

Dinamik Göz Bakışı Analizi

Yüzyüze İletişim için bir Uygulama Ortamı

Dynamic Gaze Analysis

An Application Environment for Face-to-Face Communication

Ülkü Arslan Aydın

Bilişsel Bilimler Programı, Enformatik Enstitüsü
Orta Doğu Teknik Üniversitesi
Ankara, Türkiye
ulku.arslan@gmail.com

Sinan Kalkan

Bilgisayar Mühendisliği Bölümü
Orta Doğu Teknik Üniversitesi
Ankara, Türkiye
skalkan@metu.edu.tr

Cengiz Acartürk

Bilişsel Bilimler Programı, Enformatik Enstitüsü
Orta Doğu Teknik Üniversitesi
Ankara, Türkiye
acarturk@acm.org

Özet—Dinamik ortamlarda göz bakışı analizi, görsel çevredeki nesnelerin tespit ve takip edilmesindeki güçlükler nedeniyle tam olarak çözüm üretilmemiş problemlerden birisidir. Bu çalışma, dinamik göz bakışı analizine, görsel nesneler olarak yüzlerin yer aldığı yüzyüze iletişim ortamı özelinde bir çözüm üretmektedir. Bu kapsamda geliştirilen uygulama, video kayıtlarında konuşmacıların yüzlerini otomatik olarak tespit ve takip etmekte, konuşmacıların göz bakışı konumlarını yüz görüntüleri ile eşleyerek göz kaçırma davranışlarının tespit edilmesine olanak sağlamaktadır. Analiz, iletişim sırasındaki konuşmaları bölümlendirme ve ayrıştırılmasını eşzamanlı olarak gerçekleştirmekte, kullanıcıya sözeylem etiketleri yoluyla işaretleme olanağı sağlamaktadır. Pilot çalışmalar, uygulamanın kabul edilebilir doğruluk değerleri sunması yanında analizler zamanını da önemli ölçüde azalttığını göstermektedir.

Anahtar Kelimeler—Göz bakışı analizi, konuşma analizi, otomatik yüz tanıma, otomatik konuşma bölütleme

Abstract—Gaze analysis in dynamic environments has remained an unresolved problem due to the complexities that pertain to the detection and tracking of objects in the visual environment. This study provides a solution to the problem for face-to-face communication, in which the visual objects in the environment are faces. The application that has been developed for this purpose is able to detect and track faces in video stream, and it maps gaze locations to the images, thus allowing the user to detect gaze behavior, such as gaze aversion. The application is also capable of segmentation and diarization of speech synchronously with the video stream and eye movement overlay. It allows the user to annotate speech by speech act labels. The pilot studies reveal that the application provides acceptable accuracy values in the analysis, as well as significantly reducing the time for the analyses.

Index Terms— Gaze analysis, speech analysis, automatic face detection, automatic speech segmentation

I. GİRİŞ

Yüzyüze iletişim, tarafların sözlü sinyaller yanında yüz ifadesi, göz teması ve beden hareketleri gibi sözlü olmayan sinyalleri aktarması ve yorumlamasıyla gerçekleşir. Sözlü olmayan iletişimin kapsamı sözlü iletişime kıyasla geniştir. Bu doğrultuda, sözlü olmayan iletişimin analiz yöntemleri sözlü iletişim analiz yöntemlerine (örn. doğal dil işleme analizine) göre geniş bir analiz yelpazesinin geliştirilmesini ve birbirinden farklı özniteliklere sahip kiplerin araştırılmasını gerektirmektedir [1], [2].

Sözlü olmayan iletişimde sıkça üzerinde durulan analiz yöntemlerinden birisi video verisinden yüz ifadelerinin yorumlanmasıdır. Bu kapsamdaki çalışmalar tarafların yüz ifadelerini duygu durumuna göre etiketlemeyi amaçlamaktadır. Son yıllarda iletişimde göz bakışı (ing. gaze) yoğunlukla incelenen alanlardan birisi olarak ortaya çıkmıştır. Bu kapsamdaki çalışmalar gözün yüzde nereye odaklandığından göz teması ve göz kaçırma davranışlarının modellenmesine kadar bir dizi araştırmayı kapsamaktadır [3], [4], [5], [6].

Gözün iletişimdeki rolü dili tamamlayıcı olmakla birlikte dilin kullanımını yönlendirici de olabilmektedir. Örneğin, göz teması ve göz kaçırma iletişimde kişinin ne zaman konuşmaya başlayacağı, konuşmasını ne zaman sonlandırıp karşı tarafa sıra vereceği (ing. turn taking) gibi iletişimin sürekliliğini sağlayan temel mekanizmalar üzerinde belirleyicidir. Bu çalışma göz teması ve göz kaçırma davranışına odaklanmaktadır.

Göz teması (ve dolayısıyla göz kaçırma davranışı) yüzyüze iletişimde tarafların çoğunlukla farkında olmadan oluşturdukları, sözlü olmayan iletişim biçimlerinden birisidir. Bu davranışlar dilden bağımsız olarak değerlendirilmemekte

II. YÜZ TESPİTİ VE YÜZ İZLEME

olup dil kullanımı ile ilişkisi çerçevesinde, çokkipli (ing. multimodal) bir çerçevede incelenmelidir.

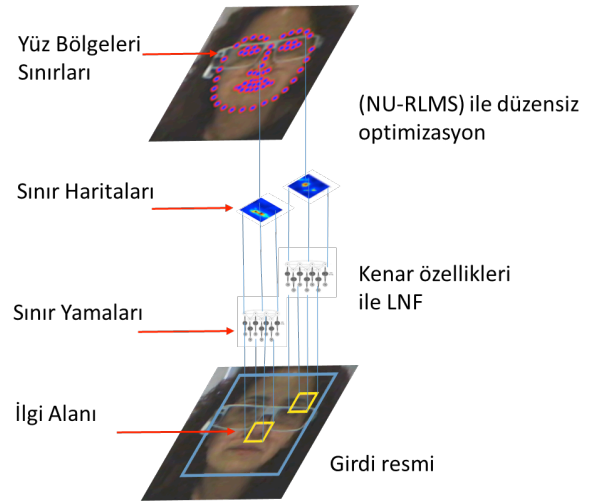
İletişimde dil kullanımını analiz etmeyi hedefleyen kuramlardan birisi sözeylem kuramıdır (ing. speech act theory) [7], [8]. Sözeylem kuramı dil kullanımını, ifadelerin temsil ettiği eylemler cinsinden inceleyen işlevsel yaklaşımlı bir kuramdır. Bu kurama göre, örneğin, düzsöz eylemi (ing. locutionary act) söylenen ifadenin doğrudan anlamını temsil ederken edimsel eylem (ing. illocutionary act) söyleyicinin niyetini temsil eder; etkisöz (ing. perlocutionary act) ise söylenen sözün, dinleyen taraf üzerindeki etkisini ifade eder. Bu doğrultuda iletişimdeki dilsel ifadeler talimat (ing. directive), yüklenme (ing. commissive), iddia (ing. assertive), dışavurum (ing. expressive), bildirim (ing. declarative) gibi kategori etiketleri ile tanımlanabilir. Biz bu çalışmada yüzyüze iletişimin dil kullanımı ile ilgili boyutlarını sözeylem kuramı çerçevesinde ele alıyoruz. Sözü edilen dil kullanımı etiketlerini gözbakışı ile eşzamanlı olarak analiz etmeyi hedefliyoruz.

Bu çalışmanın özgün boyutu, dil kullanımını göz hareketi davranışları ile birlikte analiz etmenin yanında gözbakışının dinamik olarak analiz edilebileceği bir uygulama ortamı sunmasıdır. Gözbakışının analizi ile ilgili alanyazındaki önemli problemlerden birisi dinamik ortamlarda gözbakışının analizi ile ilgili kabul gören, genel geçer bir yöntemin henüz geliştirilmemiş olmasıdır [9], [10], [11]. Mevcut teknolojide göz hareketi takip cihazı üreticileri, sundukları yazılımlarda çoğunlukla temel göz hareketi ölçütlerini, örneğin Hedef İlgi Alanı (HİG, ing. Area of Interest, kısaca İlgi Alanı İA) temelli analizleri statik arayüzlerde tanımlamakta, kullanıcının hareketli olduğu ortamlardaki gözbakışı için ise çoğunlukla göz konumu verisinin RGB görüntü kaydı üzerine yerleştirildiği video kayıtları sağlamakta ve/veya video kayıtlarından çıkarılabilecek durağan görüntü setlerinin tanımlanabildiği kullanıcı arayüzleri sunmakta, dinamik gözbakışı analizi ile ilgili, el ile analiz gerektiren kısmi çözümler sağlayabilmektedir.

Dinamik gözbakışı analizindeki güçlüğün temel nedeni, araştırmacı tarafından belirlenen bir İlgi Alanı'nın (örneğin ekran görüntüsü içerisindeki bir nesnenin) dinamik bir video kaydında sürekli olarak takip edilmesindeki teknik problemlerdir. Örneğin, futbol maçı izleyen bir seyircinin topa ne zaman ve ne kadar süre baktığının analizi, topun konumunun video kaydında sürekli takip edilmesini ve izleyicinin gözbakışı (ing. gaze) konumunun topun konumu ile örtüştüğü anların tespitini gerektirmektedir. Bu aşamada ortaya çıkan problemler (örn. düşük ışık miktarından kaynaklı nesne takip kayıpları) nesne tanıma ve nesne takip algoritmalarının geliştirilmesi ile gelecekte tama yakın çözüm üretilmesi beklenen problemlerdir.

Bizim bu çalışmadaki amacımız göreceli olarak çok çalışılmış bir alanda, yüz tanıma ve yüz takip alanında geliştirilen uygulamaların yüzyüze iletişimde dinamik gözbakışının analizi için geliştirildiği bir uygulama ortamını tanıtmaktır. Aşağıdaki bölümlerde bu uygulama ortamının bileşenleri anlatılacaktır.

Yüz tespiti ve yüz izleme bilgisayarlı gözü alanında yoğun olarak çalışılan alanlardan birisidir. Temel amaç, bir video karesinde (ing. frame) yüzü tespit ettikten sonra görüntü akışı boyunca yüzü takip etmektir. Bu çalışmada açık kaynak kodlu bir uygulama olan OpenFace uygulaması üzerinde geliştirmeler yapılmıştır. OpenFace, hazır yüz tespit kütüphaneleri yanında yapay öğrenme yöntemleri içeren *Dlib C++ Library* kütüphanesini de içeren bir uygulamadır. Bu uygulamada MMOD (Max-Margin Object-Detection) öğrenme algoritması HOG (Histogram of Oriented Gradients) öznelik çıkarımı ile kullanarak az miktarda veri ile yüz tespiti gerçekleştirilebilmektedir [12], [13]. Yüzün görüntü akışı içindeki tespitinden sonraki adım ise yüze ait sınır işaretlerinin (ing. facial landmark) tespitidir. Bu işaretlerin tespiti CLM'nin (Constrained Local Model) özelleşmiş bir hali olan CLNF (Constrained Local Neural Field) ile gerçekleştirilmekte, yüz parçalarının eşleştirilmesi NU-RLMS (Non-Uniform Regularized Landmark Mean-Shift) ile optimize edilmektedir. Şekil 1, OpenFace metodolojisinin görsel bir tarifini sunmaktadır.



Şekil 1. OpenFace metodolojisi [12], [13].

Bu yöntemde CLM (Constained Local Model) ilk aşamada verilen örneklemden yola çıkarak bir şekil modeli oluşturmaktadır. Bu amaçla bir PDM (Point Distribution Model) kullanılmakta, örneğin göz kapağı gibi belirgin konumları içeren parçalar (ing. patch) oluşturulmaktadır. Sonraki adımda belirli bir x_i noktası ile eşleşen yüze ait sınır işaretlerinin olasılığını hesaplamak üzere eşleşme haritaları (ing. response maps) hesaplanmaktadır [12], [13]:

$$\pi(x_i) = C_i(x_i; I), \quad (1)$$

ki burada I görüntüyü, C_i ise değeri 0 (eşleşme yok) ile 1 (tam eşleşme) arasında olan lojistik regresyon kesişimini ifade etmektedir.

Parçaların belirlenmesinden sonraki aşama uygunluk (ing. fitting) probleminin çözümüdür. Bu aşamada RLMS

(Regularised Landmark Mean Shift) yöntemi, CLM parametrelerini yinelemeli biçimde güncellemektedir:

$$\arg [R(p0 + \Delta p) + \sum_1^n D_i(x_i; I)], \quad (2)$$

ki burada R değeri karmaşık deformasyonları düzenleştiren bir değerdir (ing. regularization term). D_i değeri I görüntüsü için x_i konumunun eşleştirme ölçütüdür. Bu adımın sonunda OpenFace uygulaması yüz için 68 adet işaret tespit edebilmektedir.



Şekil 2. Yüz üzerinde tespit edilen 68 işaret noktası.

Bu çalışma kapsamında gerçekleştirdiğimiz geliştirmeler ile kullanıcının İlgi Alanını (İA) isteğe göre belirlemesine olanak tanınmakta, yeni girdi parametreleri ile görselleştirmeler oluşturması sağlanmakta, ve kullanıcıya isteğe göre uyarlanmış yüz tespiti ve takibi için alternatif bir geliştirme seçeneği sunulmaktadır.

Yüz tespiti ve takibi, akan görüntüde yüz ve yüz parçalarının konumlarının otomatik olarak belirlenebilmesi için gerekli bir adımdır. Bu yöntem, sonraki aşamada yüze bakan tarafın göz hareketlerini yüz verisi üzerine yerleştirerek iletişim sırasındaki göz göze gelme ve göz kaçırma davranışlarının tespit edilebilmesine olanak sağlamaktadır.

Yukarıda yüzyüze iletişim sırasında sözlü olmayan iletişim kiplerinin ancak sözlü iletişim kipleri ile birlikte değerlendirildiğinde iletişimdeki sistematik yapıların ortaya çıkarılabileceği değerlendirilmiştir. Aşağıdaki bölümde iletişim sırasında kaydedilen konuşmanın bölütlenmesi ile ilgili yöntemler sunulacaktır.

III. KONUŞMANIN BÖLÜTLENMESİ

Konuşma, birbirinden ayrılması güç parçalardan oluşan bir ses akış sinyalidir. Yüz tanıma gibi, konuşma tanıma da aktif araştırma alanlarından birisidir. Konuşma ile ilişkili ses bölütleri arasındaki sınırların belirlenebilmesi konuşma tanımada en temel problem olarak değerlendirilebilir.

Konuşma analizi birbiriyle ilişkili iki temel yöntem grubunu içermektedir: konuşmanın bölütlemeye (ing. speech segmentation) dair yöntemler ve konuşmanın sinyal

özelliklerinden yola çıkılarak konuşmacının kimliği, cinsiyeti, arka plandaki gürültülerin türleri (örn. müzik, toplantı ortamı gürültüsü, vb.) gibi bir dizi bulgunun elde edildiği ayrıştırma (ing. diarization) yöntemleri.

Bu çalışma kapsamında konuşmanın bölütlenmesi ve ayrıştırılması birlikte ele alınmıştır. Bu amaçla açık kaynak kodlu, platform-bağımsız ve kullanıcı-bağımsız bir uygulama olan *CMU Sphinx*, *Open Source Speech Recognition Toolkit* konuşma tanıma sistemi uygulamaya entegre edilmiş, dinamik gözbakışı analizine imkan verecek biçimde geliştirilmiştir.

CMUSphinx uygulaması arka planda konuşma bölütleme ve ayrıştırma için *LIUM Speaker Diarization* uygulamasını kullanmaktadır. Bu uygulama ilk aşamada sesin güç spektrumunu temsil eden MFCC (Mel-frequency Cepstral Coefficients) özneliklerinin çıkarımını yapmaktadır.

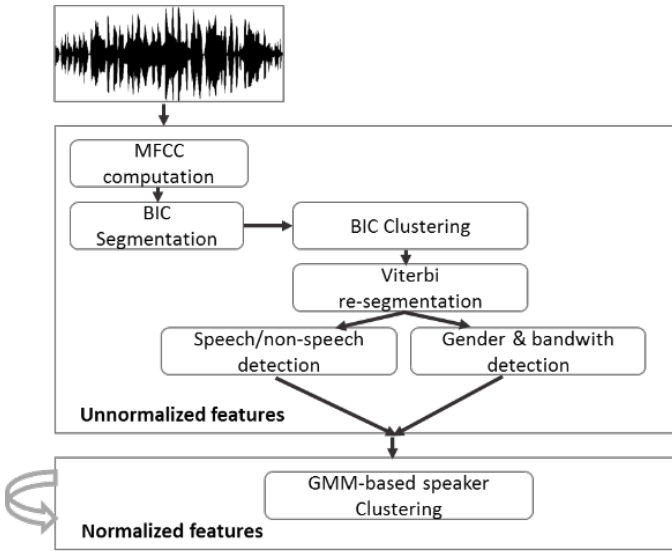
Sesin bölütlenmesi BIC (Bayesian Information Criterion) kullanılarak gerçekleştirilmektedir [14], [15]. Bölümlendirme iki adımdan oluşmaktadır. Birinci adımda olabilirlik ölçütü (ing. likelihood measure) yoluyla değişim noktaları tespit edilmekte (Generalized Likelihood Ratio, GLR), ikinci adımda ise aynı kullanıcıdan gelen ardışık segmentler biraraya getirilmekte [16] ve her segmente bir küme atanacak biçimde BIC hiyerarşik kümeleme işlemi yinelemeli olarak gerçekleştirilmektedir [15]. Her yinelemede isimleri i ve j olan iki ardışık küme için ΔBIC_{ij} değeri hesaplanmaktadır [16]:

$$\Delta BIC_{ij} = \frac{n_i+n_j}{2} \log|\Sigma| - \frac{n_i}{2} \log|\Sigma_i| - \frac{n_j}{2} \log|\Sigma_j| - \lambda P, \quad (3)$$

ki burada $|\Sigma_i|$, $|\Sigma_j|$ and $|\Sigma|$ değerleri i , j ve $(i + j)$ kümeleri ilişkili Gaussian belirteçleri (ing. determinants), n_i ve n_j , i ve j kümelerinin toplam uzunluğunu, λ kestirim kalitesini artırmayı amaçlayan bir yumuşatma (ing. smoothing) parametresini, P ise ceza katsayısını (ing. penalty factor) ifade etmektedir. Her ardışık küme için ΔBIC değerleri hesaplanmakta ve değer negatif olduğunda kümeler birleştirilmektedir.

Konuşma analizinin sonraki aşamasında bölütleme işlemi Viterbi algoritması kullanılarak tekrarlanmaktadır. Bu işlemde kümeler sekiz bileşenli bir GMM (Gaussian Mixture Model) ile temsil edilmektedir. GMM modeli, konuşmacı ayrışma işlemi için de kullanılmaktadır.

Şekil 3'te konuşmacı ayrıştırma ve konuşma bölütleme işlemlerinin iş akışı gösterilmektedir.



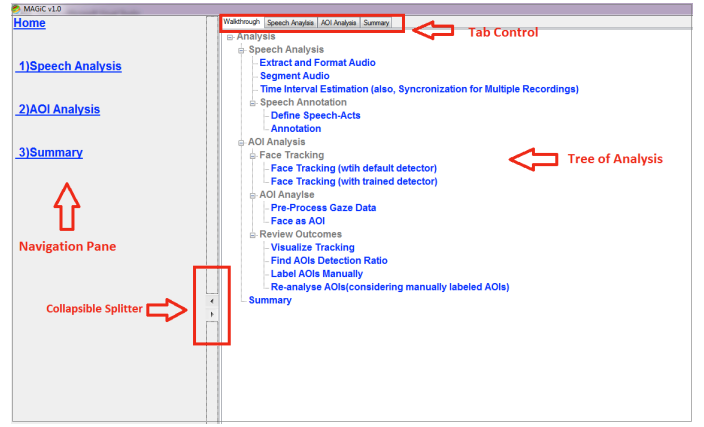
Şekil 3. Konuşmacı ayrıştırma ve konuşma bölümlendirme işlemlerinin iş akışı (LIUM Speaker Diarization Wiki sayfasından alınmıştır).

CMUSphinx uygulaması mevcut sürümünde bölütleme için öncelikle konuşmacının tespitini gerektirmekte, konuşmacı tespit edilemezse bölütleme yapmamaktadır. Diğer yandan yüzyüze konuşmada iki konuşmacı ve iki mikروفon bulunmaktadır. Bu çalışma kapsamında CMUSphinx uygulamasının kapsamı, bölütlenmemiş konuşma parçalarının tespiti ve bölütlenmesine olanak sağlayacak biçimde genişletilmiştir.

IV. YAZILIM MİMARİSİ VE BİLEŞENLERİ

Dinamik gözbakışının çokkipli bir çerçevede analiz edilebilmesi için geliştirdiğimiz uygulama bir masaüstü bilgisayar uygulaması olarak geliştirilmiş olup C# dilinde yazılmıştır. Uygulama, SoC (Separation of Concerns, Reade, 1989) tasarım prensiplerine göre geliştirilmiştir. Bu sayede uygulama kullanıcıların, sunulan fonksiyonların tamamını değil bir bölümünü de çalıştırabileceği, modüler bir yapıda tasarlanmıştır. Örneğin, uygulama sadece konuşma bölütleme ya da sadece yüz takibi için kullanılabilir.

Uygulama, son kullanıcıya programlama bilgisi gerektirmeyen bir arayüz sunmaktadır. Uygulamanın ana kontrol menüsünden erişilebilen dört adet modülü bulunmaktadır. Bu modüllerden Tam Çözüm Modülü'nün (ing. Walkthrough Module) temel amacı kullanıcıya uygulamanın işlevlerini içeren hiyerarşik bir ağaç yapısı sunmaktır. Bu yapının yaprakları kullanıcıya analizi adım adım anlatan bilgilendirme ekranlarını içermektedir. Uygulama, Tam Çözüm Modülü dışında üç temel modül daha içermektedir. Bu modüllerin işlevleri aşağıdaki bölümlerde açıklanmıştır. Şekil 4'te uygulama arayüzü ve açıklamaları gösterilmiştir.



Şekil 4. Uygulama arayüzü ve açıklamaları.

A. Konuşma Analizi Modülü (Speech-Analysis Module)

Kullanıcı arayüzünü oluşturan modüllerde analiz sırasında kullanılan ardışık fonksiyonlar akordiyon paneli şeklinde düzenlenmiştir. Konuşma Analizi Modülü de benzer biçimde aşağıdaki panelleri ve ilişkili fonksiyonları içermektedir.

- Sesin video dosyasından çıkarılması, formatlanması ve analize hazır hale getirilmesi için kullanıcıya bir panel arayüzü sağlanmaktadır.
- Konuşmanın bölütlenmesi için kullanılan panelde konuşmayı kelime-altı parçalara ayıran fonksiyonlar sunulmaktadır. Her bölütün süresi milisaniye olarak tutulmaktadır. Bölütler 0'dan başlayıp birer birer artırılan dosya isimleriyle kaydedilmektedir (örn. 199.wav).
- Zaman aralığı tahmin panelinde kullanıcı veri kaydının başladığı ve bittiği anları işaretlemekte ve arayüz iki konuşmacıdan gelen ses kaydının senkronizasyonunu sağlamaktadır. Yüzyüze konuşmada her konuşmacı için bir mikروفon kayıt almaktadır. Bu panelde kullanıcı veri kaydının başladığı anı kayıtları dinleyerek işaretledikten sonra uygulama iki kayıt arasındaki zaman farkını hesaplayarak kayıtları senkronize etmektedir. Bu adımdan sonra konuşma yeniden bölütlenerek (ing. re-segmentation) bölütleme kalitesinin artırılması sağlanmaktadır.

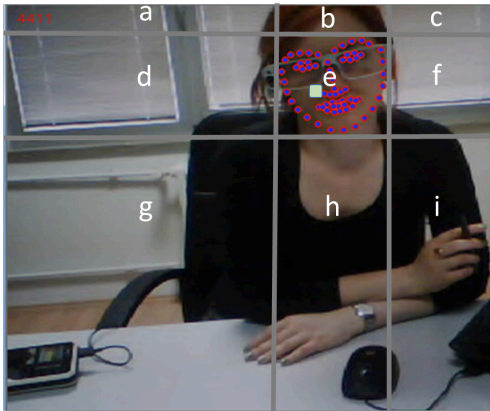
Konuşma işaretleme panelinde hangi sözeylem (ing. speech-act) etiketlerinin kullanılacağı kullanıcı tarafından belirlenmektedir. Bu kapsamda kullanıcı varolan etiket listelerini kullanabilmekte, bunları güncelleyebilmekte ya da kendi tanımladığı bir etiket listesi yükleyebilmektedir. Sonraki aşamada panel kullanıcıya sözeylem etiketlerinin işaretleme için kullanılacağı bir arayüz sunmaktadır.

Aşağıdaki bölümde arayüzün ikinci modülü olan İlgi Alanı Analiz Modülü tanıtılmaktadır.

B. İlgi Alanı Analiz Modülü (Area of Interest Analysis Module)

Bu modül kullanıcıya yüz izlemenin yapıldığı, İlgi Alanı (İA) analizinin yapıldığı ve çıktıların gözden geçirildiği üç panel ve ilişkili fonksiyonları sunmaktadır.

- Yüz izleme panelinde kullanıcı varsayılan bir yüz tanıma modelini kullanabileceği gibi kendi eğiteceği bir yüz tanıma modelini de kullanabilir. Varsayılan model, dlib kütüphanesinin sunduğu yüz tespit yöntemidir. Kullanıcı video görüntüsünü yüklemekte ve model yüzü tespit ettikten sonra takip edebilmektedir. Varsayılan yüz tanıma modeli düşük ışık miktarı gibi video kaynağından kaynaklı faktörlerden etkilenmekte, bu da kullanıcının hedeflediği doğruluk oranlarına erişememesine neden olabilmektedir. Bu nedenle arayüze modeli kullanıcının eğitebileceği fonksiyonlar eklenmiştir. Bu arayüzde kullanıcı videodan statik görüntüleri çıkarmakta, çıkarılan görüntüler arasında kullanıcı tarafından seçilenler ile model eğitilmektedir. Öğrenme modali olarak bir Destek Vektör Makinesi (Support Vector Machine, SVM) değişkesi kullanılmaktadır. Kullanıcı, bu modelin parametrelerini (örn. uç değerler için tolerans değerlerini kontrol eden C parametresini, ve risk boşluklarını belirleyen epsilon değerlerini) belirleyerek modeli eğiterek yüz tanıma doğruluğunu artırabilmektedir.
- İlgi Alanı (İA) analizi paneli, gözbakışı verisinin yüz görüntüsü üzerine yerleştirilmesini sağlamaktadır. Bu panelde kullanıcı görüntü üzerinde ilgi alanları belirleyebilmektedir. Panel fonksiyonları, kullanıcıya konuşmacının gözbakışının karşıdaki konuşmacının yüzüne odaklandığı ya da odaklanmadığı anlar hakkında bilgi vermektedir. Şekil 5'te örnek bir İA spesifikasyonu gösterilmiştir. Bu gösterimde görsel alan dokuz ilgi alanına ayrılmıştır. Konuşmacının gözbakışı resimde görüntülenen konuşmacının yüzüne odaklandığı İA, "e" ile belirtilmiş alandır. Gözbakışının diğer İA'larda olduğu durumlar ise göz kaçırma davranışını ifade etmektedir.



Şekil 5. İA spesifikasyonu. Yüzdeki işaret ("e" bölgesi) karşıdaki konuşmacının gözbakışı konumunu göstermektedir.

- İA modülü içinde sunulan üçüncü panel, kullanıcıya yüz tanıma ve İA analizinin çıktılarını gözden geçirme olanağı sağlamaktadır. Bu panelde yüz tanıma ve takip fonksiyonlarının çıktıları görselleştirilmekte, göz izleme cihazlarından kaynaklı sistematik kayma (ing. drift) hataları düzeltilebilmekte, yüz işaretlerinin konumlarının doğruluğu kontrol edilebilmekte, ve ilgi alanları (İA) kullanıcı tarafından el ile etiketlenebilmektedir.

C. Özet Modülü

Bu modülde kullanıcı konuşma analizi modülü ve ilgi alanı analizi modüllerinden elde edilen verileri tek bir çıktı dosyasında birleştirilmektedir. Çıktı dosyasındaki her satır video kaydının tek bir video karesini (ing. frame) temsil etmektedir. Her kare için, karenin video akışındaki süresi, sözeylem etiketi, konuşan kullanıcının kimliği, konuşmacıların yüzleri belirleyen dikdörtgen alanın koordinatları ve gözbakışı konumları sunulmaktadır. Bu veriler kullanılarak konuşma sırasında her iki konuşmacı için de gözbakışı davranışları belirlenebilmektedir.

V. PİLOT ÇALIŞMALARDA ELDE EDİLEN BULGULAR

Uygulamanın geliştirilmesinden sonra laboratuvar ortamında bir dizi pilot çalışma gerçekleştirilmiştir. Gözlük olarak giyilebilen iki göz izleme cihazının kullanıldığı (30 Hz, 56°x40° görsel ortam kayıt açısı) bir deneysel ortamda üç çift konuşmacıdan veri kaydı alınmıştır. Mevcut sürümde dinamik yüz tanıma doğruluk oranının %80, gözbakışı tespit ve eşleştirme doğruluk oranının %85 civarında gerçekleştiği gözlenmiştir. Verilerin işaretleme ve analizinde, uygulama kullanılmadan, elle yapıldığı durumlara kıyasla önemli ölçüde avantaj sağlandığı gözlenmiştir. Örneğin elle (kare kare) yapıldığında bir saat sürebilecek dinamik ortam gözbakışı konumu işaretleme işlemi saniyeler içinde tamamlanmıştır.

VI. SONUÇ

Dinamik ortamlarda gözbakışı analizi günümüzde henüz tam bir çözüm üretilememiş durumda olan gözbakışı analizi problemlerinden birisidir. Bu çalışma, dinamik gözbakışı analizi problemini, etkileşim ortamının niteliklerini önceden belirleyerek çözmeyi amaçlamıştır. Bu kapsamda, bilgisayar bilimleri alanında yoğunlukla araştırılmış olan yüz tanıma alanı problem çerçevesi olarak ele alınmış, yüzyüze diyalog ortamının analiz edilmesi için bir uygulama geliştirilmiştir. Geliştirilen uygulama, gözizleme alanında çalışan araştırmacıların kare kare yaptığı analizleri kolaylaştırmış, saatler cinsinden ifade edilen analiz sürelerini saniyeler mertebesine indirmiştir. Uygulamanın mevcut sürümüne Github ortamından erişilebilmektedir.¹

Uygulama, mevcut durumda ilk versiyonu ile sunulmaktadır. Sonraki versiyonlarda uygulamanın yüz

¹ <https://github.com/ulkursln/MAGiC/releases>, erişim tarihi 20 Nisan 2017.

dışındaki nesnelere de tanımayla olanak sağlayan kütüphanelerin entegre edilmesi yoluyla geliştirilmesi planlanmaktadır. Bu sayede dinamik gözbakışı analizinin kapsamının genişletilmesi hedeflenmektedir. Uygulamanın analizdeki doğruluk değerlerinin artırılacağı modüllerinden birisi konuşma bölütleme modülüdür. Mevcut sürümde dilden bağımsız olarak ses bölütleme işlemi yapan CMUSphinx uygulaması entegre olarak kullanılmaktadır. Türkçe'ye özel ses bölütleme araçlarının geliştirilmesi ve entegre edilmesiyle sözeylem işaretleme sürecinin kolaylaştırılacağı öngörülmektedir.

KAYNAKLAR

- [1] D. Archer, R. M. Akert, "Words and everything else: Verbal and nonverbal cues in social interpretation," *Journal of Personality and Social Psychology*, vol. 35, pp. 443-449, 1977.
- [2] A. Mehrabian, M. Wiener, "Decoding of inconsistent communications. *Journal of Personality and Social Psychology*," Vol. 6, pp. 109-114, 1967.
- [3] R. Bales, F. Strodtbeck, T. Mills, M. Roseborough, "Channels of communication in small groups". *American Sociological Review*, Vol. 16, pp. 461-468, 1951.
- [4] S. Duncan, "Some signals and rules for taking speaking turns in conversations," *Journal of Personal Social Psychology*, vol. 23, pp. 283-292, 1972.
- [5] J. Fasola, M. J. Mataric, "Using socially assistive human-robot interaction to motivate physical exercise for older adults," *Proceedings of the IEEE*, vol. 100(8), pp. 2512-2526, 2012.
- [6] U. J. Pfeiffer, B. Timmermans, G. Bente, K. Vogeley, L. Schilbach, "A non-verbal Turing test: differentiating mind from machine in gaze-based social interaction", *PLoS ONE*, <https://doi.org/10.1371/journal.pone.0027591>, 2011.
- [7] J. L. Austin. "How to do things with words", London: Oxford University Press, 1962.
- [8] J. R. Searle, "Speech Acts: An Essay in the Philosophy of Language", Cambridge University Press, Cambridge, UK. 1969.
- [9] K. Holmqvist, N. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, J. van de Weijer. "Eye tracking: a comprehensive guide to methods and measures", Oxford, UK: Oxford University Press, 2011.
- [10] S. M. Munn, L. Stefano, J. B. Pelz, "Fixation identification in dynamic scenes", *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization - APGV '08*, pp. 33-42, 2008.
- [11] S. Stuart, B. Galna, S. Lord, L. Rochester, A. Godfrey, "Quantifying saccades while walking: Validity of a novel velocity-based algorithm for mobile eye tracking," 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp.5739-5742, 2014.
- [12] T. Baltrušaitis, M. Mahmoud, P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic Action Unit detection". *Proceedings of Facial Expression Recognition and Analysis Challenge, IEEE International Conference on Automatic Face and Gesture Recognition*, 2015.
- [13] T. Baltrušaitis, P. Robinson, L. P. Morency, "OpenFace: an open source facial behavior analysis toolkit". *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, 2016. <https://www.cl.cam.ac.uk/~tb346/res/openface.html>, erişim tarihi 20 Nisan 2017.
- [14] C. Barras, X. Zhu, S. Meignier, J. L. Gauvain, "Multi-stage speaker diarization of broadcast news," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1505-1512, 2006.
- [15] S. Chen, P. Gopalakrishnan, "Speaker, environment and channel change detection and clustering via the bayesian information criterion," *DARPA Broadcast News Transcription and Understanding Workshop*, Landsdowne, VA, USA, 1998.
- [16] S. Meignier, T. Merlin, "LIUM SpkDiarization: an open source toolkit for diarization", *CMU SPUD Workshop*, pp. 1-6, 2010.