzaklıkları kullanılarak sınıflandırma yapılmaktadır. Toplam 5 özne üzerinde yapılan deneylerde bahsedilen yedi sınıf için %75.23 başarı elde edilirken şaşkınlık, kızgınlık, mutluluk ve yalın ifadeden oluşan dört sınıf için %100 başarı elde sağlanmaktadır. Yüzün kısmi kapatılması durumunda da başarılı sonuçlar veren sistem, gerçek-zamanlı çalışan ve kişiye ve ortama uyulanabilen bir uygulama ile gerçekleştirilmiştir.

### Abstract

Automatic recognition of emotional states is a challenging problem which takes the attention of researchers studying in many areas such as human-computer interaction, verbal-nonverbal communication, data-driven facial animation and analysis of the faces of autistic children. In this work, emotion recognition problem is investigated; a novel approach is proposed that works in real-time and recognizes surprise, anger, happiness, sadness, fear, disgust and neutral expression. In the proposed system, facial landmarks are tracked using the tracker based on the multi-resolution active shape models. Afterwards, the classification is done using the differences among the feature vectors belonging to emotional expressions to the feature vector involving high level features extracted from the locations of the landmarks and the gradients of specific facial regions. In experiments conducted on 5 subjects, success rate of 75.23% is achieved for the seven classes mentioned whereas a success rate of 100% is found when four classes (surprise, anger, happiness and neutral expression) are used. The system gives promising results also on partially occluded faces. The implementation works in real-time and includes modules for person and environment adaptation.

### 1. Giriş

Yüz ifadelerinin incelenmesi psikolojik araştırmalar, sözlü ve sözsüz iletişim, insan-bilgisayar etkileşimi gibi birçok alanda önemli bir yer tutar. Yüzümüz duygularımızın penceresi olduğu için duygu tanıma ile yüz ifadesi tanıma ifadeleri farklı çalışmalarda birbirlerinin yerine kullanılabilir. Görü-tabanlı otomatik yüz ifadesi tanıma üstüne son yıllarda birçok yöntem geliştirilmiştir. *Fasel ve Luetin* ile *Pantic ve Rothkrantz* bu yöndeki araştırmaları [1, 2]'de ayrıntılı olarak incelemişlerdir. *Paleari v.d.*'nin önerdiği gibi görsel bilgiye ek olarak ses bilgisi de kullanılarak çok-kipli duygu tanıma yapılabilmektedir [3]. Öte yandan beyin sinyalleri de duygu tanımadaki kullanılabilmektedir. *Murugappan v.d.* EEG verisinin zaman-sıklık analizine dayalı yöntemleri ile mutluluk, üzüntü, şaşkınlık ve korku duygularını tanımlamışlardır [4].

Bu çalışmanın odağı duygu tanımadaki görsel bilgiyi kullanmaktır. Yüz imgelerinden veya videolarından etkili bir betimleyici çıkarabilmek tüm yüz ifadesi araştırmalarının temelini oluşturur. Betimleyicilerin başarısı ifade ettikleri yerel konumların gürbüz bir şekilde belirlenmesine dayalıdır. *Arı ve Akarun*, çalışmalarında çok çözünürlüklü ve çok-pozlu aktif şekil modeline dayalı yüz nirengi noktası takipçisi geliştirmiş ve takip edilen noktaların zaman içindeki yörengi bilgisini kullanan Saklı Markov Modeli'ne dayalı sınıflandırıcı ile hem kafa hareketi hem de yüz ifadesi sınıflandırması yapmışlardır [5]. *Akayın ve Sankur*, betimleyici olarak yüze ait 17 nirengi noktasının takibiyle elde edilen yörengelerden hesaplanan bağımsız bileşenler analizi değerlerini ve yüzlerin arka arkaya hizalanarak oluşturdukları zaman-uzaysal kübün 3B ayırık kosinüs dönüşümü (AKD) değerlerini kullanmışlardır [6]. Farklı sınıflandırıcılar ile değerlendirdiklerinde, en iyi sonucun 3B AKD kullanıldığında bulunduğunu raporlamışlardır. *Kumano v.d.* her ifade için değişken şiddet şablonlarına (variable intensity templates) dayalı bir model eğitimi önermiş ve pozdan bağımsız olarak ifade tanınması yapmışlardır [7]. *Sebe v.d.*, çalışmalarında yüz ifadelerini Bayes ağları, destek vektör makineleri ve karar ağaçları ile incelemişler ve kullandıkları veritabanını araştırmacılara sunmuşlardır [8]. *Littlewort v.d.* ise otomatik yüz ifadesi tanıma problemi için Adaboost, destek vektör makineleri ve doğrusal ayraç analizi yöntemlerini sistematik olarak uygulamayı önermişlerdir [9]. *Shan v.d.*, istatistiksel bir model olan yerel ikilik örüntülerini (local binary patterns) dikkate almış, çıkarılan öznelikleri farklı yapay öğrenme teknikleri ile sınıflamada kullanmış ve destek vektör makinesinin sınıflandırıcılar arasında en iyi sonucu verdiğini raporlamışlardır [10]. *Busso v.d.*, nokta

takibi için ticari bir yazılım kullanmış, yüzü beş ilgi alanına bölmüş, her parça içine düşen nokta konumlarından temel bileşen analizi (TBA) katsayıları hesaplamış ve bu değerlerden elde edilen öznitelik vektörünü en yakın 3 komşu ile sınıflamışlardır [11].

Önceki çalışmalardan farklı bir yol izleyen çalışmamız üç adımdan oluşmaktadır: İlk adım, çok-çözünürlüklü aktif şekil modeli tabanlı takipçi ile yüze ait nirengi noktalarının tespitidir. İkinci adımda, takip edilen nokta konumlarının yardımıyla yüzün belli bölgelerindeki yerel değişimler (alın kırışıklığı, kaşların arasının kırışıklığı, kaşların gözlere mesafesi, yanakların kırışıklıkları, ağzın yatay ve dikey açıklıkları) hesaplanmakta ve ifadeye ait öznitelik vektörleri elde edilmektedir. Son olarak ise, elde edilen öznitelik vektörünün eğitim kümesindeki uzaklığına dayalı sınıflandırıcı ile yüz ifadesi belirlenmektedir. Bu çalışmanın katkısı, altı evrensel yüz ifadesi (şaşkınlık, kızgınlık, mutluluk, üzülmeye, korkma, iğrenme) ve yalın ifadenin gerçek-zamanlı olarak tanınabilmesidir. Ayrıca, yöntemin kısmi kapatma durumlarında da başarılı bir şekilde çalışıldığı gösterilmiştir.

Bildiri nirengi noktası takibinin anlatılması ile başlamakta ve yöntemin diğer parçalarını oluşturan öznitelik çıkarımı ve duygu sınıflandırma kısımları ile devam etmektedir (Kısım 2,3 ve 4). Gerçekleştirilen deneyler ve elde edilen sonuçlar, Kısım 5'te verilmektedir. Kısım 6 vargılarımızı özetlemektedir.

## 2. Nirengi Noktası Takibi

Cootes v.d. tarafından bulunan Aktif Şekil Modelleri (AŞM) nirengi noktası tespiti ve takibi için kullanılan en yaygın tekniklerdendir [12]. Bu yaklaşımda, işaretlenmiş imge kümesi kullanılarak AŞM eğitilir. Bir imgedeki nirengi noktalarını bulmak için öncelikle bütünsel bir yüz bulucusu (örneğin Viola-Jones) ile yüzün konumu belirlenir. Bu konuma göre hizalanan ortalama yüz şekli aramanın başlangıç konumunu oluşturur. Ardından şekil yakınsayana kadar alttaki iki aşama tekrarlanır:

- Her nokta için, o noktanın komşuluğundaki imge dokusunun gradyanını kullanarak şablon eşleştirmeyle modele en iyi oturan konum belirlenir.
- Bulunan noktaların, eğitim kümesindeki nokta konumlarından Temel Bileşenler Analizi (TBA) ile elde edilmiş şekil öz-uzayına izdüşümü alınır. Ve şekil, güven aralığına düşecek biçimde kısıtlanır.

Birinci aşamadaki bireysel şablon eşleştiriciler güvenilir konumlardan uzaklaşabilir ve elde edilen şekil, yüze benzemeyebilirken, ikinci aşamada kullanılan bütünsel yöntem, birbirinden bağımsız davranan zayıf modelleri genel bir model ile kısıtlayarak güçlendirmiş ve şekli eğitim kümesindeki şekillere benzetmiş olur. Sonucun yerel en iyilerde takılıp kalmaması için çok-çözünürlüklü yaklaşım kullanılmaktadır. Arama, imge piramidinde en düşük çözünürlüklü seviyeden başlar, kabadan ayrıntılıya doğru devam eder. Videoda takip yapılırken aramanın başlangıç konumu ortalama şekil yerine bir önceki çerçevede bulunan şekil olarak alınır, şeklin bulunamadığı durumda ise aramaya yine önce yüz bulunarak başlanır.

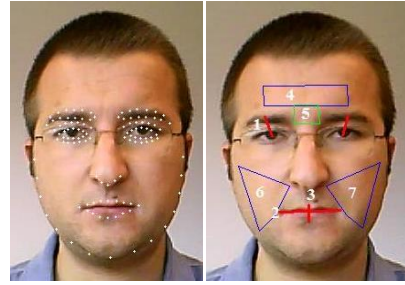
AŞM'nin kişiye özel modelle eğitildiğinde daha iyi sonuçlar verdiği bilinmektedir [5]. Sunduğumuz bu çalışmada farklı öznelerin yüzlerini içeren genel bir modelin yanı sıra öznelere ait kişisel modeller de kullanılmaktadır. AŞM takibi

We'i'nin geliştirdiği *asmlibrary* kütüphanesi kullanılarak geliştirilmiştir [13].

## 3. Öznitelik Çıkarımı

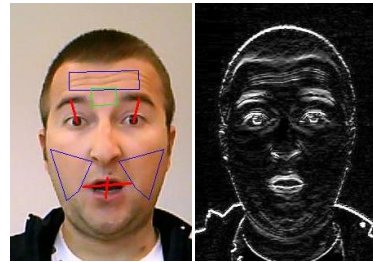
AŞM-tabanlı takipçi Şekil 1'de solda görülen 116 nirengi noktasını takip etmektedir. Takipçi her ne kadar kaş, göz, çene ve burun noktalarında gürbüz çalışsa da esnek olarak hareket edebilen dudak noktalarını tam tamına takip edemediği için bu noktaların konumlarını doğrudan öznitelik olarak kullanmak mantıklı görünmemektedir. Bu olgunun sebebi AŞM'nin nirengi noktalarının konumlarını bütünsel olarak modellemesi ve dudak noktalarının konumlarındaki küçük değişimlerin TBA ile yapılan şekil kısıtlamasında yitirilmesidir. Bunun yanı sıra dudak kenarlarındaki şiddet farkı diğer yüz bileşenlerindeki kadar ayırtedici değildir. Dolayısıyla, takip edilen nirengi noktalarının konumlarını doğrudan kullanmak yerine, bu konumlara göre yüzde belli uzunluklar ve bölgeler bulunup Şekil 1'de sağda görülen ve alta açıklanan üst düzey öznitelikler elde edilmektedir.

- Göz ortalarının kaş ortalarına uzaklığı
- Ağız genişliği
- Ağız yüksekliği
- Alın bölgesindeki dikey kenar etkinliği
- Alın çatisindeki yatay kenar etkinliği
- Sağ yanaktaki yatay/dikey kenar etkinlikleri toplamı
- Sol yanaktaki yatay/dikey kenar etkinlikleri toplamı



Şekil 1: Yüz nirengi noktaları (solda) ve öznitelik çıkarımında kullanılan ilgi bölgeleri (sağda)

İlk üç öznitelik ilgili nirengi noktalarının birbirlerine olan Öklid uzaklıklarından elde edilir. Diğer öznitelikler için, imge Gauss çekirdeği ile filtrelenerek yumuşatılır, yatay ve dikey Sobel çekirdeği ile ayrı ayrı süzgeçlenerek kenar etkinlikleri bulunur. Ardından her bölge içine düşen ortalama mutlak değer hesaplanır. Şekil 2'de şaşkınlık ifadesi durumunda dikey kenar etkinliği görülmektedir. Örneğin, alındaki mavi kutunun içine düşen ortalama dikey kenar etkinlik değeri dördüncü öznitelik oluşturur.



Şekil 2: Şaşkınlık durumundan bir görüntü (solda) ve ilgili dikey kenar etkinliği (sağda)

Öznitelik seçimindeki motivasyon için Tablo 1’de verilen duygusal ifadelerle ait hareket ipuçlarına bakılabilir. Bu ipuçları, Ekman ve Friesen’in yüz ifadelerini inceledikleri öncü çalışmalarına dayanmaktadır [14].

Tablo 1: Duygusal ifadeler ve açıklamaları

DUYGUSAL İFADE	HAREKET İPUÇLARI
<b>Şaşkınlık/Heyecan</b>	Kaşların yükselmesi, ağız açılması, çenenin hafifçe alçalması
<b>Kızgınlık</b>	Kaşların çatılması, dudakların kısılması, gözlerin firlaması
<b>Mutluluk</b>	Ağız köşelerinin yükselmesi ve alçalması
<b>Üzüntü</b>	Ağız kenarlarının alçalması, kaşların iç kısmının alçalması
<b>Korku</b>	Kaşların yükselmesi, gözlerin açılması, ağızın hafifçe açılması
<b>İğrenme</b>	Üst dudak kalkması, burun köprüsünün kırışması, yanakların çekilmesi

#### 4. Duygu Sınıflandırma

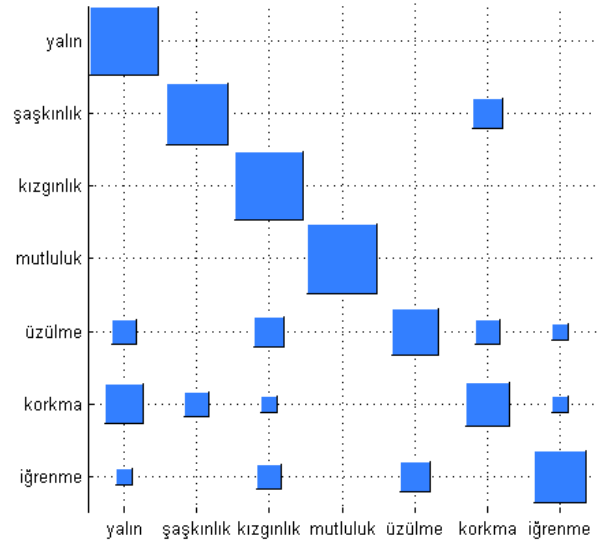
Yüz ifadeleri kişiye ve ortama (örneğin aydınlanmaya) göre değişim gösterdiği için belirli bir kişi ve ortam için kolaylıkla ayarlanabilecek bir düzenek amaçlanmıştır. Kullanılan düzenekte öznelere yalnız ifadeyle başlayarak 2 saniye kadar beklemektedir. Bu süreçte elde edilen özniteliklerden ortalama değer hesaplanır ve bunu izleyen tüm çerçevelerde, sistemi çevre koşullarından bağımsız hale getirmek amacıyla öznitelikler ortalamaya bölünerek normalleştirilir. Sistemi kullanacak kişi her ifadeyi  $T$  kere tekrar ederek toplamda  $NT$  örnek kaydetmektedir ( $N = \text{sınıf sayısı}$ ). Test aşamasında özne yine yalnız ifade ile başlayarak sistemi çevre koşullarından etkilenmemesi için ilk konuma hazırlar. İzleyen her çerçevedeki yüz için öznitelik vektörü hesaplanır. Öznitelik vektöründeki elemanların her sınıfa olan ortalama uzaklık değerleri olan  $\mathbf{d}_i$ ,  $i = 1, \dots, N$  vektörleri bulunur. Uzaklık değerleri farklılığı temsil ederken,  $\mathbf{s}_i = e^{-\mathbf{d}_i}$  değerleri benzerliği temsil eder ve benzerlik metriği olarak kullanılır. Son olarak  $\mathbf{s}_i$  vektörü, elemanları toplamı 1 edecek şekilde normalleştirilir. Bu değerler, sınıflara ait olabilirlik değerleri olarak düşünülebilir. Bu yöntem en yakın komşu (NN) ve kNN sınıflandırıcılara göre üstündür çünkü mutlak bir sınıflandırma yapmak yerine yumuşak atama (soft assignment) kullanmaktadır.

##### 4.1. Sistemin yeni öznelere için genişletilmesi

AŞM’nin yeni bir özne için eğitilmesi gerektiğinde, o kişiden fotoğraflar alınmakta ve genel model ile bu fotoğraflarda otomatik olarak nirengi noktaları işaretlenmektedir. İşaretlenen noktalar Pinotator [15] ile hızlıca ince ayara tabi tutulduktan sonra kişiye özel AŞM oluşturulmaktadır. Önceki bölümde bahsedildiği üzere, sınıflandırıcının yeni kişi ve ortam için eğitilmesi kolaylıkla yapılabilmekte ve sistem yeni kişilerin kullanımına açık olabilmektedir.

## 5. Sonuçlar

Sistemin performansını ölçmek amacıyla 5 öznenen (4 erkek, 1 kadın) 7 sınıfa ait (yalın, şaşkınlık, kızgınlık, mutluluk, üzülmeye, korkma, iğrenme) 3’er tekrar alınmıştır. Sınıflandırma başarıları %75.24 olarak bulunmuş ve Şekil 3’de görülen hata matrisi elde edilmiştir. Görüldüğü üzere yalnız ifade, kızgınlık ve mutluluk başka sınıflarla karıştırılmadan saptanabilmektedir. Hataların önemli bir kısmını korkma ve şaşkınlığın birbirleri ile karıştığı durumlar oluşturmaktadır. Bunun temel sebebi iki duyguda da gözlerin kaşlarını kaldırması ve benzer yüz ifadeleri takınmasıdır. Üzülmeye ve iğrenme duyguları da diğer duygularla karıştırılabilmektedir. Kişi duygusal ifadeleri ayırt edici biçimde yaptığında sistem tam başarı ile çalışabilmektedir. Örneğin Tablo 2’de verilen yüz ifadeler incelendiğinde sistemin 7 sınıf için bir öznenen %100 başarı ile çalıştığı gözlemlenmektedir.



Şekil 3: 7 sınıf için hata matrisi

Bu deneye ek olarak öznelere daha rahat yaptıklarını söyledikleri ilk 4 sınıfla benzer bir test gerçekleştirilmiştir. Bu deneyde genel başarı %100 olarak gözlemlenmiştir.

Başarının tam olmasının nedeni, bu sınıfların birbirinden ayrışır yüz ifadelerine denk gelmesi ve deneklerin bu ifadeleri rahat becerebilmesidir. Birbirleri ile karıştırılan duygusal ifadeler incelendiğinde kişinin bu duyguları benzer yüz ifadeleri ile yaptığı gözlemlenmektedir.

Tablo 2: Sınıflandırma başarıları

	Başarı (%)	
	4 sınıf	7 sınıf
<b>Özne #1</b>	100	100
<b>Özne #2</b>	100	71.43
<b>Özne #3</b>	100	61.90
<b>Özne #4</b>	100	66.67
<b>Özne #5</b>	100	76.19
<b>Genel</b>	100	75.23

Veri tabanındaki öznelerle ait örnek duygu tanıma sonuçları Şekil 4’de görülmektedir. İlgili duygusal ifadelerin olasılıkları arayüzde çizilmiştir. Önerilen yöntem 3GB RAM’e sahip core2duo 1.2 GHz işlemcili bir dizüstünde 320×240 çözünürlükte 30 çerçeve/saniye hızla çalışabilmekte ve dâhili web kamerası kullanılabilir. En altta görüldüğü üzere, önerilen yöntem yüzün kısmen kapatıldığı durumlarda da umut verici sonuçlar vermektedir.



Şekil 4: Duygu tanıma sonuçları (yalın, şaşkınlık, kızgınlık, mutluluk, üzülmeye, korkma, iğrenme) ve kısmi kapatma durumundaki sonuçlar (son iki resim)

## 6. Vargılar

Bu çalışmada yüz nirengi noktalarının takibine dayalı gerçek-zamanlı bir duygusal ifade tanıma yöntemi önerilmiştir. Nirengi noktası takipçisinin kesin sonuç vermediği durumda dahi etkili bir duygusal ifade betimleyici olan öznitelikler sunulmuştur. Yalın, şaşkınlık, kızgınlık, mutluluk, üzülmeye, korkma ve iğrenme durumlarındaki yüz ifadeleri sınıflandırılmış ve %75.23 genel sınıflandırma başarısı elde edilmiştir. Karıştırılan duygular özellikle şaşırma-korkma, üzülmeye-iğrenme-korkma grupları olarak gözlemlenmiş ve deneklerin daha rahat yaptıklarını ifade ettikleri ilk dört sınıfa ait başarı oranı %100 olarak bulunmuştur. Kişiyi ve ortama özel olarak kolayca ayarlanabilen yöntemin, kısmi kapanmalarda da başarılı sonuçlar verdiği gösterilmiştir. Sunulan özniteliklere ilave öznitelikler eklenerek sistemin ifade dağarcığının genişletilmesinin başarıyı arttıracığı öngörülmektedir.

## 7. Teşekkür

Bu çalışma, 108E161 numaralı proje kapsamında TÜBİTAK tarafından desteklenmiştir. Katkılarından dolayı Yunus Açıköz, Selcan Güner, Barış Evrim Demiröz ve Serhan Daniş’e teşekkür ederiz.

## 8. Kaynakça

- [1] B. Fasel ve J. Luetin, “Automatic facial expression analysis: a survey,” *Pattern Recognition*, 36:259-275, 2003.
- [2] M. Pantic ve L.J.M. Rothkrantz, “Automatic analysis of facial expressions: the state of the art,” *IEEE Trans. on*

*Pattern Analysis and Machine Intelligence*, 22:1424-1445, 2000.

- [3] M. Paleari, R. Chellali, ve B. Huet, “Features for multimodal emotion recognition: An extensive study,” *IEEE Conf. on Cyber. and Intelligent Sys.*, s.90-95, 2010.
- [4] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, D. Hazry and I. Zunaidi, “Time-frequency analysis of EEG signals for human emotion detection,” *4th Kuala Lumpur Int’l Conference on Biomedical Engineering*, 2008.
- [5] İ. Arı ve L. Akarun, “Yüz Özniteliklerinin Takibi ve İşaret Dili için İfade Tanıma,” *IEEE Sinyal İşleme ve İletişim Uygulamaları Konferansı*, 2009.
- [6] H. Ç. Akakin ve B. Sankur, “Spatiotemporal Features for Effective Facial Expression Recognition,” *IEEE 11th European Conf. on Computer Vision, Workshop on Sign Gesture Activity*, 2010.
- [7] S. Kumano, K. Otsuka, J. Yamato, E. Maeda, ve Y. Sato, “Pose-Invariant Facial Expression Recognition Using Variable-Intensity Templates,” *International Journal of Computer Vision*, c. 83, sf. 178-194, Kas. 2008.
- [8] N. Sebe, M. S. Lew, Y. Sun, I. Cohen, T. Gevers, ve T. S. Huang, “Authentic facial expression analysis,” *Image and Vision Computing*, c. 25, sf. 1856-1863, 2007.
- [9] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, ve J. Movellan, “Dynamics of facial expression extracted automatically from video,” *Image and Vision Computing*, c. 24, sf. 615-625, 2006.
- [10] C. Shan, S. Gong, ve P.W. McOwan, “Facial expression recognition based on Local Binary Patterns: A comprehensive study,” *Image and Vision Computing*, c. 27, sf. 803-816, May. 2009.
- [11] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, ve S. Narayanan, “Analysis of emotion recognition using facial expressions, speech and multimodal information,” *Proceedings of the 6th international conference on Multimodal interfaces - ICMI '04*, New York, 2004.
- [12] T. F. Cootes, C. J. Taylor, D. H. Cooper ve J. Graham, others, “Active shape models-their training and application,” *Computer vision and image understanding*, c. 61, sf. 38-59, 1995.
- [13] Y. Wei, “Research on Facial Expression Recognition and Synthesis,” *Y.L. Tezi*, Department of Computer Science and Technology, Nanjing University, 2009. <http://code.google.com/p/asmlibrary/>
- [14] P. Ekman ve W. Friesen, “Facial Action Coding System”, *Consulting Psychologists Press*, 1978.
- [15] İ. Arı ve Y. Açıköz, “Pinotator ile Hızlı İmge İşaretleme”, *IEEE Sinyal İşleme ve İletişim Uygulamaları Konferansı*, 2011. [gönderildi]