

# Filmlerde Otomatik Duygu Analizi için Sentetik Veri Üretimi ve Çokkipli Birleştirme

## Movie Emotion Estimation with Multimodal Fusion and Synthetic Data Generation

Nihan Karslıoğlu  
Computer Engineering  
Boğaziçi University  
Istanbul, Turkey  
Email: nihankarslioglu@gmail.com

Heysem Kaya  
Computer Engineering  
Namık Kemal University  
Çorlu, Turkey  
Email: hkaya@nku.edu.tr

Albert Ali Salah  
Information and Computing Sciences  
Utrecht University, Utrecht, the Netherlands,  
Computer Engineering  
Boğaziçi University, Istanbul, Turkey  
Email: a.a.salah@uu.nl

**Özetçe** —Bu çalışmada film sahnelerinden otomatik duygu tanıma üzerine bir yöntem öneriyoruz. Bu problemin büyük film ve video veritabanlarının otomatik endekslenmesi, görsel malzemenin otomatik olarak özetlenmesi, duygu uyandıran görsellerin otomatik seçilmesi gibi uygulamaları bulunmaktadır. Önerdiğimiz yöntem izleyicide uyandırılan duygunun olumluluk ve şiddet değerlerini otomatik olarak çıkartmayı amaçlamaktadır. Yöntemimizde film kesitlerinden ses ve görüntüyü kullanarak farklı öznitelikler çıkardık ve bu öznitelikleri özetleyici fonksiyonlar, temel bileşenler analizi ve Fisher vektörü kodlaması kullanarak özetledik. Bunun yanı sıra kanonik korelasyon analizi bazlı öznitelik seçimi ile önemli öznitelikleri seçtik. Sınıflandırıcı olarak ekstrem öğrenme makinesi ve destek vektör makinesi kullandık. Yöntemimizi duygu işaretlemesi yapılmış LIRIS-ACCEDE veritabanı ile test ettik. Veri dağılımındaki dengesizliği az örneği bulunan sınıf için sentetik veri üretmekle çözdük. İyi sonuç veren öznitelikleri skor ve öznitelik seviyesinde birleştirerek, özellikle duygu olumluluğu tahmininde iyi sonuçlar elde ettik.

**Anahtar Kelimeler**—Film analizi, duyuşsal bilişim, çokkipli tümleştirme.

**Abstract**—In this work, we propose a method for automatic emotion recognition from movie clips. This problem has applications in indexing and retrieval of large movie and video collections, summarization of visual content, selection of emotion-invoking materials, and such. Our approach aims to estimate valence and arousal values automatically. We extract audio and visual features, summarize them via functionals, PCA, and Fisher vector encoding approaches. We used feature selection based on canonical correlation analysis. For classification, we used extreme learning machine and support vector machine. We tested our approach on the LIRIS-ACCEDE database with ground truth annotations. The class imbalance problem was solved by generating synthetic data. By fusing the best features at score and feature level, we obtain good results on this problem, especially for the valence prediction.

**Keywords**—Movie analysis, affective computing, multimodal fusion.

### I. GİRİŞ

Bu çalışmada videodan duygu tanıma için bir yaklaşım öneriyoruz. Her ne kadar filmlerde duygu aktarımı için kullanılan teknikler çeşitlilik gösterse de; sinema literatüründe bazı sinematografik yöntemler ve düzenlemeler çokça kabul görmüşlerdir. Bunlara örnek vermek gerekirse, aktörlerin yüz ifadeleri ve sesleri, ses ve görüntü efektleri, müzikler, renkler ve sahne dekoru gibi faktörleri sayabiliriz [1], [2].

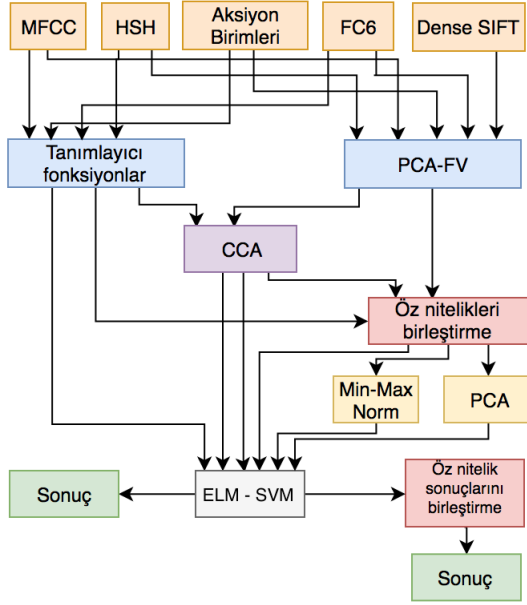
Bu çalışmada duyguları tanımlamak için duygu olumluluğu (valence) ve duygunun şiddeti (arousal) eksenlerini (VA uzayı) kullandık. Şiddet duygunun sakinliği ya da heyecanını tanımlarken, olumluluk duygunun pozitif veya negatifliğini tanımlar. Filmlerde duygu tahmini yapan çalışmalar sinematografik öğelerden yararlanmıştır. Bunlar arasında ses verisinden Mel Frekanslı Kepstrum Katsayıları (MFCC) ve enerji öznitelikleri, görüntü verisinden de renk ve ışık özniteliklerinin öne çıktığını görüyoruz [3], [4], [5], [6], [7]. MFCC ve renk öznitelikleri derin yapay sinir ağları ile işlenebilmektedir [8]. Çok miktarda görsel özniteligi özetlemek için Fisher vektörü kodlaması önerilmiştir [9]. Veri dengesizliği problemi, verisi az olan sınıfın örneklerini sentetik örnekleme artırma yoluyla çözülmeye çalışılmıştır [10]. Duygu tahminlerinde gürültüyü engellemek için Kaiser penceresi ile öznitelik yumuşatma kullanılmıştır [3].

### II. SİSTEM TANIMI

Geliştirdiğimiz sistem ile MediaEval 2017 duygu tanımlama yarışmasında [11] kullanılan LIRIS-ACCEDE veri setindeki filmlerin onar saniyelik sahnelerinde duygu olumluluğu ve şiddetinin değerlerini bulmaya çalıştık. Bunun için öznitelik çıkarma, seçme, sınıflandırma ve elde ettiğimiz en iyi modelleri birleştirme kısımlarını geliştirdik (bkz. Şekil 1).

#### A. Öznitelik çıkarma

Filmlerde duyguyu aktaran ses, aktörlerin yüz ifadeleri, sahne renkleri, sahne dekorasyonu ve videodaki hareketler gibi birçok öğe bulunmaktadır. Modelimizde ses bilgisi için



Şekil 1: Deney sistemi tasarımı

32 milisaniyelik pencere ve yüzde 50 çıkışan video sahnelerinden Mel Frekans Kepstrum Katsayıları (MFCC 0-12) öz niteliklerini çıkardık. Görsel bilgi olarak ise dört farklı öz nitelikten faydalandık: 1) Her kare için 33 kademeli renk tonu (hue) öz niteliği ve 31 kademeli renk doygunluğu (saturation) öz niteliğini birleştirerek 1023 boyutlu ton-doygunluk histogramı (HSH); 2) 4,6,8 ölçülerinde ve 7 piksel aralıklı olarak her 30 video çerçevesinden Dense-SIFT öz niteliği [12]; 3) VGG derin sinir ağlarından çıkartılan öz nitelikler [13] (altıncı ful konvolüsyon katmanı - VGG-FC6); 4) Sahnelerdeki kişilerin yüz hareket birimleri (AU). VGG-FC6 öz niteliğini Mediaeval 2017 organizasyonu [11] sağladı. [14] çalışmasının kütüphanesi kullanılarak, film sahnelerinde mevcut olan yüz hareket birimleri çıkartıldı.

### B. Öz nitelik özetleme

Videoyu tanımlayan öz niteliklerin boyutları toplamda çok büyük olduğu için bunları özetlemek gerekmektedir. Bunun için bir grup öz niteliği fonksiyonlar yolu ile özetledik. MFCC ve AU öz niteliklerini minimum ve maksimum fonksiyonları ile, HSH öz niteliklerini minimum, maksimum ve ortalama değerleri ile, VGG-FC6 için ise yalnızca ortalama değer fonksiyonu ile özetledik. Dense-SIFT öz nitelikleri bu şekilde bir özetlemede anlamlı sonuç üretmeyeceği için bu fonksiyonları kullanmadık.

Tanımlayıcı fonksiyonların yanı sıra farklı bir özetleyici yöntem olarak temel bileşenler analizi (PCA) ile öz niteliklerin boyutunu küçültüp ardından Fisher vektörü (FV) [15] ile öz nitelikleri kodladık. Fisher vektörleri öz niteliklerin standart bir olasılık modelinden ne kadar saptığını Gauss karışımlarına bakarak hesaplamaktadır. Küme sayısı MFCC, AU ve Dense-SIFT için 16, HSH ve VGG-FC6 için ise 1 olarak alındı. Ardından Fisher vektörleri kare fonksiyon ve L2 normu ile normalize edildi.

### C. Öz nitelik seçimi

Öz nitelikleri özetlemenin ardından rastgele örneklere karşı etiketlerin kanonik korelasyon analizi [16] algoritması sıralama tabanlı bir öz nitelik seçimi olarak kullanıldı. Bu yaklaşımın ana fikri öz nitelikler ile hedefteki sınıflara kanonik korelasyon analizi uygulayıp ardından projeksiyon ağırlıklarının mutlak değerlerini sıralamaktır. Kullandığımız sistemde toplam ağırlığın yüzde 90'ını özetleyen öz nitelikler seçilmiştir. Daha önce çokkipli tümleştirmede etkin kullanıldığı için bu yaklaşım seçilmiştir [16].

### D. Regresyon

VA uzayında regresyon için Ekstrem Öğrenme Makineleri (ELM) [17] ve Destek Vektör Makineleri (SVM) [18] kullanıldı. Analitik çözümü olan ELM yaklaşımı SVM'den çok daha hızlı eğitilen ve parametre optimizasyonuna göre performansı büyük ölçüde değişmediğinden SVM'den pratikte daha yüksek başarı elde eden bir yaklaşımdır [17]. Bu sınıflandırıcılar için dairesel temelli (radial basis function - RBF) ve doğrusal (linear) çekirdek fonksiyonları seçildi. ELM ve SVM'de en iyi parametreleri bulmak için ızgara arama uygulandı. Düzenleştirici katsayı ve RBF çekirdeği ölçüm parametresi [0.01,1000] aralığında üssel adımlarla arandı. Ölçüm metriği olarak Pearson korelasyon katsayısı (PCC) ve kare hata (MSE) kullanıldı. PCC, tahmin edilen değer ile gerçek değer arasındaki doğrusal korelasyonun yönünü ve şiddetini; MSE ise bu iki değer arasındaki ortalama kare farkı ölçer.

Kullandığımız regresyon yaklaşımı zaman kavramını dikkate almamaktadır. Fakat filmlerde duyguda anlık değişimler yerine bir akış mevcuttur. Yani olumsuz bir duygu vermeye çalışan bir film yapımcısı bu duyguyu aktarmaya çalışırken akışı buna göre oluşturur. Bu sebeple yumuşatma kullanarak bu durumu programlamaya çalıştık. 20 pencerelek, yani etiketleme prosedürüne göre yaklaşık olarak 1.5 dakikaya denk gelen bir süre ile sinyali düzenledik.

### E. Örnekleme artırma

Kullandığımız veritabanında MediaEval yarışmasının standart değerlendirme protokolünü kullandık ve veriyi geliştirme kümesi, eğitim kümesi, doğrulama kümesi ve test kümesi olarak ayırdık. Model geliştirme için ayırdığımız kısımdaki VA uzayı etiketlemelerini incelediğimizde değer dağılımında dengesizlik görüyoruz. [-1, 1] duygu şiddeti aralığını üç anlamlı parçaya böldük. Bunları [-1, -0.35] aralığı "düşük", (-0.35, 0.35) "nötr" ve (0.35, 1] "yüksek" olmak üzere tanımladık. Bu parçalama yöntemine göre geliştirme veri setindeki düşük/nötr/yüksek örnek oranı [4,30,7] şeklinde oluştu. Yani nötr örnek sayısı fazla ve negatif örnek oldukça azdı. Duygu olumluluğu eksenini aynı şekilde bölümlendirdiğimizde bu oran [1,2,1] şeklinde oluştu. Burada ise nötr sayısı daha fazlaydı.

Eğitim ve doğrulama setlerindeki film sayısı %60 - %40 olarak ayırmaya çalıştık. Eğitim setindeki toplam etiket sayısının doğrulama setinden yüksek olmasına özen gösterdik.

Bu şekilde sınıflara böldüğümüz VA uzayında ortaya çıkan sınıf dengesizliği sayıca çok olan sınıfa (yani eğitim kümesinde daha baskın olan duygu aralığına) ağırlık verilmesine sebep oluyordu. Bu sonuçların sağlıklı olmaması sebebiyle

ADASYN adı verilen ve sayıca çok olan sınıfın örnek sayısına göre diğer sınıflarda örnekleme yapabilen bir algoritmadan yararlandık [19]. Bu tekniği sınıflandırma işleminden önce kullanıyoruz. Etiket eksenini veri sayısı analizindeki gibi üç parçaya böldüğümüzde etiket ekseninde yetersiz veri bölgeleriyle karşılaşırız. Bu yüzden yedi eşit parçaya bölerek etiketlerimizin daha dengeli bir şekilde dağılmasını sağladık. Böylece (-1, 1) aralığında her bölgede yeterli sayıda veri oluşturduk. Ekseni daha küçük aralıklara bölmeyi denediğimizde, aralarında önemsiz farklar olan çok fazla benzer grup oluştuğunu gördük. Bu da sınıflandırıcımızın daha iyi tahminler yapmasına katkı sağlamadığı gibi sistem hafızası gibi teknik sınırlara takılmamıza sebep oldu. Bu sebeple etiket eksenini yedi parçayla sınırlandırmayı yeterli gördük.

#### F. Öznitelikleri birleştirme

Bir filmin izleyicide duygu oluşturan öğeleri birbirinden bağımsız değildir. Yönetmen ses ve görüntü gibi birden fazla öğeyi uyumlu bir şekilde kullanarak bazı duyguları tetikleyebilir. Sistemimizde bunu modelleyebilmek için öznitelik birleştirme yaklaşımı kullandık.

Bir birleştirme yöntemi olarak regresyon sonucu gelen skorlarda çok terimli regresyon yöntemini kullandık. Bunun dışında öznitelikleri sınıflandırıcıya vermeden önce de birleştirdik ve min-max normalizasyonu uygulayarak farklı özniteliklerin sayısal olarak dengeli olmasını sağladık. Ayrıca bu birleştirmeyi PCA ile boyut düşürerek de denedik.

### III. DENEY SONUÇLARI

MediaEval 2017 duygu tanımlama yarışmasının geliştirme setinde 30 film vardır. Bu sette etiketli veri sayısı toplamı 5274'tür. Test setinde ise 17 film ve 5708 etiketli veri vardır. Şekil 1'de sistemimizde uyguladığımız deneyler şematik olarak görülmektedir. PCC ve MSE olarak iki farklı ölçütümüz olsa da, temelde modellerimizi PCC sonuçlarına göre belirledik.

Dengesiz dağılımlı veri setimizden, öznitelikler çıkarıp ELM uyguladık. Bu değerlerin sonuçları Tablo I'dedir. Her bir deneyin sonucu ayrı bir satırda raporlandı. Kullanılan öz nitelik ile MSE ve PCC ölçütlerinde elde edilen değerler farklı sütunlarda verildi. MSE'nin düşük, PCC'nin yüksek olması tercih edilir. Duygu olumluluğunda en iyi sonucu AU ile aldık. Duygu şiddetinde en iyi sonucu ise HSH ile aldık. Yalnız tahmin edilen sonuçları incelediğimizde her öznitelik için ayrı olarak, sonuçlar hep belli bir değere yakın sonuç üretti. Dengesiz dağılımlı veri setinde çalışmamız bu probleme sebep olmuş olabilir. ADASYN [16] örnekleme artırımı yöntemiyle, az örneği bulunan sınıflardan sentetik veriler üretildi. Sentetik verilerin eğitim setine katılmasıyla veri dağılımında eşitlik sağlandı.

Her bir öznitelik için deneylediğimiz yöntemlerden, aldığımız en iyi ve önemli sonuçların değerleri Tablo II ve Tablo III'te yer almaktadır. Bu tablolarda her bir deneyin sonucu ayrı bir satırda raporlandı. Deneylenen sınıflandırıcı, yöntem, öz nitelik ile MSE ve PCC ölçütlerinde elde edilen değerler; farklı sütunlarda verildi. Tablo II'de duygu olumluluğunda AU öznitelikleri PCC ölçütüne göre, görece daha iyidir. Fakat AU öz niteliğiyle gelen sonuçların grafiğini incelediğimizde sonuçların sifra yakın tahminlerin yapıldığını gözlemledik. Performans sonucu olarak en yüksek sonucu AU

TABLE I: Dengesiz dağılımlı veride; öz niteliklerin tanımlayıcı fonksiyonlar ile tanımlanıp, ELM uygulanarak edilen sonuçları

Öznitelik	Valence		Arousal	
	PCC	MSE	PCC	MSE
HSH	0.1175	0.0847	<b>0.0984</b>	0.1398
MFCC	-0.1886	0.0915	-0.0261	0.1682
AU	<b>0.1294</b>	0.0865	0.0314	0.1079
FC6	-0.1762	0.0954	-0.2055	0.1247

TABLE II: Duygu olumluluğu uzayında elde edilen en iyi performans değerleri

Sınıflandırıcı	Yöntem	Öznitelik	PCC	MSE
ELM	PCA, FV	VGG-FC6	0.1896	0.0826
	Tanıml. fonk.	VGG-FC6	0.1454	0.0982
	PCA, FV, CCA	VGG-FC6	0.1386	0.0978
	Tanıml. fonk.	AU	<b>0.2195</b>	0.0930
	Tanıml. fonk.	HSH	0.1284	0.0823
SVM	İstatist., CCA	VGG-FC6	0.1697	0.0952
	PCA, FV, CCA	MFCC	0.0999	0.0925

ile alsak da, gerçek değerlere benzer korelasyonu görmemiz beklenirdi. VGG-FC6 öz niteliği için ELM ile sınıflandırmada, tanımlayıcı fonksiyonlar yerine PCA ve FV özetleme yöntemleri kullanıldığında; korelasyonlarda yükselme, MSE değerlerinde düşme gerçekleşti. Bu öz niteliğe kanonik korelasyon analizi bazlı öz nitelik seçimi eklenmesi; PCC değerlerini düşürdü, MSE değerlerini yükseltti. ELM sınıflandırmasıyla alınan en iyi sonuç, SVM ile alınan en iyi sonuçtan yüksek çıktı.

Duygu şiddeti için, ELM sınıflandırıcısı kullanılarak elde edilen en iyi sonucu; HSH öz niteliğine tanımlayıcı fonksiyonlar ve kanonik korelasyon öznitelik seçimi uygulayarak elde ettik (Tablo III). Diğer öznitelikler için ELM sınıflandırıcısıyla alınan sonuçların zayıf olduğunu görüyoruz. Buna rağmen; MFCC, HSH, AU ve Dense-SIFT için gerçek değerler ve tahminler grafiğini incelediğimizde, pozitif korelasyonu olan kısımları gördük. Duygu şiddeti uzayında da SVM sınıflandırılmalarında alınan sonuçların, ELM'de elde edilen en iyi sonuçtan düşük olduğunu gözlemledik.

Özniteliklerden elde ettiğimiz sonuçların ardından birleştirme yöntemleri ile sistemimizi iyileştirmeye çalıştık. Skor tabanlı birleştirmeyi ELM tahmininden gelen değerler için

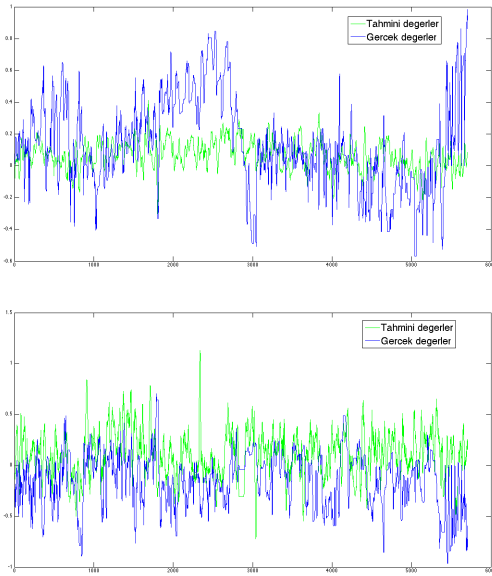
TABLE III: Duygu şiddeti uzayında elde edilen en iyi performans değerleri

Sınıflandırıcı	Yöntem	Öznitelik	PCC	MSE
ELM	Tanıml. fonk., CCA	HSH	<b>0.1264</b>	0.1409
	PCA, FV	MFCC	0.0237	0.1396
	PCA, FV	Dense Sift	-0.1080	0.0940
	PCA, FV, CCA	AU	-0.0452	0.1280
SVM	Tanıml. fonk.	MFCC	0.1205	0.1814
	PCA, FV	VGG-FC6	0.0690	0.1729

denedik. En iyi sonuç aldığımız deneylerin skorlarını seçtik. Duygusal olumluluk uzayı için VGG-FC6 özneteliği, PCA ve FV yöntemlerinde iyi sonuç vermişti. Bu sonucu AU ve HSH için tanımlayıcı fonksiyonlardan aldığımız skorlar ile birleştirdik. VGG-FC6 ve AU birleşimi ile 0.1903 PCC değeri ve 0.0824 MSE değeri elde ettik. Skor tabanlı birleştirmeyi SVM tahmininden gelen değerler için de denedik. Duygusal olumluluk uzayı için VGG-FC6 özneteliği tanımlayıcı fonksiyonlar ile özetlenip kanonik korelasyon bazlı öznetelik seçildiğinde iyi sonuç vermişti. Bu sonucu MFCC için PCA ve FV yöntemleri ile özetleyip, kanonik korelasyon bazlı öznetelik seçtiğimiz skorla birleştirdik. VGG-FC6 ve MFCC birleşimi ile 0.1821 PCC değeri ve 0.1947 MSE değeri elde ettik. ELM tahmininden gelen sonuçlarda skor tabanlı birleştirmeye, SVM tahmini için yapılan skor tabanlı birleştirmeden daha iyi bir sonuç verdi.

Duygu olumluluğu için en iyi çalışan modelimizdeki gerçek etiketler ve tahminler grafiğini incelediğimizde -0.2 ile 0.4 arasında tahminler yapıldığını gördük (bkz. Şekil 2). Bu tip sistemlerde otomatik sınıflandırıcının ortalama değere yakın kalması, yani tutucu tahminler yapması, problemdeki belirsizliğin yüksek olduğuna işaret eder.

Duygu şiddeti için SVM sınıflandırıcısıyla en iyi sonucu MFCC özneteliği, tanımlayıcı fonksiyonlar ile özetlenip kanonik korelasyon bazlı öznetelik seçildiğinde iyi sonuç vermişti. VGG-FC6 özneteliğine, PCA ve FV yöntemi uygulanmasıyla duygu şiddeti için pozitif korelasyonu olan kısımlar görmüştük. Bu iki özneteliği birleştirdiğimizde 0.1025 PCC değeri ve 0.1375 MSE değeri elde ettik. Duygu şiddeti için ELM sınıflandırıcısıyla, en iyi sonucu HSH özneteliğinin tanımlayıcı fonksiyonlar ile tanımlanıp kanonik korelasyon ile seçildiği teknikte elde ettik. VGG-FC6 dışındaki öznetelikler ile birleştirme yaptık. AU ile birleştirmede bu boyuttaki en iyi sonucu elde ettik (0.1322 PCC ve 0.1670 MSE, bkz. Tablo III).



Şekil 2: Olumluluk (üstte) ve şiddet (altta) uzaylarında en iyi tahminler ve gerçek değerler.

#### IV. VARGILAR

Çalışmamızda film sahnelerinde duygu tahmini için çeşitli öznetelikler ve farklı özetleme teknikleri uyguladık. Sonuçlarımızdan duygu şiddeti ve olumluluğunu iki farklı problem gibi ele alarak, değişik yaklaşımlarla çözmek gerektiğini görüyoruz. Daha karmaşık işleme ve birleştirme yöntemlerinden beklenen sonuçları alamamış olmamız, bu problem için daha büyük işaretli veri setleri ile eğitim yapılması gerektiğine işaret ediyor olabilir. Özellikle olumluluk boyutunda sınıflandırıcıların ortalama değerden uzaklaşmaması, kısıtlı veriden kaynaklanan bir soruna işaret etmektedir.

#### KAYNAKÇA

- [1] G. M. Smith, *Film structure and the emotion system*. Cambridge University Press, 2003.
- [2] C. Plantinga and G. M. Smith, "Passionate views: Film," *Cognition, and Emotion*, 1999.
- [3] A. Hanjalic and L.-Q. Xu, "Affective video content representation and modeling," *IEEE transactions on multimedia*, vol. 7, no. 1, pp. 143–154, 2005.
- [4] M. Muszynski, L. Tian, C. Lai, J. Moore, T. Kostoulas, P. Lombardo, T. Pun, and G. Chanel, "Recognizing induced emotions of movie audiences from multimodal information," *IEEE Transactions on Affective Computing*, 2019.
- [5] M. Soleymani, G. Chanel, J. J. Kierkels, and T. Pun, "Affective characterization of movie scenes based on multimedia content analysis and user's physiological emotional responses," in *Multimedia, 2008. Tenth IEEE Int. Symposium on*, 2008, pp. 228–235.
- [6] Y. Timar, N. Karslioglu, H. Kaya, and A. A. Salah, "Feature selection and multimodal fusion for estimating emotions evoked by movie clips," in *Proc. ICMR*. ACM, 2018, pp. 405–412.
- [7] N. Karslioglu, Y. Timar, A. A. Salah, and H. Kaya, "Boun-nku in mediaeval 2017 emotional impact of movies task," in *Proc. MediaEval*, 2017.
- [8] E. Acar, F. Hopfgartner, and S. Albayrak, "A comprehensive study on mid-level representation and ensemble learning for emotional analysis of video material," *Multimedia Tools and Applications*, vol. 76, no. 9, pp. 11 809–11 837, 2017.
- [9] T. Anastasia and H. Leontios, "AUTH-SGP in MediaEval 2016 emotional impact of movies task," in *Proc. MediaEval*, 2016.
- [10] D. McDuff and M. Soleymani, "Large-scale affective content analysis: Combining media content features and facial reactions," in *Proc. IEEE FG*, 2017, pp. 339–345.
- [11] E. Dellandréa, M. Huigsloot, L. Chen, Y. Baveye, and M. Sjöber, "The MediaEval 2017 emotional impact of movies task," in *Proc. MediaEval*, 2017.
- [12] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *Proc. IEEE CVPR*. IEEE, 2007.
- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [14] T. Baltrušaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," in *Proc. IEEE FG*. IEEE, 2015.
- [15] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *Proc. IEEE CVPR*. IEEE, 2007.
- [16] H. Kaya, T. Özkaptan, A. A. Salah, and F. Gürgeç, "Random discriminative projection based feature selection with application to conflict recognition," *IEEE Signal Processing Letters*, 22(6), pp. 671–675, 2015.
- [17] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Systems, Man, and Cybernetics-B*, vol. 42, no. 2, pp. 513–529, 2012.
- [18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [19] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IJCNN*. IEEE, 2008, pp. 1322–1328.