

# Seyrek Öğrenme ile Yüz İşaret Yaması Bazlı Eylem Birimi Saptama

## Facial Landmark Patch Based Action Unit Detection Using Sparse Learning

Duygu Çakır, Nafiz Arıca  
Mühendislik ve Doğa Bilimleri Fakültesi, Bahçeşehir Üniversitesi, İstanbul, Türkiye  
{duygu.cakir, [nafiz.arica](mailto:nafiz.arica@eng.bau.edu.tr)}@eng.bau.edu.tr

### Öz

Yüz Eylem Kodlama Sistemi, bir kişinin hatta bir hayvanın duygusal dışı vurum ya da acı gibi yüz ifadelerini tanımlamada en çok kullanılan ve en yaygın olarak kabul gören standarttır. Bu sistemde yüzlerce Eylem Ünitesi kombinasyonu kullanılarak yüzdeki kas hareketleri tanımlanmaktadır. Bu çalışmada yüz eylem ünitelerini yakalamak amacıyla yüz işaret noktalarından elde edilen imge yamalarıyla seyrek öğrenme tabanlı bir yöntem önerilmektedir. Önerilen yöntem bire karşı hepsi yaklaşımıyla tüm eylem üniteleri için ayrı ayrı en ayırt edici yüz işaret noktalarını çıkarmaktadır. CK+ veri kümesi üzerinde yapılan deneylerde, önerilen yöntemin son yıllarda yapılan yama bazlı çalışmaların çoğundan daha başarılı sonuç verdiği gözlenmiştir.

**Anahtar Kelimeler** — eylem ünitesi saptama, seyrek öğrenme.

### Abstract

Facial Action Coding System is the most used and widely accepted standard to describe a human's, or even an animal's facial actions such as emotional expressions and pain. The system uses hundreds of Action Unit combinations to define the facial muscle movements. In this study, we propose a sparse learning based method to detect facial action units from the facial landmark patches. The proposed method uses a one-vs-all approach to learn the most representative facial landmarks for each action unit.

*The experiments on the CK+ dataset show that the proposed method leads to better results in AU detection when compared to the recent state-of-the-art studies.*

**Keywords** — action unit detection, sparse learning.

### 1 Giriş

İnsan yüzü tarafından oluşturulan duygu gibi ifadeler yüzdeki kasların kasılması ile tanımlanır. Yüz Eylem Kodlama Sistemi (Facial Action Coding System – FACS) [1]-[3], yüzdeki kas eylemlerinin tanımlanmasında en çok kullanılan ve en yaygın kabul gören standarttır. Taksonomi, yüzdeki görsel olarak algılanabilen tüm kas hareketlerinin Eylem Üniteleri ( Action Unit - AU) tarafından tanımlanmasını içerir. Yüzlerce AU'nun kombinasyonu sonucu ortaya çıkan ifadeler, duygu saptama ve tanıma, acı seviyesi ölçümü, depresyon analizi gibi çalışmalarda [4]-[6] kullanılmaktadır. AU saptama aynı zamanda hayvan araştırmaları [7] ve robotik [8] gibi alanlarda da kullanılır. İnsan gözü yüzdeki çok ufak değişiklikleri bile saptayabilecek düzeyde iken, bilgisayarla görü çalışmaları, her kişideki AU farklı yoğunlukta ifade edildiği ve bu AU'ların yüzlerce kombinasyonu olduğu için henüz AU'ları saptama ve yorumlama konusunda yeterli başarıya ulaşamamıştır.

İlk AU saptama araştırmaları, yüzü bir bütün olarak ya da alt ve üst yarılarını ayrı ayrı inceleyen çalışmalarla başlamıştır [9,10]. Son yıllarda yapılan araştırmalara bakıldığında, yüzün tamamına ya da bir yarısına odaklanmak yerine yüz parçalarından alınan yamalar (patch) üzerinde çalışmak, başarı performansını yükseltmektedir [11]-[14]. Bu çalışmalarda kullanılan yamalar, yüzün tamamının dikey ve yatayda eşit parçalara bölünerek elde

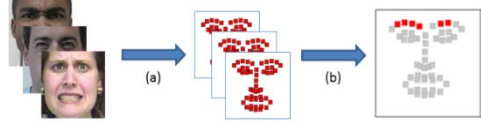
Gönderim ve kabul tarihi : 07.12.2016 - 01.07.2017

edildiği gibi, yüzdeki işaret noktalarının etrafından eşit büyüklükte parçalar alınarak da elde edilmektedir. Yama kullanılmaktaki en önemli amaç, sınıflandırmada etkisiz ya da kötü etkili / gürültülü yamaları ortadan kaldırarak sınıflandırıcıya etkisi en fazla olan tanımlayıcı yamaları bulmak ve onlar üzerinde çalışmaktır. Bunun için de yapılacak en iyi yöntem, yüz imgesini eşit parçalara bölerek izgara şeklinde yamalarla betimlemek yerine her bir AU için ayrı kas grubu hareketine maruz kalan yüz bölgeleri üzerinde çalışarak AU'ları saptamaktır.

Bu çalışmada, AU'ları oluşturan kas grubu hareketlerinin etkilediği yüz bölgelerinin otomatik olarak belirlenmesi amacıyla yüz işaret noktaları kullanılmakta ve her AU için en önemli işaret noktaları seyrek öğrenme yaklaşımıyla belirlenmektedir. Dolayısıyla bir AU için aktif yüz işaret noktaları manuel olarak seçilmek yerine öğrenme algoritmasının yapısı içinde otomatik olarak bulunmaktadır. İşaret noktaları etrafından eşit büyüklükte alınan yamalardan çıkarılan öznelik vektörleri kullanılarak, seyrek öğrenme algoritması içinde bire karşı hepsi yaklaşımıyla AU saptayıcılar geliştirilmektedir. Yüzün tamamından alınan işaret noktaları yerine aktif işaret noktalarının seçilmesi, AU saptama başarısını olumlu yönde etkilemektedir. CK+ [18] veri kümesi üzerinde yapılan deneylerde yüz işaret noktalarının betimlenmesinde en uygun yama boyutu incelenmiş ve önerilen yöntemin son yıllarda yapılan yama bazlı çalışmalardan daha başarılı sonuç verdiği gözlenmiştir. Çalışmanın ikinci bölümünde önerilen yöntemin kuramsal ayrıntıları anlatılmakta, üçüncü bölümde de deney kurulumu ve sonuçları verilmektedir. Son bölümde sonuçlar değerlendirilmekte ve gelecek çalışmalara nasıl yön verebileceği yorumlanmaktadır.

## 2 Kuramsal Altyapı

Yüz ifadeleri, yüzde bulunan kas gruplarının farklı ölçeklerdeki hareketleri sonucu oluşur. Her kas grubu yüzün değişik bölgesini harekete geçirdiği için söz konusu hareketlerin yüzde oluşturduğu farklı görsel değişimler Eylem Birimleri (AU) olarak gruplandırılmıştır. Dolayısıyla yüzdeki ifadelerin alfabeti olarak değerlendirilebilecek AU'ların tespit edilmesi herhangi bir yüz ifadesinin açıklanması için önemlidir.



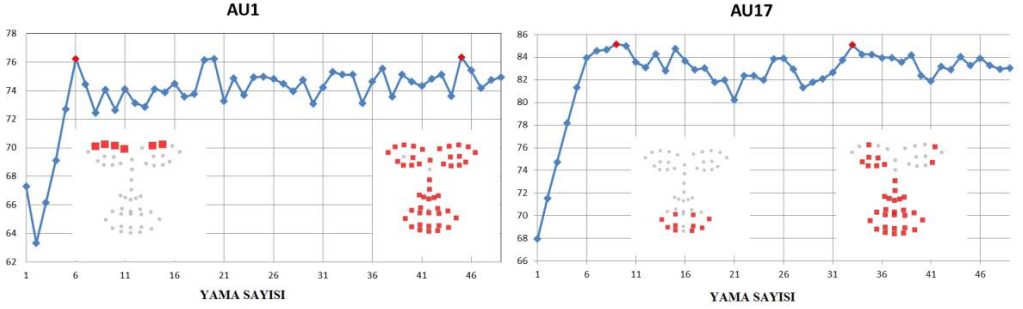
**Şekil 1:** AU1 (İç kaşı kaldırma) için örnek akış diyagramı. (a) AU'ya ait imge dizisi örneklerinin yüz işaret noktalarının etrafından 11x11 boyutlarında yamalar çıkarılır. (b) Seyrek öğrenme yöntemi ile ilgili AU için en tanımlayıcı işaret noktalarına ait yamalar belirlenir.

AU'ların tespit edilmesinde ilk yaklaşım yüz imgesinin bir bütün ya da yarım olarak ele alınıp tamamının betimlenmesi ve elde edilen öznelik vektörleriyle AU'lerin sınıflandırılmasıdır [9, 15]. Her AU'nun görsel değişime neden olduğu yüz bölgesinin manuel olarak belirlendiği ve sadece bu bölgeler kullanılarak ilgili AU'ların tespit edildiği çalışmalar da mevcuttur [11, 12]. Diğer bir yaklaşım ise yüzdeki işaret noktalarını kullanılmasıdır [10]. Son zamanlarda öne çıkan yöntemlerde [12, 14, 16, 17], çoklu görev seyrek öğrenme yaklaşımıyla AU'ların sınıflandırılmasında kullanılacak yüz imgesi bölgeleri eşit aralıklı yüz yamaları üzerinden belirlenmektedir.

Bu çalışmada, AU'lar yüzün işaret noktaları etrafında oluşan değişimlerle gözlemlendiğinden, AU'ların betimlenmesi yüz işaret noktaları ile gerçekleştirilmektedir. Ayrıca seyrek öğrenme yöntemiyle her AU için aktif olan yüz işaret noktaları belirlenmektedir. Yöntem akış diagramı Şekil (1)'de gösterilmektedir.

Bu amaçla, yüzdeki her  $p$  işaret noktasının  $m$ -boyutlu öznelik vektörü bulunmaktadır.  $N$  örneklem bulundurduğu varsayılan bir kümenin her örneği  $(x_i, y_i^l)$  çiftleri şeklinde temsil edilir ve her örneğin  $(i=1, \dots, N)$   $K$ -boyutlu öznelik vektörleri çıkarılır ( $K = p \cdot m$ ). Bu öznelik vektörleri  $(x_i \in R^K)$  yüzdeki işaret noktası yamalarına aittir, ve her etiket  $y_i^l \in \{-1, +1\}$  şeklinde (sınıflandırıcıya ait ya da değil) ifade edilir.  $\omega$ ,  $K$ -boyutlu bir ağırlık vektördür ve  $\omega$ 'ya ait  $k$ . öznelik  $\omega_k$  şeklinde temsil edilir. Sınıflandırıcılar için kullanılan maliyet fonksiyonu, Denklem (1)'de verilmiştir.

$$J(\omega, x_i, y_i^l) = \ln(1 + \exp(-y_i^l x_i \omega)) \quad (1)$$



**Şekil 2:** İki örnek AU için (AU1: İç kaşı kaldırma; AU17: Çeneyi kaldırma) en yüksek F1 skorlarını veren yama sayısı ve ilintili olduğu işaret noktaları.

AU'lara etki eden işaret noktalarının seyrek öğrenilmesi sırasında [12] çalışmasında kullanılan yöntem ele alınmış, fakat [12] çalışmasından farklı olarak çoklu görev yerine tek görev çalışıldığı için denklem aşağıdaki şekilde değiştirilmiştir, Denklem (2). Denklemde kullanılan  $\lambda$ , grup bazlı seyrekleme katsayısıdır,  $\omega_{G_j}$  ise  $j$  numaralı işaret noktasının ağırlık katsayısı grubunu temsil eder.

$$\min_{\omega} \sum_{i=1}^N J(\omega, x_i, y_i^t) + \lambda \sum_{j=1}^p \|\omega_{G_j}\|_2 \quad (2)$$

Denklem uygulandığında,  $\omega$ 'ya ait birçok sütun sıfırlanmış olacak, kalan sütunlara denk düşen işaret noktaları da, sınıflandırıcıya etkisi en yüksek olan işaret noktaları olarak kullanılacaktır.

### 3 Uygulama

#### A - Veri Kümesi

CK+ [18] veri kümesi, 123 öznenin 593 önden çekilmiş hareketli imgesini bulunduran ve literatürde en çok kullanılan verisetlerindedir. Veri kümesinde bazı öznelere duygu ve AU'ları da etiketlenmiştir ve videodaki yüzlerin işaret noktaları da temin edilmiştir. Bu çalışma kapsamında, veri kümesi tarafından temin edilen işaret noktaları yerine, [19]'da önerilen ve hem ön yüz hem de yandan çekilmiş yüzlerde başarılı olduğu saptanan bir işaret noktası çıkarıcı kullanılarak  $p=49$  işaret noktası takip edilmiştir. İşaret noktalarının etrafından çıkarılan yamalar  $11 \times 11$  boyutlarında çalışılmıştır.

#### B - Öznitelik Çıkarımı ve Sınıflandırma

Bu çalışmada öznitelik çıkarım yöntemi olarak laboratuvar ortamında kayıt edilmiş ya da bozuntuya uğramamış imge serilerinde başarılı çıkarımlar yaptığı ispatlanmış *LBP-TOP* [20] yöntemi kullanılmış, elde edilen öznitelik vektörünü  $1 \times 177$ -B şeklinde alınmıştır ( $m=177$ ). Veri kümesi çok fazla örnek içermediği için, eğitime ve sınıflandırmada *Leave-One-Subject-Out (LOSO)* yöntemi kullanılarak her katlamada bir özne ve özneye ait tüm imge serileri dışarıda bırakılmış, kalanlar eğitimde kullanılarak toplam N katlama yapılmıştır. Her AU ikili bir Destek Vektör Makinesi (SVM) [21] ile eğitilmiştir.

Seyrek öğrenme yöntemine giriş yapmadan önce gerçekleştirilen ilk deneyde,  $p=49$  işaret noktasından  $11 \times 11$ 'lik yamalardan çıkarılan öznitelikler ardarda birleştirilerek oluşturulan vektörle ( $K=p \cdot m$ ) sınıflandırma yapılmıştır. Çizelge-1 sonuçlarında tüm işaret noktaları kullanıldığında elde edilen F1 skorları, diğer çalışmalarla karşılaştırılmıştır.

#### C - Seyrek Öğrenme ve İşaret Noktalarından AU Saptama

[12] çalışmasında önerilen yama bazlı seyrek öğrenme yöntemi, yüzü satırlar ve sütunlar şeklinde eşit boyutlu yamalara bölerek gerçekleştirilmiştir. Fakat ifadeye etki etmesi mümkün olmayan, denegin yüzü dışındaki yamalar da deneylerde dikkate alınmıştır. Bu çalışmada önerilen yöntemde ifade tanıma yerine AU saptama kullanılarak çok daha büyük kapsamlı bir probleme çözüm aranmaktadır. Bu sebeple [12] çalışmasında kullanılan çok görevli seyrek öğrenme için önerilen problem çözümü tek göreve indirgenerek başarılı

AU sınıflandırıcıları elde edilmeye çalışılmıştır. Yapılan ilk 11x11-boyutlu yama deneylerinde 49 işaret noktasının hepsini deneye katmak yerine, AU ile en çok ilintili noktaları deneye katıldığında sonuçların daha yüksek olduğu görülmüştür. Şekil (2)'de iki adet AU için örnek F1 sonuç grafiği verilmiş ve en yüksek F1 skorunda hangi işaret noktalarının alındığı işaretlenmiştir.

Çizelge (1)'de son yıllarda yapılan çalışmalarla bu araştırmadan elde edilen sonuçlar kıyaslanmıştır. Kıyaslama sonucundan da görüleceği üzere özellikle alt yüze ait AU'larda, seyrek öğrenme ile yama bazlı yüz eylem birimi saptama daha başarılı sonuç vermiştir.

Tabloda koyu renkle vurgulanan sonuçlar karşılaştırmadaki en başarılı sonuçlardır. Sonuçlardan da görüleceği üzere, tüm yüz üzerinde çalışmak yerine işaret noktaları üzerinde çalışarak bir nevi *parazit* yamalar saf dışı bırakılmış, ele alınan tüm hareketli noktalar başarı oranlarını arttırmıştır.

## 4 Sonuçlar ve Öneriler

AU saptama çalışmalarında son zamanlarda yapılan çalışmalara bakıldığında, yüzün tamamında çalışmak yerine yamalar üzerinden gitmenin daha başarılı saptama sonucu verdiği görülmüştür. Bu doğrultuda yapılan çalışmalar yüzü tamamen eşit boyutlu karelere bölerek ya da işaret noktaları etrafından yamalar keserek öznitelik çıkarma yoluna gitmişlerdir. Bu çalışmada, yüzün işaret noktaları etrafından alınan yamalar seyrek öğrenme kullanılarak ilgili AU sınıflandırıcısı eğitilmiş ve yüzdeki tüm işaret noktalarını kullanmak yerine sınıflandırıcıya katkısı en büyük olan işaret noktalarından çıkarılan öznitelik vektörleri eğitmeye sokulmuştur. Farklı AU'lar üzerinde test edilen yöntem, özellikle alt yüz birimleri tarafından tetiklenen AU'larda diğer çalışmalardan daha başarılı sonuç vermiştir. Bu yöntemde, bir nevi *parazit* yamalar saf dışı bırakılmış, ele alınan sınıflandırıcıyla ilintili işaret noktaları başarı oranlarını arttırmıştır.

**Çizelge-1:** CK+ veri kümesinde F1 skorlarına göre AU saptama sonuçlarının karşılaştırmaları.

	[22]	[23]	[24]	[25]	[26]	49 işaret noktası	Yama Bazlı Seyrek Öğrenme
AU1	86,86	80,72	<b>89,12</b>	87,41	87,5	74,63	76,35
AU2	85,47	79,18	83,7	81,78	<b>85,5</b>	62,66	66,01
AU4	72,58	69,93	<b>79,79</b>	79,7	51,43	74,68	75,85
AU6	61,74	<b>69,81</b>	67,16	68,48	72,65	66,94	67,83
AU9	85,91	---	---	---	---	<b>87,14</b>	<b>87,14</b>
AU11	26,51	---	---	---	---	26,09	<b>43,14</b>
AU12	72,65	77,83	53,24	81,04	<b>85,95</b>	84,73	85,38
AU15	72,53	65,7	<b>77,83</b>	60,33	74,21	70,3	73,37
AU17	81,66	85,2	<b>85,97</b>	84,29	75,44	83,7	85,15
AU20	47,46	---	---	---	---	71,94	<b>72,22</b>
AU24	35,29	---	---	---	---	45,36	<b>55,14</b>
AU25	92,57	---	---	---	---	94,87	<b>95,56</b>

## 5 Bilgilendirme

Bu çalışma 115E310 no'lu Tübitak-1001 projesi kapsamında Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından desteklenmektedir.

## Kaynakça

- [1] Hjortsjö, Carl-Herman. *Man's face and mimic language*. Studen litteratur, 1969.
- [2] Ekman, Paul, Wallace V. Friesen. *Facial action coding system*. 1977.
- [3] Hager, J. C., P. Ekman, and W. V. Friesen. *Facial action coding system*. Salt Lake City, UT: A Human Face. ISBN 0-931835-01-1, 2002.
- [4] Lawrence Ian Reed, Michael A Sayette, and Jeffrey F Cohn. *Impact of depression on response to comedy: a dynamic facial coding analysis*. Journal of abnormal psychology, 116(4):804, 2007.
- [5] Amanda C Lints-Martindale, Thomas Hadjistavropoulos, Bruce Barber, and Stephen J Gibson. *A psychophysical investigation of the facial action coding system as an index of pain variability among older adults with and without alzheimer's disease*. Pain Medicine, 8(8):678-689, 2007.

- [6] Patrick Lucey, Jeffrey F Cohn, Iain Matthews, Simon Lucey, Sridha Sridharan, Jessica Howlett, and Kenneth M Prkachin. *Automatically detecting pain in video through facial action units*. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 41(3):664–674, 2011.
- [7] Marina Davila-Ross, Goncalo Jesus, Jade Osborne, and Kim A Bard. *Chimpanzees (pan troglodytes) produce the same types of ‘laugh faces’ when they emit laughter and when they are silent*. PloS one, 10(6):e0127337, 2015.
- [8] Takuya Hashimoto, Sachio Hitramatsu, Toshiaki Tsuji, and Hiroshi Kobayashi. *Development of the face robot saya for rich facial expressions*. In SICE-ICASE, 2006. International Joint Conference, pages 5423–5428. IEEE, 2006.
- [9] Gianluca Donato, Marian Stewart Bartlett, Joseph C Hager, Paul Ekman, and Terrence J Sejnowski. *Classifying facial actions*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 21(10):974–989, 1999.
- [10] Ying-li Tian, T Kanada, and Jeffrey F Cohn. *Recognizing upper face action units for facial expression analysis*. In Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, volume 1, pages 294–301. IEEE, 2000.
- [11] Caifeng Shan, Shaogang Gong, and Peter W McOwan. *Facial expression recognition based on local binary patterns: A comprehensive study*. Image and Vision Computing, 27(6):803–816, 2009.
- [12] Lin Zhong, Qingshan Liu, Peng Yang, Bo Liu, Junzhou Huang, and Dimitris N Metaxas. *Learning active facial patches for expression analysis*. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2562–2569. IEEE, 2012.
- [13] Li-Rong Zhong, Quanwei Liu, Ping Yang, Jie Huang, and Dimitris N Metaxas. *Learning multiscale active facial patches for expression analysis*. 2014.
- [14] Kaili Zhao, Wen-Sheng Chu, Fernando De la Torre, Jeffrey F Cohn, and Honggang Zhang. *Joint patch and multi-label learning for facial action unit detection*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2207–2216, 2015.
- [15] Marian Stewart Bartlett, Joseph C Hager, Paul Ekman, and Terrence J Sejnowski. *Measuring facial expressions by computer image analysis*. Psychophysiology, 36(02):253–263, 1999.
- [16] Ping Liu, Joey Tianyi Zhou, Ivor Wai-Hung Tsang, Zibo Meng, Shizhong Han, and Yan Tong. *Feature disentangling machine-a novel approach of feature selection and disentangling in facial expression analysis*. In Computer Vision–ECCV 2014, pages 151–166. Springer, 2014.
- [17] Sima Taheri, Qiang Qiu, and Rama Chellappa. *Structure-preserving sparse decomposition for facial expression analysis*. Image Processing, IEEE Transactions on, 23(8):3590–3603, 2014.
- [18] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. *The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression*. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pages 94–101. IEEE, 2010.
- [19] Xiangxin Zhu and Deva Ramanan. *Face detection, pose estimation, and landmark localization in the wild*. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2879–2886. IEEE, 2012.
- [20] Guoying Zhao and Matti Pietikainen. *Dynamic texture recognition using local binary patterns with an application to facial expressions*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(6):915–928, 2007.
- [21] Corinna Cortes and Vladimir Vapnik. *Support-vector networks*. Machine learning, 20(3):273–297, 1995.
- [22] Ziheng Wang, Yongqiang Li, Shangfei Wang, and Qiang Ji. *Capturing global semantic relationships for facial action unit recognition*. In Computer Vision (ICCV), 2013 IEEE International Conference on, pages 3304–3311. IEEE, 2013.
- [23] Andreas Damianou, Carl Ek, Michalis Titsias, and Neil Lawrence. *Manifold relevance determination*. arXiv preprint arXiv:1206.4610, 2012.

- [24] Raquel Urtasun, Ariadna Quattoni, Neil Lawrence, and Trevor Darrell. *Transferring nonlinear representations using gaussian processes with a shared latent space*. 2008.
- [25] Stefanos Eleftheriadis, Ognjen Rudovic, and Maja Pantic. *Discriminative shared gaussian processes for multiview and view-invariant facial expression recognition*. Image Processing, IEEE Transactions on, 24(1):189–204, 2015.
- [26] Xiao Zhang, Mohammad H Mahoor, S Mohammad Mavadati, and Jeffrey F Cohn. *A  $l_p$ -norm mtmkl framework for simultaneous detection of multiple facial action units*. In Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on, pages 1104–1111. IEEE, 2014.