

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

VIDEO ANALİZİ İLE BAŞ HAREKETİ SINIFLANDIRMA

YÜKSEK LİSANS TEZİ

FİLİZ GÜRKAN

Elektrik ve Elektronik Mühendisliği Anabilim Dalı

Haberleşme Mühendisliği Programı

MAYIS 2015

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

VIDEO ANALİZİ İLE BAŞ HAREKETİ SINIFLANDIRMA

YÜKSEK LİSANS TEZİ

**Filiz GÜRKAN
(504131310)**

Elektrik ve Elektronik Mühendisliği Anabilim Dalı

Haberleşme Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Bilge GÜNSEL

MAYIS 2015

İTÜ, Fen Bilimleri Enstitüsü'nün 504131310 numaralı Yüksek Lisans Öğrencisi **Filiz GÜRKAN**, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “**VIDEO ANALİZİ İLE BAŞ HAREKETİ SINIFLANDIRMA**” başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Prof. Dr. Bilge GÜNSEL**

İstanbul Teknik Üniversitesi

Jüri Üyeleri : **Prof. Dr. Melih PAZARCI**

İstanbul Teknik Üniversitesi

Yrd. Doç Dr. Koray KAYABOL

Gebze Teknik Üniversitesi

Teslim Tarihi : **4 Mayıs 2015**
Savunma Tarihi : **27 Mayıs 2015**

Anneme ve Babama,

ÖNSÖZ

Yüksek Lisans eğitiminin boyunca her konuda bana yardımcı olan ve beni çalışmaya sevk eden danışman hocam Prof. Dr. Bilge GÜNSEL'e teşekkürlerimi sunarım. Bu aşamada her zorluğa düştüğümde her zaman olduğu gibi yanımda olup bana destek veren aileme, kardeşime ve bana hep güç katan SABİT'e desteklerinden dolayı teşekkür ederim. Tez aşaması sırasında bilgilerini benimle paylaşan İnci BAYTAŞ ve Erdem BAŞYURT arkadaşlarıma teşekkürlerimi sunarım.

Bu tezin oluşabilmesi için gerekli tüm çalışmalar İTÜ Elektronik ve Haberleşme Mühendisliği Bölümü bünyesinde yer alan "Çoğulortam Sinyal İşleme ve Örüntü Tanıma Laboratuvarı"nda Bilimsel Araştırma Projelerini destekleme kapsamında alınan bilgisayar ortamında gerçekleştirilmiştir.

Tez çalışması için İTÜ BAP 38155 numaralı proje desteği alınmıştır.

Mayıs 2015

Filiz Gürkan
Elektrik Elektronik Mühendisi

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	vii
İÇİNDEKİLER	ix
KISALTMALAR	xi
ÇİZELGE LİSTESİ.....	xiii
ŞEKİL LİSTESİ.....	xv
ÖZET.....	xvii
SUMMARY	xix
1. GİRİŞ	1
2. GRAF-TABANLI YÖNTEMLERDE OPTİK AKIŞ HESABI	7
2.1 Olasılıksal Yaklaşımla Optik Akış Modellenmesi	8
2.2 Enerji Enküçükmeye Dayalı Graf-Tabanlı Modeller	11
2.3 İnanç Yayılım Algoritması ile Optik Akış Kestirimi	12
3. SIFT AKIŞ YÖNTEMİ İLE OPTİK AKIŞ KESTİRİMİ	19
3.1 SIFT Özniteliklerinin Çıkarımı	19
3.2 SIFT Akış Kestirimi	21
3.3 Çiftli Markov Rastgele Alanı	22
3.4 SIFT Akış Algoritması Parametreleri Ve Etkileri.....	24
4. BAŞ HAREKET VE KONUM BİLGİSİ TÜMLEŞTİRİLEREK İLGİLİLİK SINIFLANDIRMA.....	33
4.1 Baş Dönüş Hareketi Sınıflandırma.....	35
4.1.1 Karar ağacı	38
4.1.2 Rastgele orman algoritması.....	39
4.2 Baş Konum Sınıflandırma	41
5. BAŞARIM TESTLERİ.....	45
5.1.1 Pointing'04 veri tabanında başarımlar	46
5.1.2 Test video kayıtlarında başarımlar	50
5.2 Baş Konum Sınıflandırma Başarım Testleri	51
5.3 Karar Tümlleştirme ile Baş Dönüş Hareketi Sınıflandırma Başarımı.....	52
6. SONUÇ VE ÖNERİLER.....	55
KAYNAKLAR	59
ÖZGEÇMİŞ.....	63

KISALTMALAR

SIFT	: Ölçeklemeyle Değişmeyen Öznitelik Dönüşümü (Scale Invariant Feature Transform)
MRF	: Markov Rastgele Alanı (Markov Random Field)
TRW	: Yeniden Ağırlıklandırılmış Ağaç (Tree re-weighted)
DS	: Doğru Tahmin Edilen Gözlem Sayısı
GS	: İlgili sınıftaki gerçek gözlem sayısı
TS	: İlgili sınıfta olduğu tahmin edilen gözlem sayısı
F1	: F ölçümü
HK	: Hareket Kestirimi
YS	: Yüz Sezme

ÇİZELGE LİSTESİ

Sayfa

Çizelge 3.1 : SIFT akış algoritması parametreleri	25
Çizelge 4.1 : Eğitim setinde yer alan açılar	35
Çizelge 4.2 : Eğitim ve test seti sınıf etiketleri	37
Çizelge 4.3 : Karar ağacı sınıflandırma eğitim veri seti	39
Çizelge 4.4 : Violen-Jones sınıf etiketleri	41
Çizelge 5.1 : Pan ve tilt hareketleri için elde edilen ortalama başarıml (F ölçütü)....	47
Çizelge 5.2 : Tilt açısı sabit tutularak pan hareketi için elde edilen ortalama başarıml (F ölçütü)	47
Çizelge 5.3 : Pan açısı sabit tutularak tilt hareketi için elde edilen ortalama başarıml (F ölçütü)	48
Çizelge 5.4 : Pointing'04 eğitim ve test seti ile pan hareketi için çerçeve bazında elde edilen baş hareketi ikili sınıflandırma başarıml (Test A)	49
Çizelge 5.5 : Pointing'04 eğitim ve test seti ile tilt hareketi için çerçeve bazında elde edilen baş hareketi ikili sınıflandırma başarıml (Test B)	49
Çizelge 5.6 : Pointing'04 veri tabanında eğitilen sınıflandırıcılar ile test video Kayıtlarında dönüş sınıflandırma başarıml (Test C).....	50
Çizelge 5.7 : Yüz konum sınıflandırma başarıml	51
Çizelge 5.8 : Violen-Jones ve SIFT akış öznitelikleri, tümleşik sınıf etiketleri	52
Çizelge 5.9 : Violen-Jones ve SIFT akış tümleşik başarıml (Test E).....	52
Çizelge 5.13 : Karar tümleştirme ve baş dönöl sınıflandırma başarıml (Ortalama F ölçütü).....	53

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1 : İnanç yayılımı için grafiksel model.....	16
Şekil 3.1 : Pointing04 veri tabanından alınan video çerçevesi ve karşı düşen renk kodlu SIFT görüntüsü	21
Şekil 3.2 : MRF gösterimi	23
Şekil 3.3 : İki katmanlı inanç yayılım algoritması	24
Şekil 3.4 : Akış vektörleri hesaplanan video çerçeveleri, SIFT ve optik akış ile hesaplanan akış vektörleri renkli gösterimi	25
Şekil 3.5 : Varsayılan parametre değerleri ile elde edilen SIFT akış sonuçları	26
Şekil 3.6 : α parametresinin etkisi	27
Şekil 3.7 : η parametresinin etkisi.... ..	28
Şekil 3.8 : nLevels parametresinin etkisi	29
Şekil 3.9 : Seviyelerde döngü sayısına göre maliyet fonksiyonu değişim grafikleri	30
Şekil 3.10:patchSize parametresinin etkisi	31
Şekil 4.1 : Kullanılan yöntemin blok diyagramı.....	34
Şekil 4.2 : Pointing'04 veritabanı örnekleri.....	36
Şekil 4.3 : Eğitim ve test setine ait video çerçeveleri	38
Şekil 4.4 : Eğitim setine uygun olarak oluşturulmuş karar ağacı... ..	39
Şekil 4.5 : Violen-Jones öznitelikleri	43
Şekil 4.6 : İntegral resminin öznitelik hesaplamadaki kolaylığı	43
Şekil 4.7 : Violen-Jones kademeli sınıflandırıcı	44
Şekil 4.8 : Violen-Jones yüz sezme algoritması çıktı örnekleri	44
Şekil 5.1 : Gerçek video çerçeveleri örnekleri	50

VIDEO ANALİZİ İLE BAŞ HAREKETİ SINIFLANDIRMA

ÖZET

Güncel çoğul ortam uygulamalarında gerçek zamanlı hareket analizi ve hareketli nesnelerin izlenmesi önem kazanmıştır. Hareketli nesnelerin gerçek zamanlı ve çevrim içi yakalanmasını gerçekleyen sistemlerde, ortamdaki ışıklılık değişimlerine, nesnelerin kameraya yaklaşmasından uzaklaşmasından kaynaklanan ölçek değişimlerine ve nesne hareketinden kaynaklı örtüşmeye karşı gürbüz algoritmaların geliştirilmesi son dönemde üzerinde çalışılan problemlerdir. Bir diğer önemli konu ise günlük yaşamda kullanılan kameralar ile yapılan 2 boyutlu video kayıtlarından hareket kestiriminin yüksek doğrulukla yapılabilmesidir.

Tez kapsamında, ekrana belli uzaklıktaki kişilerin baş pozisyonu ve duruşu kestirimi ile ekranda gösterilen içerik ile ilgili ya da ilgisiz olduklarını otomatik olarak sınıflandıran bir yöntem önerilmektedir. Bu amaçla video dizilerinden hareketli nesnelere ilişkin konum belirlenmesi ve iki boyutlu hareket analizi yapılmaktadır. Yapılan analizler sonucu elde edilen, baş hareketini ve konumunu belirleyen öznitelikler kullanılarak her iki yöntem için ayrı ayrı ve iki yöntem birlikte karar tümleştirme yapılarak hareket kestirimi, yani ilgililik-ilgisizlik sınıflandırılması yapılmaktadır.

Bu amaç için, kamera düzlemine paralel bulunan ve 0 derece açı ile ekrana bakan baş pozisyonu referans alınmak üzere ± 15 , ± 30 , ± 45 , ± 60 derecelik açılara sahip video çerçeveleri eğitim ve test seti olarak kullanılmakta, hareket akış vektörleri ve konum bilgileri yardımıyla ilgililik durumu kestirilmektedir.

Baş konum bilgisi, Violen Jones yüz sezici ile belirlenen baş bölgesinin, başlangıç pikseli konumu ve alan özniteliklerini içeren 3 boyutlu öznitelik vektörü ile modellenmektedir. Konum sınıflandırma, iki sınıflı bir sınıflandırma problemi olarak ele alınmakta ve baş konumunun başlangıç pikselleri x, y ve yüz alanı A öznitelikleri kullanılarak sezilen baş bölgesi ‘ortada’ veya ‘yanda’ olarak sınıflandırılmaktadır. Bu algoritmada baş konumu başlangıç pikselleri x, y ; başın ekrana uzaklığının yakın-uzak-ortada olma durumuna göre sırasıyla video çerçevesinin ortada yer alan %5, %12, %25’lik kısmının içinde yer alıyorsa ‘ortada’ olduğu kabul edilir. Amaçlanan ilgililik sınıflandırma olduğu için ‘ortada’ ve ‘yanda’ sınıfları sırasıyla ‘olası ilgili’ ve ‘olası ilgisiz’ sınıflarını temsil etmektedir.

Baş hareketlerini sınıflandırma, olasılıksal bir yöntem olan SIFT akış hareket kestirimi ile hesaplanan akış vektörleri yardımıyla yapılmaktadır. Burada akış vektörlerinin açı ve genlik bilgisini içeren 2 boyutlu öznitelik vektörleri kullanılmaktadır. Bu yöntemde başarıyı arttırmak amacıyla, video kareleri gri düzeyi bilgilerinin doğrudan kullanılması yerine, SIFT öznitelikleri ile hareket kestirimi yapılmaktadır. Kullanılan SIFT akış algoritması enerji enküçüklemeye dayalı bir yöntem olarak ele alınmakta ve enerji fonksiyonunu enküçükleyen değer akış vektörü olarak belirtilmektedir. Burada enküçükleme yöntemi olarak inanç yayılım algoritması kullanılmaktadır.

İlgililik sınıflandırma başarımını arttırmak amacıyla önerilen konum ve hareket sınıflandırma karar tümleştirme yöntemi için öncelikle konum sınıflandırma ile “orta” olarak etiketlenen video çerçeveleri bulunur. Daha sonra belirlenen bu video çerçeveleri için hareket kestirimi ve sınıflandırma gerçekleştirilerek baş hareketinin ± 15 derece ve ± 30 derece olarak belirlendiği durumlar ‘ilgili’, ± 45 ve ± 60 derece olarak belirlendiği durumlar ‘ilgisiz’ olarak etiketlenir. Baş konum sezme ile ‘yanda’ olarak etiketlenen video çerçeveleri ise hareket kestirimi yapılmaksızın ‘ilgisiz’ sınıfına atanmaktadır.

Karar tümleştirme ile ekrana ilgililik sınıflandırma başarımını hesaplamak için, hareket kestirimi eğitim kümesi olarak literatürde kullanılan Pointing’04 veritabanı kullanılmaktadır. Baş hareket açısının göz ile yüzde yüz olarak doğru tahmin edilmesi ve etiketlenmesi zor bir problem olduğu için, bu amaca yönelik olarak açı değerleri hesaplanıp etiketlenmiş bir veritabanı kullanılması gerekliliği doğmuştur. Baş konumu sınıflandırmada ise böyle bir problem ile karşılaşılması için amaca yönelik olarak kaydedilen gerçek videolar kullanılmaktadır.

Hareket kestiriminde başarımlar testleri öncelikli olarak, veritabanı eğitim ve test setleri ile yapılmaktadır. Bu amaçla, her açı ayrı bir sınıfa atanır ve bu sınıflar ikili olarak gruplandırılarak ikili(binary) eğitim-test sınıflandırma yapılır. Baş hareketi yatay doğrultuda olabileceği gibi düşey ve çapraz doğrultularda da olabileceği için her doğrultu için ikili sınıflandırma yapılarak ve başarımlar hesaplanmaktadır.

Veritabanı testlerinin yanı sıra hem hareket kestirimi hem de baş konum kestirimi sınıflandırma için çekilen gerçek videolar ile de testler yapılarak başarımlar raporlanmaktadır.

Sınıflandırma algoritması olarak Rastgele Orman algoritması kullanılmıştır. Yapılan denemelerde bu algoritmanın başarımlar ve hız açısından literatürde kullanılan algoritmalarından daha iyi olduğu saptanmıştır.

Başarımlar raporlama için ‘presicion’, ‘recall’ ve ‘F ölçütü’ parametrelerinden yararlanılmıştır.

Pointing04 veritabanı eğitim ve test veri setleri ile yapılan ikili sınıflandırma sonuçları ortalama olarak “pan” hareketi için %72, “tilt” hareketi için %62 ve çapraz doğrultularda %62 olarak hesaplanmıştır. Doğruluğun en yüksek olduğu “pan” hareketi için gerçek videolarla testler yapılmıştır ve yapılan testler sonucu görülmüştürki, gerçek videolarda sadece hareket kestirimi öznelikleri kullanılarak yapılan sınıflandırmada %62 olan ekranla ilgililik saptama başarımı (F ölçütü) öznelikleri tümleştirildiğinde %71’e yükseltilmiştir. Sadece baş konum sezme algoritmasıyla hesaplanan başarımlar ise %71’dir. Ancak bu algoritmada sınıflar “olası ilgili” ve “olası ilgisiz” olarak belirtildiği için tek başına istenilen ilgililik sınıflandırma problemi için yeterli değildir.

Önerilen tümleştirme algoritmasının ve SIFT akış yönteminin başarımlarını göstermek için gelecek çalışmalar olarak, farklı çözünürlük, farklı doğrultu ve farklı kişiler ile oluşturulmuş videolar ile testler yapılarak başarımların arttığı gösterilecektir.

HEAD ROTATION CLASSIFICATION VIA VIDEO ANALYSIS

SUMMARY

The problem of automatically measuring the engagement level of people is an area of great interest across a wide variety of fields. Engagement level measuring can be used to monitor drivers to guarantee safety of driving. Another application can be measurement of involvement levels of computer game players. In this thesis we focus on classification of head rotations to automatically measure the engagement level of viewers to the displayed video content.

In the literature a number of methods are proposed for measurement of engagement level. Some of them use expensive sensors such as eye trackers while some others require complicated installments. In this thesis we use an on the shelf Web camera with a simple setting. We define the problem as a two class classification problem, i.e., “involved in the screened content” or “not involved.” Engagement classification has been performed by integrating head location and head movement features. The head movement based classification is activated where the location detection fails. 2-D feature vectors that comprise amplitude and angle of the estimated optical flow vectors extracted by the SIFT flow algorithm are used for motion classification. Head location is represented by 3-D feature vectors that consist of the initial head coordinates detected by Viola-Jones face detector and area of the detected region.

SIFT flow is a probabilistic optical flow estimation method and it is preferred for motion estimation because the probabilistic methods are more powerful than gradient based methods. This mainly because probabilistic methods produce useful extensions of the standard quadratic gradient techniques for computing optical flow, including an automatic gain control mechanism, and the incorporation of a prior bias on the flow. It provides (two-dimensional) flow vector confidence information, allowing later stages of processing to weight their use of the vectors accordingly, and it provides a framework for properly combining flow information with probabilistic information from other sources.

We use images which have different head pose angle in different directions to report performance of engagement level classification of TV viewers. There are 8 angles, ± 15 , ± 30 , ± 45 , ± 60 degrees and 6 angles, ± 15 , ± 30 , ± 60 degrees of pan movement (x direction) and tilt movement (y direction), respectively. In addition to all these, there are different head pose combinations of pan and tilt angles for diagonal directions. Each angle is represented with one class for each direction. This means that there are 8 and 6 classes for pan and tilt movement. We classify head pose according to the angle classes. After classification we separate result classes in two groups (involved in the screened content or not), in order to report performance of engagement level classification. In this thesis, classification and performance reports are made in each direction separately.

SIFT flow algorithm is used to get motion features to classification of head movement. Before applying SIFT flow algorithm to the frames, we get SIFT image from the RGB image. SIFT is a local descriptor to characterize local gradient information. SIFT descriptor is a sparse feature representation that consists of both feature extraction and detection. But in the SIFT flow algorithm, only use the feature extraction component. Per pixel in SIFT image, is represented with 128 dimensional vector.

SIFT flow algorithm has an energy function similar to that of optical flow to estimate SIFT flow from two SIFT images. This energy function is designed to make flow vectors smooth, with discontinuities agreeing with object boundaries. Based on these two criteria, the energy function of SIFT flow is formulated. SIFT flow algorithm is a probabilistic based method as mentioned before. Therefore, minimum value of energy function (flow vector) is not calculated directly. On the contrary, probabilistic based optical flow, SIFT flow uses graph based methods which decrease complexity of the algorithm. To find the optimum solution of energy functions, tree-reweighted belief propagation algorithm is used. The optimum solution of that energy function, which is called “belief”, is flow vectors in x and y direction. In classification problem we do not use flow vectors directly. We evaluate angle and magnitude of vectors and these are 2-D features to classify head movement.

In this thesis, we discuss the problem as a binary classification. So in motion classification, we evaluated 28 binary datasets from 8 classes of the pan movement (x direction) and 15 binary datasets from 6 classes of tilt movement (y direction). To classify, each video frame classifies with each binary dataset. After classification, video frames are labeled in terms of angle class with majority voting. According to angle class labels, we decided involvement level as “involve” or “not involve”. Class ± 15 or ± 30 degree means “involve”, whereas class ± 45 or ± 60 degree means “not involve”.

Head location features are calculated through the instrument of Viola Jones face detector. Fast feature computation and efficient feature selection properties of the Viola - Jones algorithm is important advantages. We get 3-D feature which is x, y and A. x and y are the coordinates of beginning location and A is the area of the head location which is found with Viola-Jones face detector. Head location estimation problem is considered as a binary classification and location is classified “middle” or “side”. In Viola-Jones algorithm, according to value of area A, if beginning pixels of location x, y is located respectively in 5%, 12%, 25% of middle of image pixels, classify as a “middle”, otherwise “side”. In this thesis main goal is involvement level classification, so class “middle” represent “probable involve” and class “side” represent “probable not involve”.

In our thesis, we try to increase performance of classification of involvement level, involve or not involve. This is achieved by integrating head location and head movement features. Firstly, head location features are calculated. Video frames are classified according to these features. After that, motion estimation and classification is applied to video frames of which head locations can't be found and video frames are in class “middle”. With motion classification, if head movement is ± 15 degree or ± 30 degree, these frames are labeled as class “involve”. If head movement is ± 45 degree or ± 60 degree, these frames are labeled as class “not involve”. Motion estimation and classification is not applied to video frames, which are classified as

“side” with head location classification. These frames are labeled as “not involve” directly.

Pointing04 database is used as a training and test dataset for motion classification. This database consists of 15 sets of images. In each set, there are 2 series (one of them for training and one of them for test) of 63 images of the same person with varying pan and tilt movement. The reason why we used database for training set is the difficulty of specifying the class label of video frames visually. But it is not true for head location classification, so the recorded real videos are used for training set in head location classification. In addition to the Pointing04 test database, recorded real videos are also used for the test set.

As a classification method, Random Forest is preferred. Random forest is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution of all trees in the forest.

To report classification performance, accuracy has been the most common metric. But in real datasets where class distributions are not specific, a classifier that always predicts the most common class may deceptively yield high accuracy. So we used recall, precision and F1 score, which is the harmonic mean between precision and recall. Precision is the fraction of retrieved instances that are relevant while recall is the fraction of relevant instances that are retrieved.

First of all, motion classification test was made with Pointing04 train and test set. The results were calculated for each direction separately in terms of F1 score. For pan movement (x direction) and tilt movement (y direction) binary classification results were 72% and 62% on average, respectively. For diagonal movement (both x and y direction) binary classification result was 62% on average. In addition to these tests, for pan movement which gives best results in motion classification test was also made with recorded real test video. When only motion features were used for classification, performance was 62% in terms of F1 score. When motion estimation and face detector are applied and obtained features are integrated, performance increased to 71%. When only head location features are used, classification performance was 71%. But classification, which is made with only head location features, gives that whether a class label is “probable involved” or “probable not involved”. These classes are not enough to classify video frames involvement level.

As a future work, to show the success of integration of head location and head movement features and SIFT flow algorithm, we will make some test with different videos which have different resolution, different direction and different people. Besides, to increase the performance of classification we have the option of adding some different features in addition to 5-D features which we used now.

1. GİRİŞ

Günümüzde farklı kaynaklardan yapılan video yayınları hem içerik hem de uzunluk açısından izlenmesi zor boyutlara ulaşmıştır. İçerik sağlayıcılar izleyicinin ilgisini çekmek amacıyla teknolojik ve psikolojik yaklaşımları olabildiğince kullanmaktadır. Sağlanan içerik ile ilgililik düzeyinin ölçülmesi, raporlanması ve ilgililiği artırıcı teknolojilerin geliştirilmesi günümüz sistemlerinde oldukça önemlidir. İlgililiğin doğru bir şekilde belirlenebilmesi eğitim videolarının hazırlanmasından ürün algılamaya kadar birçok alanda önem kazanmıştır. Bu nedenlerle son dönemde örüntü analizi alanında izleyicinin ekrandaki içerik ile “ilgili” veya “ilgisiz” olma durumunun otomatik olarak izlenmesi problemi üzerinde çalışılmaktadır.. Bu amaçla ekran karşısındaki izleyicinin kamera ile kaydedilen görüntülerinden elde edilen öznitelikler ile sınıflandırma yapılarak ilgililik durumunun saptanması konusunda son dönemde literatürde çeşitli çalışmalar yayınlanmaktadır [1,2].

Tez kapsamında, baş pozisyonu ve duruşu kestirilerek ekrana belli uzaklıktaki izleyicilerin ekranda gösterilen içerik ile ilgili ya da ilgisiz olduklarını, otomatik olarak sınıflandıran bir yöntem önerilmektedir. Bu amaçla video dizilerinde hareketli nesnelere ilişkin konum belirlenmesi ve iki boyutlu hareket analizi yapılmaktadır. Öncelikle baş hareketini ve konumunu belirleyen öznitelikler çıkarılmaktadır. Bu öznitelikler ile, sadece konum ve sadece hareket bilgisi kullanılarak ve elde edilen sonuçlar üzerinde karar tümleştirme uygulanarak izleyicinin ekran ile ilgili ya da ilgisiz olduğu sınıflandırması yapılmaktadır. İzleyicinin 0 derece açı ile ekrana bakan baş pozisyonu referans alınmak üzere üzere, baş duruşunun ± 30 derece altında hareketlilik göstermesi ve baş konumunun ekranda belli bir alan kaplaması, ekranda gösterilen içerik ile ilgililik, bunların dışındaki durumlar ilgisizlik şeklinde sınıflandırılmaktadır. Çalışmada ekranda izlenen içeriğe ilgililik-ilgisizlik sınıflandırmada, hareket sınıflandırmanın konum bilgisi ile tümleştirilerek kullanılmasının başarımı arttıracığı gösterilmektedir.

İlgililik izleme birçok farklı uygulama için önemli bir işlemdir. Araba kullanımında sürücünün yola ilgililiğinin izlenmesi veya sanayide tehlikeli cihazların kullanımında

kişilerin ilgililiğinin izlenmesi ile güvenlik sağlanabilmektedir [3]. Bu tarz uygulamalarda ilgililiğin pahalı sensörlerle ölçülmesi ve bu sensörleri kişinin üstünde taşıması gerekliliği nedeniyle geliştirilen algoritmaların tezde ilgilenilen problemin çözümünde kullanılması pek olanaklı görünmemektedir. Bunun yanı sıra kullanılan baş hareketi kestirim algoritmalarında 3 boyutlu hareket kestirimi yapılması işlemsel karmaşıklıkğı arttırmaktadır.. Gerçek zamanlı hareket analizi ve hareketli nesnelerin izlenmesinde kullanılan algoritmaların ortamdaki ışıklılık değişimleri, nesnelerin ekrana uzaklaşıp yaklaşmasından kaynaklanan ölçek değişiklikleri ve nesne hareketinden kaynaklı örtüşmeye gürbüz olması gibi problemler üzerinde son dönemde araştırmalar yürütülmektedir.

Tez çalışmasında ilgililik izlemede kullanılan video verisinin toplanmasında günlük hayatımızda yer alan elde edilmesi kolay ve ucuz kameralar kullanılmaktadır. Ayrıca önerilen hareket kestirimi algoritması da doğası gereği ışıklılık ve ölçek değişimlerine karşı gürbüz bir yapıdadır.

Tez çalışması kapsamında problem ekrandaki içerik ile “ilgili”, “ilgisiz” şeklinde iki sınıflı bir sınıflandırma problemi olarak ele alınmış, sınıflandırmanın baş hareket ve baş konum öznitelikleri yardımıyla yapılması amaçlanmıştır. Benzer bir problem ile ilgilenilen [2]’de önerilen çözümde farklı öznitelikler üzerinde durulmuştur. Göz, burun ve ağız arasındaki açılar ile baş hareketinin ve ilgililik seviyesinin kestirilmesi çalışılmıştır. Bunun yanı sıra sınıflandırma yöntemlerinde de farklılıklar vardır. [2]’de SVM (destek vektör makinası) sınıflandırma yöntemi kullanılmıştır. Tez çalışmasında SVM sınıflandırıcılar kullandığımız Rasgele Orman (Random Forest) sınıflandırma yöntemi ile neredeyse aynı başarıyı sağladığı ve karmaşıklığı çok daha fazla olduğu için tercih edilmemiştir.

Bu tez çalışmasında sınıflandırma için kullanılacak öznitelikler “baş konum” sınıflandırmada kullanılan özniteliklerin çıkarılmasında Violen-Jones yüz sezme algoritması kullanılmıştır [4]. Violen-Jones algoritması video çerçevesi gri düzey farklılıklarına göre yüz piksellerini bulmaktadır. Tez kapsamında algoritmanın bulduğu yüz konum bilgisine göre baş konum x, y yüz başlangıç koordinatları ve A yüz alanı bilgisini içeren 3 boyutlu özniteliklerle modellenmektedir. Bölüm 5.3 de daha ayrıntılı olarak anlatılacak olan baş konum sezme algoritma sonucunda x, y ve A değerlerine göre, yüz konumu ekranın “ortasında” veya “yanında” olarak etiketlenir. Tezde yapılması planlanan, ilgililik sınıflandırma olduğundan, Violen-Jones

algoritmasında “orta” olarak etiketlenen sınıflar “olası ilgili”, “yan” olarak etiketlenen sınıflar “olası ilgisiz” olarak adlandırılır.

Baş hareket kestirimi algoritması olarak 2011 yılında MIT tarafından geliştirilmiş olan SIFT akış yöntemi kullanılmıştır [5]. Bu yöntemde kamera ile kaydedilen video çerçeveleri SIFT yöntemine benzer özniteliklerle modellenir [5,6] ve hareket kestirimi SIFT öznitelikleri üzerinden yapılır. SIFT öznitelikleri video çerçevesinde yer alan her bir pikselin 128 boyutlu bir vektör ile temsil edilmesiyle oluşur. Bu sayede SIFT akış hareket kestirim algoritmasının ışıklılık ve ölçek değişimlerine karşı daha gürbüz bir algoritma olması amaçlanmıştır [5]. Bölüm 4 de ayrıntılı olarak açıklanacak olan SIFT akış algoritması olasılık tabanlı bir hareket kestirim yöntemi sunmaktadır. Bu yöntemde video çerçevesine ait piksel SIFT öznitelikleri, Markov rastgele alanları ile modellenir . Problemin çözümü için, elde edilmek istenilen akış vektörlerinin marjinal dağılımlarının üstel bir fonksiyona uyduğu kabul edilir ve üstel enerji fonksiyonu olarak ifade edilir. MAP tahmin yöntemine göre marjinal dağılımı enbüyükleyen ve bu durumda enerji fonksiyonunu en küçükleyen akış vektörleri bulunur . Enküçükleme işlemi için kullanılan algoritma İnanç Yayılım Algoritmasıdır [5,9]. Bölüm 3.1.2’de ayrıntılı olarak anlatılacak bu algoritmada, pikseller (oluşturulan MRF şemasında düğümlere karşılık düşer) arası mesaj iletimi ile inanç olarak isimlendirilen ve akış vektörlerine karşılık gelen değerler hesaplanır [5,7] SIFT akış yönteminde akış vektörleri her video çerçevesi için hesaplanabilir. Hesaplanan x ve y yönlü akış vektörlerinden elde edilen açı ve genlik değerleri öznitelik olarak kullanılarak 2 boyutlu baş hareketi sınıflandırılması yapılır. Bu sınıflandırma sonucu video çerçeveleri x doğrultusu için, ± 15 , ± 30 ± 45 ± 60 derece hareket etiketli sınıflara y doğrultusu için ± 15 , ± 30 , ± 60 derece hareket etiketli sınıflara atanır. Çapraz yönler için x ve y doğrultularının kombinasyonu olacak şekilde baş pozisyonlarını gösteren sınıflar bulunmaktadır.

Tez çalışmasında, 384x288 boyutlarında bir çift video çerçevesi için akış vektörleri hesaplanması, 2.67 GHz Intel Core i5 CPU’lu bir bilgisayarda MATLAB programı ile yaklaşık 10 saniye sürmektedir. Bu süre şuan için gerçek zamanlı bir sınıflandırmaya olanak vermemektedir. Akış vektörleri şuan her bir video çerçevesi için hesaplanmaktadır. İleriki aşamalarda, hareket kestirimi belirlenen uygun sayıdaki video çerçevesi atlanarak daha az sayıda uygulanabilir. Bu işlemsel

karmaşıklığın ve sürenin azalmasını sağlayarak gerçek zamanlı bir uygulamaya olanak sağlar.

Bu çalışmada Violen-Jones algoritmasında baş konum başlangıç pikselleri x,y ve alan bilgisi A olmak üzere 3 ve SIFT akış algoritmasında akış vektörleri genlik ve açısı olmak üzere 2, toplamda 5 öznitelik kullanılmıştır.

Bu iki algoritma ile öznitelikleri elde edilen video çerçeveleri, SIFT akış algoritması için veritabanı ve Violen-Jones baş konum sezme algoritması için gerçek videolarla eğitilmiş ve Bölüm 5.2’de ayrıntılı olarak anlatılan Rastgele Orman Algoritması ile sınıflandırılmaktadır [10].

Amacımız baş hareket kestirim öznitelikleri ile gerçekleştirilen ekrana ilgililik, ilgisizlik sınıflandırma başarımını baş konum sezme ve baş hareket kestirimi algoritmalarını tümleştirerek arttırmaktır. Bu amaçla video çerçevelerinden öncelikle baş konum sezme algoritması ile konum bilgisi elde edilir. Baş hareketinin örtüşme yarattığı durumlarda veya baş konumunun video çerçevesinin dışına çıktığı durumlarda baş konum sezme algoritması sonuç bulamamaktadır. Baş konum sezme algoritmasının konum sezemediği video çerçevelerine doğrudan SIFT akış algoritması ile hareket kestirimi yapılmaktadır. Baş konum sezme algoritması ile konumun sezilebildiği durumlarda ise elde edilen öznitelikler ile sınıflandırılan ve baş konumu “yanda” olarak belirlenen video çerçeveleri, ekrana ilgisiz olduğu düşünülerek “ilgisiz” olarak etiketlenir. Asıl tümleştirme işlemi baş konum sezici öznitelikleri ile sınıflandırılmada “ortada” yani “olası ilgili” olarak etiketlenmiş video çerçevelerinde gerçekleşir. Bu video çerçeveleri ve baş konum sezici ile konum sezilemeyen video çerçeveleri için SIFT akış yöntemi ile hareket öznitelikleri bulunur ve ikili sınıflandırma yapılır. Sınıflandırma sonucu ± 15 derece ve ± 30 derece baş hareketi saptandı şeklinde etiketlenen sınıflar “ilgili”, ± 45 derece ve ± 60 derece baş hareketi saptandı şeklinde etiketlenen sınıflar “ilgisiz” sınıfına atanır.

Tez çalışmasında hareket kestirim ve sınıflandırma algoritmalarında kullanılan video çerçeveleri Pointing04 veritabanından alınmıştır [11]. Göz ile kafa baş duruşunun saptanarak etiketlenmesi zor olduğundan çekilen videolar ile eğitim seti oluşturmak yerine, veritabanından yararlanılarak, sonuçların doğruluğunun artırılması amaçlanmıştır. Pointing04 veritabanında x,y ve çapraz doğrultularda farklı bakış açılarında konumlandırılmış 15 farklı insanın farklı saç, aksesuar veya kıyafette

görüntülediği eğitim ve test setleri bulunmaktadır. Veritabanında yer alan her bir doğrultu ve açı için baş hareketi sınıflandırma başarımları hesaplanmıştır. Testler Pointing04 veritabanı eğitim seti, gerçek videolar test kaydı olarak kullanılarak yapılmıştır. Baş konum sınıflandırma algoritması eğitimi için her bir sınıfa uygun olarak hazırlanmış eğitim videoları kullanılmıştır. Bu videoların konum sezici ile öznetelikleri elde edilmiş, etiketlenmiş ve eğitim setleri oluşturulmuştur. Test verisi olarak hareket sınıflandırma algoritmasında da kullanılan gerçek videolardan yararlanılmıştır.

Baş hareketi sınıflandırmada “pan” şeklindeki hareket için 8 sınıf ve “tilt” şeklindeki hareket için 6 sınıf göz önüne alınmaktadır. Açığırkı bu sınıflandırmanın çok-sınıflı bir sınıflandırma işlemi olarak gerçekleşmesi başarımları düşürür. Öte yandan sınıf sayısının ikiye düşürülmesi de düşünülebilir, ancak bu durumda da varyans çok yüksek olacağından istenilen başarıma ulaşamayabilir. Bu nedenle tez kapsamında “pan” ve “tilt” kombinasyonu sınıflar ikili olarak sınıflandırmaktadır. 8 sınıfın ikili gruplandırılması sonucu elde edilen 28 eğitim veri setinin ve 6 sınıfın ikili gruplandırılması sonucu elde edilen 15 eğitim veri setinin her biri ile video çerçeveleri sınıflandırılmış, en çok karar verilen sınıf asıl sınıf olarak o video çerçevesine atanmıştır. Açısal olarak sınıflandırılma yapıldıktan sonra ise baş hareketinin ± 15 derece ve ± 30 derece olarak kestirildiği video çerçeveleri “ilgili”, ± 45 derece ve ± 60 derece olarak kestirilen video çerçeveleri “ilgisiz” olarak etiketlenmiştir.

Yapılan testlerin başarımlarının raporlanmasında literatürde sıklıkla kullanılan ‘presicion’, ‘recall’ ve ‘F ölçütü’ oranları kullanılmıştır. Bu ölçütler sınıfların doğru bulunma ve yanlış bulunma oranlarını göz önüne aldığından başarımların daha doğru bir şekilde raporlanmasını sağlarlar.

Başarımlarında Pointing04 veritabanında yer alan eğitim ve test setlerinin hareket kestirimi için ikili testler yapılmıştır. Bu testler “pan” hareketi, “tilt” hareketi ve her iki hareketi de içeren video çerçeveleri için yapılmıştır. Pan hareketi yani x doğrultusunda yer alan açılar için yapılan test sonuçlarında başarımlar F ölçütü cinsinden ortalama olarak %72 , tilt hareketi yani y doğrultusunda yer alan açılar için yapılan test sonuçlarında başarımlar ortalama olarak %62 olarak hesaplanmıştır.

Çapraz doğrultudaki sınıfların ortalama başarımı ise %62'dir. Bu başarımlar hesaplanırken aynı üst sınıfa ait ikili sınıfların başarımlarını hesaba katılmamıştır. Gerçek videolarda yapılan testlerde sadece hareket sınıflandırılarak %62 olan ekranla ilgililik saptama başarımı (F ölçütü), önerilen yüz konum sezme ve hareket kestirimi öznitelikleri tümleştirme yöntemi ile %71'e yükselmiştir.

Sonuçlar literatürdeki benzer çalışmalarla karşılaştırılırsa [2]'de sadece baş duruşu bilgisiyle yapılan sınıflandırmalarda başarımlar %62.37 olarak hesaplanmıştır. Baş duruş bilgisinin yanı sıra, göz burun ve ağız noktaları arası açı bilgisi sınıflandırmaya katılırsa başarımlar %73.57, genlik sınıflandırmaya katılırsa başarımlar %78.76 olmaktadır. [1]'de yapılan çalışmada ise eğitim seti olarak USF 3B yüz veritabanından elde edilen 2B video çerçevelerinin ve test seti olarak Pointing'04 veri tabanının kullanılmıştır. Burada yüze ait SIFT öznitelikleri çıkarılıp boyut küçükleme yöntemleri ile öznitelikler elde edilmiş, sınıflandırma ve regresyon yöntemleri kullanılarak 6.05°'lik hata ile baş hareketi saptanmıştır.

Bunun yanı sıra hareket kestirimi ve baş konum kestirimi algoritmalarından elde edilen özniteliklere ek olarak, farklı özniteliklerde eklenebilir. Örneğin bir sınıfta kalma sürelerine bakılarak çok kısa süreli sınıf değişikliklerinin hatalı olduğu varsayılabilir ve bu sınıflandırmalar düzeltilerek başarımlar artırılabilir. [2]'de belirtilen ve başarımlarını arttırdığı gözlemlenen yüz önemli noktaların arasındaki açı bilgilerini de öznitelikler arasına katılarak başarımların artırılabilceği düşünülmektedir.

2. GRAF-TABANLI YÖNTEMLERDE OPTİK AKIŞ HESABI

Görsel bir veriden hareket bilgisinin kestirilmesi önemli bir konudur ve literatürde önerilen birçok yöntemde video çerçevelerinden 2-Boyutlu (2-B) hareket kestirimi gerçekleştirilmektedir. Optik akış, hareket kestiriminde kullanılan en yaygın yöntemlerden birisidir. Optik akış ile 2-B hareket kestiriminde gradyan tabanlı eniyileme yöntemleri yaygın olarak kullanılır. Gradyen tabanlı yöntemlerle optik akış kestirimi yoğun hareket bilgisi sağlaması nedeniyle hareketin blok eşleme ve benzeri yöntemlerden daha doğru hesaplanabilmesine olanak tanır [12,13]. Öte yandan literatürdeki kestirimin bir eniyileme problemi olarak ele alındığı iteratif yöntemlerde [13] komşu pikseller ile ilişkinin kestirimde kullanılması doğruluğu arttırmaktadır. Açıklık (aperture) problemi ve performansın ayrık türev hesaplamada kullanılan komşuluk ilişkilerine oldukça bağlı olması bu yöntemlerin temel eksikliğidir. Literatürde son dönem çalışmalarında optik akışın olasılıksal modelleme ile enbüyük sonsal olasılık dağılımı (MAP) kestirimi problemi olarak tanımlanması ve çözülmesi önerilmiştir [8]. Bu sayede ışıklılık değişimi ve gürültünün bozucu etkileri modele dahil edilebilmekte ve akış kestirim başarımları arttırılmaktadır. Bu yöntemlerin temel sorunu MAP kestiriminin işlemsel karmaşıklığının yüksek olmasıdır. Bu problemi gidermek amacıyla MAP kestiriminde graf-tabanlı eniyileme algoritmaları kullanılmaktadır [5].

Tez çalışması kapsamında hareket kestirimi optik akış kestirimi problemi olarak ele alınmakta ve graf-tabanlı olasılıksal hareket modelleri kullanılarak MAP kestirimi inanç yayılımı algoritması ile gerçekleştirilmektedir. Graf tabanlı algoritmalar, rastgele değişkenler arası koşullu bağımlılık yapısını modelleyen ve eniyileme ile sonsal olasılık dağılımını enbüyükleyen öngörü (MAP) yapılmasını sağlayan yöntemlerdir. Graf tabanlı modeller esnek olmaları, gösterimlerindeki kolaylık ve büyük ağlardaki yüksek performansları nedeniyle son yıllarda yaygınlık kazanmıştır. Günümüzde kullanımı en yaygın olan graf tabanlı modeller yönlü ve yönsüz graf modelleri olarak iki ana gruba ayrılabilir [14,15]. Tez kapsamında yönsüz graflar kullanılarak videodan hareket kestirimi problemi ile ilgilenilmektedir. Bu kapsamda graflar, video

çerçevelerindeki herbir piksel grafa ilişkin bir uç (node) olarak alınarak oluşturulur. Hareket kestirimi, bu yönsüz grafa karşı düşürülen maliyet fonksiyonunun eniyilenmesi ile gerçekleştirilmektedir [16,17]. İstenilen çözüme farklı eniyileme algoritmaları ile ulaşılabilir. Bunların başlıcaları graf kesim (graph cut) ve inanç yayılım (belief propagation) algoritmalarıdır [16]. Tez kapsamında enerji tabanlı yönsüz graflar ile model oluşturulmuş ve inanç yayılım algoritmasını temel alan SIFT akış yöntemi [5] kullanılmıştır.

Bu bölümde öncelikle optik akış kestiriminin olasılıksal modeli gradyen tabanlı yöntemler ile ilişkilendirilerek formüle edilmektedir. Ardından graf-tabanlı yöntemler ile optik akış kestirimi anlatılmaktadır.

2.1 Olasılıksal Yaklaşımla Optik Akış Modellenmesi

$f(x,y,t)$ t anında (x,y) pikselinin gri düzeyini göstermek üzere, optik akış yöntemi ile hareket kestiriminde video çerçevesindeki (x,y) pikselinin gri düzeyinin konuma ve zamana bağlı değişiminin sadece hareketten kaynaklandığı varsayılır ve bu nedenle 2.1 denklemindeki optik akış kısıtının sağlandığı kabul edilir [12,13].

$$\mathbf{f}_s \cdot \mathbf{v} + f_t = 0 \quad (2.1)$$

2.1 denkleminde görülen $\mathbf{v} = [v_x, v_y]$ optik akış vektörünü, $f_t(x,y)$ pikselinin $f(x,y,t)$ gri düzeyinin zaman değişkenine göre türevini göstermektedir. f_x ve f_y sırasıyla $f(x,y,t)$ 'nin x ve y yönlerindeki birinci ayrı türevlerini göstermek üzere, denkleminde $\mathbf{f}_s = (f_x \ f_y)^T$ vektörel gösterimi kullanılmaktadır.

2.1 denkleminin çözümünde gradyen tabanlı optik akış yöntemi eşitlik 2.2 de görülen karesel hatayı enküçüklemeyi gerçekler [12,13].

$$E(\mathbf{v}) = [\mathbf{f}_s \ \mathbf{v} + f_t]^2 \quad (2.2)$$

2.2 deki karesel hata eşitliğinin \mathbf{v} ye göre türevi sıfıra eşitlenirse eşitlik 2.3 elde edilir.

$$\nabla E(\mathbf{v}) = \mathbf{M} \cdot \mathbf{v} + \mathbf{b} = 0 \quad (2.3)$$

Denklemin 2.3 de görülen \mathbf{M} , $\mathbf{f}_s \cdot \mathbf{f}_s^T = (f_x^2 \ f_x f_y; \ f_x f_y \ f_y^2)$ matrisini temsil etmekte iken \mathbf{b} vektörü $(f_x f_t \ f_y f_t)^T$ vektörünü göstermektedir.

Görüldüğü gibi 2.3 eşitliğinden $\mathbf{v} = [v_x, v_y]$ optik akış vektörünün çözülmesi için tersi alınması gereken M matrisi tekil matristir, tersi alınamaz. Bu problemi aşabilmek için 2.3 eşitliğine bazı ek terimler eklenmesi önerilmiştir. Horn ve Schunk [13] 1981 yılında yayınladıkları bir çalışma ile 2.2 de görülen hata fonksiyonuna ek bir düzgünlük kısıtlaması getirerek bu problemi çözmüştür.

Optik akış hesaplamada önerilen başka bir çözüm ise problemi kestirim problemi olarak göz önüne almaktır. Kestirim problemi olarak ele alındığında amaç, ∇f optik akış vektörünün video çerçevesinin x, y doğrultuları ve t zamanına göre gradyanını göstermek üzere, $P(\mathbf{v}|\nabla f)$ şeklinde elde edilen sonsal olasılık dağılımını en büyükleyen optik akış vektörünün bulunması olarak ifade edilebilir [8].

$P(\mathbf{v}|\nabla f)$ koşullu olasılığını elde etmek için 2.1 eşitliğindeki optik akış kısıtlamasından yararlanabiliriz. Ancak 2.1 eşitliği ortam ideal yani gürültüsüz olarak kabul edilerek yazılmıştır. Pratikte kamera gürültüsü, örtüşme ve bozucu etkilerden kaynaklı hatalar söz konusudur. Ölçümler hatasız yapılabilsen bile 2.1 de bahsedilen eşitlik; ışıklılık koşullarındaki değişiklik ve yansımadan kaynaklı bozulmalar nedeniyle tam olarak sağlanamaz. Bu iki tip bozulmayı modele katmak için eşitliğimize denklem 2.4'te görülen bağımsız Gauss gürültü terimleri \mathbf{n}_1 ve \mathbf{n}_2 eklenmelidir [8].

$$\mathbf{f}_s \cdot (\mathbf{v} - \mathbf{n}_1) + f_t = n_2 \quad n_i \sim N(0, \lambda_i) \quad (2.4)$$

Denklem 2.4'deki \mathbf{n}_1 ışıklılık değişimleri, yansıma vb. hataları modellerken, \mathbf{n}_2 kamera gürültüsü, örtüşme ve benzeri ölçüm hatalarını modellemektedir ve her ikisi de Gauss dağılımı olarak kabul edilir.

2.4 teki eşitlik $P(f_t|\mathbf{v}, \mathbf{f}_s)$ koşullu olasılığını tanımlamaktadır. Amacımız optik akış vektörü \mathbf{v} 'yi bulmak olduğundan $P(\mathbf{v}|f_t, \mathbf{f}_s)$ koşullu olasılığını elde etmek gerekmektedir. O yüzden Bayes eşitliğinden yararlanılarak 2.5 teki eşitlik yazılır .

$$P(\mathbf{v}|f_t, \mathbf{f}_s) = \frac{P(f_t|\mathbf{v}, \mathbf{f}_s) \cdot P(\mathbf{v})}{P(f_t)} \quad (2.5)$$

Bu eşitlikte $P(\mathbf{v})$ akış vektörünün önsel dağılımını göstermektedir ve 0 ortalamalı λ_p varyanslı Gauss dağılımı olduğu varsayılmıştır [8]. $P(f_t)$ ise \mathbf{v} ye bağlı bir değişken içermediğinden normalizasyon terimi olarak kullanılır.. 2.5 de verilen eşitlikte, eşitliğin sağ tarafında çarpılan iki dağılımın da Gauss kabul edilmesi nedeniyle

açıkça görülmektedir ki ulaşılmak istenilen $P(\mathbf{v}|f_t, \mathbf{f}_s)$ dağılımı da Gauss dağılımına uymaktadır. Bu eşitlikte görülen harekete ilişkin önsel dağılım eşitlik 2.6 daki üsttel dağılımla modellenenir. Olabilirlik dağılımı ise 2.7 deki gibi modellenenir [8].

$$P(\mathbf{v}) \propto \exp(-\frac{1}{2} \cdot \mathbf{v}^T \lambda_p^{-1} \cdot \mathbf{v}) \quad (2.6)$$

$$P(f_t|\mathbf{v}, \mathbf{f}_s) \propto \exp(-\frac{1}{2} \cdot (\mathbf{f}_s^T \mathbf{v} + f_t)^T \cdot (\mathbf{f}_s^T \lambda_1 \mathbf{f}_s + \lambda_2)^{-1} \cdot (\mathbf{f}_s^T \mathbf{v} + f_t)) \quad (2.7)$$

Sonsal dağılım 2.8'deki gibi Gauss dağılımlı olarak bulunur.

$$P(\mathbf{v}|f_t, \mathbf{f}_s) \propto \exp(-\frac{1}{2} \cdot (\mathbf{v} - \mu_v)^T \cdot \lambda_v^{-1} \cdot (\mathbf{v} - \mu_v)) \quad (2.8)$$

2.5, 2.6, 2.7 ve 28 denklemleri kullanılarak 2.9 eşitliği yazılabilir.

$$\exp(-\frac{1}{2} \cdot (\mathbf{v} - \mu_v)^T \cdot \lambda_v^{-1} \cdot (\mathbf{v} - \mu_v)) = \exp(-\frac{1}{2} \cdot \mathbf{v}^T (\mathbf{f}_s \mathbf{f}_s^T \lambda_1 \mathbf{f}_s + \lambda_2)^{-1} \mathbf{f}_s^T + \lambda_p^{-1}) \cdot \mathbf{v} - f_t (\mathbf{f}_s^T \lambda_1 \mathbf{f}_s + \lambda_2)^{-1} \mathbf{f}_s^T \cdot \mathbf{v} - \frac{1}{2} f_t \mathbf{f}_s^T \lambda_1 \mathbf{f}_s + \lambda_2)^{-1} f_t \quad (2.9)$$

Eşitleme sonucu kestirilen optik akış vektörünün ortalama ve varyansı 2.10 eşitlikleri ile ifade edilebilir.

$$\lambda_v = (\mathbf{f}_s (\mathbf{f}_s^T \lambda_1 \mathbf{f}_s + \lambda_2)^{-1} \mathbf{f}_s^T + \lambda_p^{-1})^{-1}$$

$$\mu_v = -\lambda_v \mathbf{f}_s (\mathbf{f}_s^T \lambda_1 \mathbf{f}_s + \lambda_2)^{-1} f_t \quad (2.10)$$

$P(\mathbf{v}|f_t, \mathbf{f}_s)$ sonsal olasılığını enbüyükleyen optik akış vektörü MAP tahmin yöntemine göre bulunur ve bu en büyükleme işlemi ve 2.11 eşitliği ile formüle edilir. Burada eşitlik 2.5 te de belirtilen Bayes kuralına göre $P(f_t)$ normalizasyon terimi olarak alınıp eşitliğe katılmamıştır

$$\operatorname{argmax}_v P(\mathbf{v}|f_t, \mathbf{f}_s) = \operatorname{argmax}_v (P(f_t|\mathbf{v}, \mathbf{f}_s) \cdot P(\mathbf{v})) \quad (2.11)$$

Sonuçta MAP tahmin yöntemi ile bulunan \mathbf{v} optik akış vektörü 2.12 deki gibi gösterilebilir.

$$\mathbf{v}_{\text{MAP}} = \operatorname{argmax}_v (P(f_t|\mathbf{v}, \mathbf{f}_s) \cdot P(\mathbf{v})) \quad (2.12)$$

Bölüm 2.1 de bahsedildiği gibi 2.12'deki çarpım, dolayısıyla optik akış vektörlerinin dağılımı, Gauss dağılımına sahiptir. Gauss dağılımında açıkça görülebilirki fonksiyonu maksimum yapan değer ortalama, yani μ_v vektördür. Sonuç olarak söylenebilirki 2.10'daki μ_v vektörü $P(\mathbf{v}|f_t, \mathbf{f}_s)$ fonksiyonunu maksimum yapan optik akış vektörü, yani MAP çözümüdür [8].

Bulduğumuz $P(\mathbf{v}|f_t, f_s)$ olasılığının ortalama değeri olan μ_v değeri tek bir piksele aittir. Tüm video çerçevesinde optik akış vektörleri hesaplanmak istenildiğinde her bir piksel bağımsız kabul edilerek komşu piksellere göre ortalama değer vektörleri hesaplanmalıdır. Bu vektörler yaklaşık olarak optik akış vektörlerine eşit olacaktır.

Probleme olasılıksal olarak bakmanın gradyan tabanlı metodlara göre bazı üstünlükleri vardır. Değerlerin dağılımsal olarak yazılması optik akışın doğasında bulunan bazı belirsizliklerin gösterilebilmesine olanak sağlar. Bu zorlukların en başında dağılımların başlangıç parametrelerini seçmek gelir. Seçilen başlangıç parametrelerine göre kestirilen vektörler değişiklik göstermektedir [8].

2.12 eşitliğiyle formülize edilen akış vektörlerinin MAP kestiriminin bulunması yüksek işlemsel karmaşıklıktır. Bu nedenle teknik yazında işlemsel karmaşıklığı daha düşük eniyileme yöntemleri kullanılmaktadır. Bu yöntemler bölüm 2.2’de anlatılmaktadır.

2.2 Enerji Enküçülemeye Dayalı Graf-Tabanlı Modeller

Bölüm 2.1 de eşitlik 2.5’teki dağılımların Gauss dağılımı olduğu kabul edilmiş ve MAP tahmin yöntemine göre optik akış vektörünün μ_v ‘ye yakınsadığı analitik olarak gösterilmiştir. Tez kapsamında kullanılan graf-tabanlı eniyileme yaklaşımında literatüre uygun olarak, önsel ve sonsal dağılımların Gibbs dağılımı olduğu varsayılmıştır [18]. Gibbs dağılımına uyan optik akış vektörleri için, önsel olasılıksal yoğunluk fonksiyonu 2.13 eşitliğinde verilmektedir.

$$P(\mathbf{v}) = P(v_1, v_2, \dots, v_{m \times n}) = \exp(-E(\mathbf{v})) \quad (2.13)$$

Bu eşitlikte $P(\mathbf{v})$ optik akış vektörlerinin önsel dağılımını, $E(\mathbf{v})$ önsel dağılımın Gibbs dağılımına uygunluğunda kullanılan enerji modelini göstermektedir. Amacımız akış vektörünü elde etmek olduğundan, ‘y’ gözlemlenen video çerçevesi bilgisi olarak adlandırılırsa, $P(\mathbf{v}|\mathbf{y})$ sonsal olasılığı eşitlik 2.5 te belirtilen Bayes eşitliğinden de yararlanarak eşitlik 2.14’deki şekilde iki olasılık yoğunluk fonksiyonunun çarpımı olarak yazılabilir. 2.14 eşitliğinde Gibbs dağılımlı önsel ve olabilirlik dağılımlarının çarpımı sonucunda oluşan üstel terim iki dağılıma ilişkin enerjilerin toplamı şeklinde görülmektedir.

$$P(\mathbf{v}|\mathbf{y}) \propto P(\mathbf{v}, \mathbf{y}) = P(\mathbf{v}).P(\mathbf{y}|\mathbf{v}) = \exp(-(E(\mathbf{v}) + E(\mathbf{y}|\mathbf{v}))) \quad (2.14)$$

Eşitlik 2.12’de de bahsedilen MAP öngörü sonucuna benzer bir şekilde optik akış vektörünü elde etmek için 2.14 eşitliğini maksimum yapan hareket vektörünün hesaplanması gerekir. Üstel negatif durumda olduğu için dağılımı en büyükleyen akış vektörü değerinin, üsteli en küçükleyen değer olduğu görülebilir.

2.14 eşitliğindeki $E(\mathbf{v})$ terimi akış vektörlerinin önsel olasılık dağılımına ilişkin enerjiyi modeller. Bu nedenle literatürde ‘şekillendirme’ ya da ‘yumuşatma’ terimi olarak adlandırılır [19]. $E(\mathbf{y}|\mathbf{v})$ terimi ise görüldüğü gibi gözlemlenen verinin olabilirlik fonksiyonunu belirler ve literatürde ‘veri terimi’ olarak adlandırılır. Bu durumda 2.14 eşitliğindeki üstel terimi genel olarak 2.15 eşitliğindeki gibi iki toplamsal terim ile ifade edilebiliriz [17,19].

$$E(\mathbf{v}|\mathbf{y})=E_{veri}(\mathbf{v})+E_{yum}(\mathbf{v}) \quad (2.15)$$

2.15 eşitliğinde görülen enerji terimlerinin tanımı kullanılacak modele ve probleme göre değişkenlik gösterebilir. Kullandığımız enerji fonksiyonu ayrıntılı olarak bölüm 3.2 de açıklanacaktır. Bunun yanında genel olarak söyleyebiliriz ki, veri terimi gözlemlenen veri ile tahmin edilen veri arasındaki uyumlulukla ilgilidir. Yumuşatma terimi ise belirlenen komşuluk seviyesinde akış vektörünün komşular arası uyumluluğunu kontrol eder [5]. Açıkça 2.15 eşitliğinin çözümü olan akış vektörlerinin hesaplanması olası tüm durumlar için enerji hesabını ve en küçük enerji durumunun belirlenmesini gerektirir. İşlem yükü nedeniyle kestirimde bir eniyileme algoritması kullanılır. Bu sayede 2.15’de görülen bu fonksiyon, en uygun etiket, yani en uygun \mathbf{v} akış vektörü için enküçükleme olur. Enküçükleme için farklı algoritmalar önerilmiştir [16]. Bunların en çok kullanılanlarının başında inanç yayılım (belief propagation) algoritması ve graf kesim (graph cut) algoritması gelir. Tez kapsamında inanç yayılım algoritması kullanıldığından ileriki altbölümlerde buna ilişkin ayrıntılı formülasyon sunulmaktadır.

2.3 İnanç Yayılım Algoritması ile Optik Akış Kestirimi

Şekil 2.1’de basit bir örneği görülen graf modeli düğümlerden oluşur [14]. Tez çalışmasında ilgilenilen problemde düğümler her bir pikseli temsil eder. Bulunmak istenilen optik akış vektörleri olduğundan optik akışa ilişkin Bölüm 2.1’de açıklanan olasılıksal dağılımlar kullanılır.

Graf tabanlı bir modeli oluşturan düğümlerin ortak olasılık fonksiyonu, faktör olarak adlandırılan fonksiyonların çarpımı olarak yazılabilir. Tez kapsamında bu fonksiyonların akış vektörleri ve piksel değerlerinin olasılıksal fonksiyonlarının üstelleri olduğu varsayılır. Herbir faktöre ilişkin dağılımın (olasılık fonksiyonlarının) Gibbs olması durumunda logaritmik domende ortak olasılık eşitlik dağılımı 2.14'deki gibi üstellerin toplamı şeklinde yazılabilir. Denklem 2.16'da f , g , h fonksiyonları faktör olarak adlandırılır. Bu eşitlikte a, b, c değişkenlerinin f ; a, b değişkenlerinin g ; d değişkeninin ise h faktörüne bağlı olduğu varsayılmıştır [20].

$$P(a, b, c, d) = \exp (f(a, b, c) + g(c, d) + h(d)) \quad (2.16)$$

2.16 eşitliğini, eşitlik 2.14 ile karşılaştırdığımızda $E(\mathbf{v})$ ve $E(\mathbf{y}|\mathbf{v})$ fonksiyonlarına faktörlerin karşılık geldiğini görebiliriz.

Bu ortak yoğunluk fonksiyonunu maksimum yapan parametreleri bulmak zor bir eniyileme problemidir. Bu işlem iteratif bir şekilde mesaj iletim algoritmaları ile çözebilir [7]. Eniyilemede amaçlanan enerji olarak ifade edilen üstel terimi enküçükleyen ve bu sayede ortak olasılığı enbüyükleyen çözümün elde edilmesidir. Bu algoritmalar literatürde enerji enküçüklemeye dayalı graf tabanlı modeller olarak adlandırılır [16]. Yapılan tez çalışmasında eniyileme için inanç yayılım algoritmasının bir versiyonu olan Ağaç Yeniden Ağırlıklandırma (Tree Reweighted-TRW) algoritması kullanılmış olup bu bölümde algoritmanın detayları anlatılmaktadır [21,22].

İnanç yayılım algoritması ilk olarak Judea Pearl tarafından 1982 yılında graflara karşı düşen ağaç gösterimlerinde optimum çözümü hızlı hesaplamak amacıyla ortaya atılmıştır [9]. Matematiksel notasyon ile ifade edildiğinde, $\mathbf{m}_{ij}(\mathbf{v}_j)$ i pikselinden j pikseline gönderilen ve j . piksele ait akış vektörü bilgisini içeren bir mesaj vektörü olarak tanımlansın. Mesaj kısaca, i pikselinin j pikseline, o piksele ait hareket/akış vektörünün ne olduğu hakkında düşüncesini gönderdiği bir bilgidir [23]. Bu nedenle mesaj vektörü akış vektörüyle aynı boyuttadır.

İnanç yayılım algoritmasında başlangıç mesajı rasgele üretildikten sonra iteratif olarak güncellenerek yakınsanılan mesaj hesaplanır ve kestirilmek istenilen akış vektörünün sağladığı $P(\mathbf{v}_i|\mathbf{y})$ marjinal dağılımı bu mesajlar aracılığı ile elde edilir. Marjinal dağılım i pikseli için $\mathbf{b}_i(\mathbf{v}_i)$ olarak gösterilir, inanç olarak adlandırılır [14]. Hareket kestiriminde kullanılan video çerçevelerinin her bir pikseline ait gri düzeyleri

ve bu piksellere karşı düşen akış vektörleri ayrı ayrı raslantı değişkenleri ile modellenir. Herbir video çerçevesi i piksel içermekte ise, gözlem kümeleri sırasıyla $\mathbf{y}=\{y_1, y_2, \dots, y_i\}$ ve $\mathbf{v}=\{v_1, v_2, \dots, v_i\}$ olarak yazılabilir. Bu gösterimle v_i ve y_i değişkenleri modelimizde i . düğüme karşı düşmektedir. Eşitlik 2.14'de gösterilen ve faktörlerin çarpımı olarak yazılabilen olasılık yoğunluk fonksiyonu gösterimi hareket kestirim problemine uyarlandığında denklem 2.17'de görülen eşitlik elde edilir. Denklemde görüldüğü üzere, $P(\mathbf{v}|\mathbf{y})$ koşullu olasılık dağılımına ulaşılmakta kullanılan $P(\mathbf{v}, \mathbf{y})$ ortak olasılık dağılımı faktör çarpım ile ifade edilmiştir.

$$P(\mathbf{v}|\mathbf{y}) \approx P(\mathbf{v}, \mathbf{y}) = \frac{1}{Z} \prod_{(ij)} \Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j) \prod_{(i)} \Phi_i(\mathbf{v}_i, \mathbf{y}_i) \quad (2.17)$$

Eşitlikte Z normalizasyon terimi ve (ij) komşu pikselleri temsil etmektedir. Tez çalışmasında kullanılan modelde en yakın komşuluklu Markov Raslantı Alanı (MRF) modeli ele alındığından en yakın 4 komşu piksel üzerinden işlem yapılmaktadır [5].

Denklem 2.17'de görülen $\Phi_i(\mathbf{v}_i, \mathbf{y}_i)$ her piksele karşı düşen y_i ve v_i değerleri arasındaki uyumluluk fonksiyonudur (faktör) [14]. Komşu piksellerin ortak hareket edeceği varsayımı ile komşu piksellerin hareketlerinin uyumluluğunu kontrol eden bir $\Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j)$ hareket uyumluluk fonksiyonu (faktör) da eşitlikte görülmektedir. Piksellere ait gözlemlerin birbirinden bağımsız kabul edilmesi varsayımı altında eşitlikte görülen çarpım ifadelerine ulaşılır ve ortak yoğunluk fonksiyonu eşitlik 2.17'deki gibi elde edilir [14].

Eşitlik 2.14 ve eşitlik 2.17 ilintilendirildiğinde $P(\mathbf{v}) = \prod_{(ij)} \Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j)$ ve $P(\mathbf{y}|\mathbf{v}) = \prod_{(i)} \Phi_i(\mathbf{v}_i, \mathbf{y}_i)$ yazılabileceği görülmektedir. Bir başka deyişle inanç yayılım algoritmasında istenilen akış vektörlerinin önsel dağılımı ve akışın piksel gri düzeylerine bağlı koşullu dağılımı uyumluluk fonksiyonları ile modellenmekte ve iteratif olarak güncellenen mesajlar ile $P(\mathbf{v}|\mathbf{y})$ sonsal dağılımı (marjinal dağılım) kestirilmeye çalışılmaktadır.

Ψ_{ij} ve Φ_i fonksiyonları Bölüm 2.2.1'de tanımlandığı üzere Gibbs fonksiyonu olarak kabul edilir ve $J_{ij}(\mathbf{v}_i, \mathbf{v}_j) = \ln \Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j)$, $h_i(\mathbf{v}_i, \mathbf{y}_i) = \ln \Phi_i(\mathbf{v}_i, \mathbf{y}_i)$ olarak tanımlanırsa eşitlik 2.18 elde edilir [14].

$$\ln P(\mathbf{v}|\mathbf{y}) \approx - \sum_{(ij)} J_{ij}(\mathbf{v}_i, \mathbf{v}_j) - \sum_{(i)} h_i(\mathbf{v}_i, \mathbf{y}_i) \quad (2.18)$$

Eşitlik 2.18'deki toplamsal terimler eşitlik 2.15'te gösterilen enerji fonksiyonlarına karşı düşmektedir.

İnanç yayılım algoritmasında akış vektörüne karşılık gelen $P(\mathbf{v}|\mathbf{y})$ olasılığı i . piksel için hesaplanan inanç terimi olarak adlandırılır ve eşitlik 2.19 ile tanımlanır [14].

$$\mathbf{b}_i(\mathbf{v}_i) = P(\mathbf{v}_i|\mathbf{y}_i) = k \Phi_i(\mathbf{v}_i, \mathbf{y}_i) \prod_{(j \in N)} \mathbf{m}_{ji}(\mathbf{v}_i) \quad (2.19)$$

Bu eşitlikte k normalizasyon terimi ve N ise i 'nin komşuluklarını temsil eder. Denklem 3.5 ile denklem 3.7 karşılaştırıldığında k normalizasyon teriminin $1/z$ ye karşılık geldiği görülebilir.

2.19'da k normalizasyon terimi ve N ise i pikselinin komşuluklarını temsil eder.

Denklem 2.17 ile denklem 2.19 karşılaştırıldığında k normalizasyon teriminin denklem 2.17'deki $1/Z$ 'ye karşılık geldiği görülebilir. 2.19'da görülen $\prod_{(j \in N)} \mathbf{m}_{ji}(\mathbf{v}_i) = \prod_{(j \in N(i))} \Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j)$ olduğu görülebilir. Burada inancın hesaplanması için gerekli olan ve bilinmeyen değer mesaj değişkenidir ve eşitlik 2.20'de verilmektedir. Bu eşitlik iteratif olarak ulaşılan mesaj bilgisidir. İlk durumda mesaj değişkenine rastgele bir ilk değer atanır. Daha sonra gerekli iterasyon sayısı sonunda eşitlik 2.20'deki değere yakınsanır [14] [23].

$$\mathbf{m}_{ij}(\mathbf{v}_j) \leftarrow \max_{\mathbf{v}_i} \Phi_i(\mathbf{v}_i, \mathbf{y}_i) \Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j) \prod_{(k \in N \setminus j)} \mathbf{m}_{ki}(\mathbf{v}_i) \quad (2.20)$$

Eşitlik 2.20'de görüldüğü gibi i ye j den gelen mesajı hesaplamak için, j hariç diğer komşularından gelen mesajların çarpımını alırız.

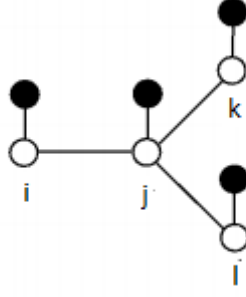
Bu sistemde i, j pikseline ve ona komşu pikselleri kapsayan yeni bir düğüm noktası olsun. Bu durumda i düğümündeki inancı hesaplamak istersek eşitlik 2.21'den yararlanabiliriz [14].

$$\mathbf{b}_i(\mathbf{v}_i) = k \Phi_i(\mathbf{v}_i, \mathbf{y}_i) \prod_{(j \in N(i))} \mathbf{m}_{ji}(\mathbf{v}_i) \quad (2.21)$$

2.21'de $\mathbf{m}_{ji}(\mathbf{v}_i)$, i düğümüne komşularından gelen mesajlardır. Burada j tek bir pikseli değil komşu pikselleri temsil eder.

Eşitlik 2.21 deki inanç formülasyonunun marjinal olasılığa eşitliğini göstermek için şekil 2.1'deki gibi bir grafiksel modelden yararlanabiliriz [14] .

i . pikseldeki inancı hesaplamak için eşitlik 2.20 ve eşitlik 2.21'den yararlanarak eşitlik 2.22 yazılabilir.



Şekil 2.1: İnanç yayılımı için grafik model.

$$\mathbf{b}_i(\mathbf{v}_i) = k \Phi_i(\mathbf{v}_i, \mathbf{y}_i) \mathbf{m}_{ji}(\mathbf{v}_i)$$

$$\mathbf{b}_i(\mathbf{v}_i) = k \Phi_i(\mathbf{v}_i, \mathbf{y}_i) \sum_{\mathbf{v}_j} \Phi_j(\mathbf{v}_j, \mathbf{y}_j) \Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j) \mathbf{m}_{kj}(\mathbf{v}_j) \mathbf{m}_{lj}(\mathbf{v}_j) \quad (2.22)$$

Sonuçta i. pikseldeki inanç şekil 2.1 için eşitlik 2.23 ile hesaplanır.

$$\begin{aligned} \mathbf{b}_i(\mathbf{v}_i) &= k \Phi_i(\mathbf{v}_i, \mathbf{y}_i) \sum_{\mathbf{v}_j} \Phi_j(\mathbf{v}_j, \mathbf{y}_j) \Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j) \dots \\ &\dots \sum_{\mathbf{v}_k} \Phi_k(\mathbf{v}_k, \mathbf{y}_k) \Psi_{jk}(\mathbf{v}_j, \mathbf{v}_k) \sum_{\mathbf{v}_l} \Phi_l(\mathbf{v}_l, \mathbf{y}_l) \Psi_{jl}(\mathbf{v}_j, \mathbf{v}_l) \end{aligned} \quad (2.23)$$

$\mathbf{v}_k, \mathbf{v}_j$ ve \mathbf{v}_l üzerinden toplam aldığımızda inanç marjinal olasılığı eşitlik 2.23' deki şekilde hesaplanabilir [14].

$$\mathbf{b}_i(\mathbf{v}_i) = k \sum_{\mathbf{v}_j, \mathbf{v}_k, \mathbf{v}_l} p(\{\mathbf{v}\}) = p_i(\mathbf{v}_i) \quad (2.24)$$

Yukarıdaki formülasyon tek bir düğüm (piksel) için verilmiştir. Bahsettiğimiz gibi inanç yayılım algoritması graf model olarak bütün video çerçevesindeki görüntüyü kullanır, bir başka deyişle tek bir piksel için tanımlanan işlemler tüm pikseller için gerçekleşir. İşlemsel karmaşıklığı düşürmek amacıyla, tez kapsamında video çerçevesinin alt parçalara ayrılmasıyla elde edilen görüntü parçaları kullanılmaktadır [22].

Son dönemde işlemsel karmaşıklığı düşürmek amacıyla eniyileme Ağaç Yeniden Ağırlıklandırma (TRW) algoritması ile yapılmaktadır. TRW Bölüm 2.2'de bahsettiğimiz inanç yayılım algoritmasının hızlandırılması amacıyla önerilmiş olan özelleşmiş halidir [16,21,22]. Problemi TRW açısından yeniden özetlersek, enerji en küçükleme metotlarında ana amaç, \mathbf{Y} video çerçevesini modelleyen grafi göstermek üzere, $E(\mathbf{v}) = \min_{\mathbf{v}} E(\mathbf{v}|\mathbf{Y})$ olan akış vektörünü bulmaktır. \mathbf{Y} grafi eşitlik 2.25 deki gibi alt graflara, yani alt ağaçlara bölünebilir [21].

$$\mathbf{Y} = \mathbf{Y}^1 + \mathbf{Y}^2 \dots + \mathbf{Y}^n \quad (2.25)$$

Bu durumda minimum enerji için n grafın bölündüğü parça sayısı olmak üzere eşitlik 2.26 yazılabilir [22,24].

$$E(\mathbf{v}^i) = \min_{\mathbf{v}^i} E(\mathbf{v}^i | Y^i) \quad i = 1 \dots n \quad (2.26)$$

Bu eşitliklere göre enerji fonksiyonunu en küçükleyen akış vektörünü bulmak için $E(\mathbf{v}^i)$ değerlerinin birlikte kullanılması gerekmektedir.

Hareket kestirim problemi göz önüne alındığında i video çerçevesine ait bir piksel olmak üzere i pikselini içeren tüm alt ağaçlarda inanç yayılım algoritmasını uyguluyoruz. Daha sonra bulunan sonuçların ağırlıklı ortalaması bize yakınsanan akış vektörü sonucunu verir. Burada kullanılan temel algoritma yine inanç yayılım algoritmasıdır. Ancak algoritma tüm video çerçevesinden oluşturulmuş bir graf yapısında işletilmek yerine parçalara ayrılmış alt graflarda işletilir. Bu ise sistemin hızının iki katına varan düzeylerde artmasını sağlar [21].

3. SIFT AKIŞ YÖNTEMİ İLE OPTİK AKIŞ KESTİRİMİ

Tez kapsamında Ölçeklemeyle Değişmeyen Öznitelik Dönüşümü (SIFT - Scale Invariant Feature Transform) akış algoritması ile Bölüm 2’de formüle edilen olasılıksal modelleme kullanılarak video verisinden optik akış kestirimi yapılmaktadır. SIFT akış algoritması 2004 yılında MIT tarafından geliştirilmiştir [5] ve TRW BP yöntemiyle enerji en küçükleme yapılarak akış vektörlerin kestirimini gerçekler. Algoritma birçok hareket kestirim yönteminden farklı olarak video çerçevesinin gri düzeyi bilgisini değil SIFT görüntülerini giriş verisi olarak kullanır. Bu bölümde SIFT akış algoritması ile gerçekleştirilen optik akış kestirimi anlatılmaktadır.

3.1 SIFT Özniteliklerinin Çıkarımı

SIFT algoritması 2004 yılında Lowe tarafından geliştirilmiştir ve bir görüntüdeki özniteliklerin belirlenmesi ve eşleştirilmesinde kullanılır [6]. Tez kapsamında algoritma tek bir video çerçevesini giriş olarak alır ve bu görüntünün özniteliklerini belirler. Öznitelikler konum, ölçek, yönelim bilgisi ve bunların istatistiklerine ilişkin bilgi içermektedir ve her video çerçevesi için belirli anahtar noktalarda hesaplanır. Bu anahtar noktalarda özniteliklerin bulunmasında ilk adım video çerçevesinin eşitlik 3.1 deki gibi farklı σ varyans değerlerindeki Gauss filtrelerinden geçirilerek farklı ölçeklerde yumuşatılmış görüntülerin elde edilmesidir [25]. Burada i, j piksel koordinatlarını simgeler.

$$g_{\sigma}(i, j) = G_{\sigma}(i, j) * f(i, j) \quad (3.1)$$

Filtreleme sonucu elde edilen yumuşatılmış görüntülerin 3.2’de formüle edilen şekilde piksel bazında farkları alınarak DOG (Gaussların farkı) filtrelemesi gerçekleştirir .

$$g_{\sigma_1}(i, j) - g_{\sigma_2}(i, j) = (G_{\sigma_1}(i, j) - G_{\sigma_2}(i, j)) * f(i, j) \quad (3.2)$$

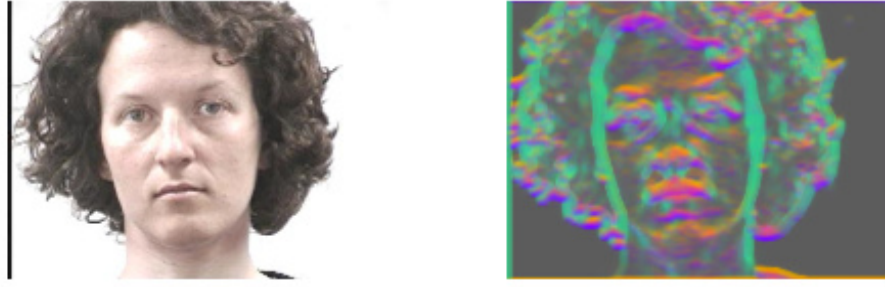
DOG filtreleme sonrası her piksel gri düzeyi kendine komşu 8 piksel ve komşu ölçeklerdeki 9 piksel yani toplam 26 piksel gri düzeyi ile karşılaştırılır. Bu

karşılaştırma sonucu eğer o piksel gri düzeyi o bölgenin yerel maksimum ya da minimumu ise piksel anahtar olmaya aday olarak işaretlenir ve piksel koordinatları ile birlikte işlem yapılan ölçek bilgisi σ değeri de kaydedilir [6,26]. Aday anahtar noktaların sayısı uygulanan ek eşikleme işlemleri ile azaltıldıktan sonra kesinleştirilen anahtar noktaların konum ve ölçekleri kaydedilir.

Bir sonraki aşamada görüntü dönme etkisine karşı bağımlılığı ortadan kaldırmak amacıyla, her anahtar noktaya yerel özelliklerine bağlı yönelim bilgisi atanır. Her anahtar noktanın, merkezinde kendisinin olduğu ve komşu pikseller ile gradyan yönelim bilgisi içerdiği bir alanda olduğu düşünülürse, görüntüde oluşacak bir dönmenin sebep olduğu çevre gradyan yönelimleri her bileşen için eşit dönmeye sahip olacaktır ve bu şekilde dönme bozulmasına dayanıklılık sağlanacaktır [26].

Sonuçta her bir piksele atanan SIFT öznitelik vektörü 128 boyutludur., Anahtar noktanın kendisi, 45'lik 8 farklı yönelim açısı ve 4 x 4 lik örnekleme alanı içerisindeki bilgi tek bir vektörde toplandığından $8 \times 4 \times 4 = 128$ boyutlu öznitelik vektörleri elde edilir [6,26]. Bu öznitelikler bakış açısına, ışıklılık değişimine ve ölçek değişimine karşı dayanıklıdır. Bu yüzden hareket vektörlerini belirlerken video çerçevelerinin gri düzey bilgisi yerine SIFT özniteliklerini kullanmak doğruluğu artırır.

SIFT akış yönteminde amaç hareket kestirimi olduğundan, görüntü eşleme amacıyla geliştirilen klasik SIFT yönteminden farklı olarak, öznitelik vektörleri belirli anahtar noktaları için değil video çerçevesine ait her piksel için hesaplanır ve akış kestiriminde kullanılır. Bir video çerçevesinin RGB uzayında $n \times m \times 3$ boyutlu olduğunu göz önüne alırsak, SIFT öznitelikleri çıkarılmadan önce her pikselde R-G-B değerlerin ortalaması alınıp maksimum ortalama ile normalize edilerek normalize görüntü elde edilir [5]. Bu sayede video çerçevesinden en yüksek düzeyde bilgi elde edilebileceği düşünülmüştür. Ardından $n \times m \times 128$ boyutlu SIFT gösterimi elde edilir. SIFT özniteliklerinin modelleme yeterliliği hakkında bir fikir vermesi açısından 128 boyutlu SIFT vektörleri temel bileşen analizi kullanılarak 3 boyuta indirildiğinde. elde edilen $n \times m \times 3$ 'lük gösterim renk kodlaması ile ifade edilirse Şekil 3.1'de görülen gösterim elde edilir. Görüldüğü gibi SIFT gösterimleri video çerçevesinin görsel içeriği hakkında yüksek oranda bilgi içermektedir.



Şekil 3.1 : Pointing04 veri tabanından alınan video çerçevesi ve karşı düşen renk kodlu SIFT görüntüsü.

3.2 SIFT Akış Kestirimi

İki video çerçevesinin SIFT gösterimlerini kullanarak en doğru akış vektörünü elde etmek için optik akış algoritmasına benzer bir maliyet fonksiyonu tanımlanabilir [5,27]. Bölüm 2’de anlatılan olasılıksal modellem ile MAP kestirimi tanımlamalarına uygun olarak maliyet fonksiyonu bir enerji fonksiyonu olarak tanımlanıp enerjiyi en küçükleyen değerler kestirilen akış vektörleri olarak adlandırılır.

Enerji fonksiyonu oluşturulurken, akış vektörlerinin nesne sınırlarında süreksizlik göstermesi koşuluyla düzgün olarak dağılması gerekliliği göz önüne alınmıştır. Eşitlik 3.3 de görülen enerji fonksiyonunda, $i=(x,y)$ piksel koordinatları, $\mathbf{v}(i)=(x(i),y(i))$ i noktasındaki akış vektörü, s_1 ve s_2 ise ardarda gelen iki video çerçevesinde i . piksele karşı düşen SIFT vektörlerini göstermektedir[5].

$$E(\mathbf{v}) = \sum_p \min (||s_1(i) - s_2(i + \mathbf{v}(i))||, t) + \sum_p \eta(|x(i)| + |y(i)|) + \sum_{(i,j) \in \epsilon} \min(\alpha|x(i) - x(j)|, d) + \min(\alpha|y(i) - y(j)|, d) \quad (3.3)$$

Bu eşitlikte t ve d e bulunduğu toplamsal terimde eşik değeri görevi yapan sabitlerdir. Yani toplamların belirli bir değerün üstüne çıkmasını engellerler. ϵ p pikseline ait komşu piksellerin kümesini göstermektedir. SIFT akış algoritmasında komşuluk en yakın 4 komşu piksel olarak tanımlıdır [5].

Eşitlik 2.15 de bahsedildiği gibi enerji fonksiyonu veri ve düzgünleştirme olmak üzere iki ana terimden oluşur. Bu terimler sırasıyla önsel dağılıma ve marjinal dağılıma karşılık gelmektedir. Bu dağılımların Gibbs dağılımı olduğu varsayılırsa eşitlik 3.4 deki şekilde enerji fonksiyonu Gibbs dağılımının üsteli olur ve çarpım durumunda üsteller toplam şeklinde ifade edilir.

$$P(\mathbf{v}).P(\mathbf{y}|\mathbf{v}) = \exp(-(E(\mathbf{v}) + E(\mathbf{y}|\mathbf{v}))) \quad (3.4)$$

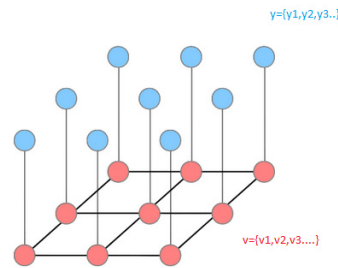
Bu eşitlikte gördüğümüz gibi terimlerden bir tanesi giriş verisi ile (veri terimi) bir diğeri ise akış vektörü ile (düzgünleştirme terimi) ilgilidir.

Eşitlik 3.3 ve 3.4 birlikte değerlendirildiğinde, eşitlik 3.3'deki SIFT akış enerji fonksiyonuna bakıldığında birinci terimin veri ile ilgili terim olduğunu diğeri iki terimin ise akış vektörüne kısıtlar getirdiği yani düzgünleştirme terimi olduğu görülür [5]. Eşitlik 3.3 de ilk terim olarak gözlemlediğimiz veri terimi, SIFT gösterimlerinin akış vektörüne göre uyumluluklarını modeller. Gözlemlenen SIFT gösterimi ile tahmin edilen akış vektörüne göre oluşturulan SIFT gösteriminin L1 normuna göre uzaklığına göre hatabulunarak uyumluluk hesaplanmış olur.

Eşitlik 3.31 de ikinci terim olarak gözlemlediğimiz düzgünleştirme terimi, komşu piksellerde yer alan akış vektörlerinin birbirine olan farkını sınırlamak içindir. Bu terimde farklar ayrıştırılmış olarak yani x ve y vektörü kendi içinde hesaplanmıştır. Bu enerji fonksiyonunu en küçükmek için kullanılan algorithmada kolaylık sağlamak amacı ile yapılmıştır. Eşitlik 3.3 de üçüncü terim olarak gözlemlediğimiz terim, akış vektörünün x ve y yönündeki bileşenlerinin birbirinden çok farklı olmasını engellemek içindir. Eşitlik 3.3 de görülen enerji fonksiyonu bir sonraki başlıkta anlatılan Markov Rastgele Alan (MRF) graf modeline göre oluşturulmuştur.

3.3 Çiftli Markov Rastgele Alanı

Markov rastgele alanı (MRF) literatürde en yaygın kullanılan olasılıksal grafik modellerin arasında yer almaktadır [16]. MRF yapısı şekil 3.2 de görüldüğü gibi video çerçevesinden gözlemlenen bir y_i değişkeni ve tahmin edilmek istenilen v_i değişkenlerini temsil eden düğümlerden oluşur. Alt indis olarak verilen i bir pikseli temsil edebildiği gibi, bir piksel grubunu da gösterebilecek şekilde seçilebilmektedir. Tez kapsamında ilgilenilen hareket kestirim probleminde v_i değişkeni i . pikseldeki akış vektörünü y_i ise gözlemlenen video çerçevesinden elde edilen SIFT öz niteliklerini temsil eder.



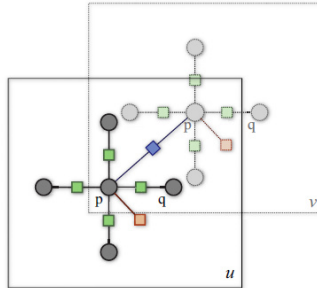
Şekil 3.2: MRF gösterimi.

MRF yapısında, her i değeri için \mathbf{v}_i ve \mathbf{y}_i arasında olasılıksal bir bağımlılık olduğu varsayılır ve bu $\Phi_i(\mathbf{v}_i, \mathbf{y}_i)$ fonksiyonu ile modellenir. Bunun yanı sıra MRF modellerinde tahmin edilmek istenilen \mathbf{v}_i değişkenlerinin de belirli bir komşuluk içinde birbirleriyle ilintili olduğu varsayılır ve bu ilinti uyumluluk fonksiyonu $\Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j)$ ile modellenir. ij indisi komşu pikselleri göstermektedir.

Bunların ışığında \mathbf{v}_i ve \mathbf{y}_i için ortak olasılık yoğunluk fonksiyonu Z normalizasyon terimi olmak üzere eşitlik 3.5 deki şekilde yazılabilir [15].

$$P(\mathbf{v}, \mathbf{y}) = \frac{1}{Z} \prod_{(ij)} \Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j) \prod_{(i)} \Phi_i(\mathbf{v}_i, \mathbf{y}_i) \quad (3.5)$$

Bölüm 2.2’de bahsedildiği gibi enerji tabanlı yöntemler ile işlem yapıldığı için eşitlik 3.5’de belirtilen marjinal olasılık fonksiyonunun eşitlik 2.13’deki şekilde Gibbs dağılımına uyduğu varsayılır. Bu durumda çarpım halinde bulunan terimlerin her biri toplam fonksiyonlarına dönüşür. MRF modelinde $\Psi_{ij}(\mathbf{v}_i, \mathbf{v}_j)$ olarak gösterilen uyumluluk fonksiyonu SIFT akış yöntemine küçük yer değiştirme terimi ve düzgünleştirme terimi, $\Phi_i(\mathbf{v}_i, \mathbf{y}_i)$ olarak gösterilen fonksiyon ise veri terimi olarak aktarılmıştır. Böylelikle eşitlik 3.3’de görülen maliyet fonksiyonu elde edilir. Tez çalışmasında eşitlik 3.3’de belirtilen enerji fonksiyonunu enküçükleme için Bölüm 2.2.2’de açıklanan inanç yayılım algoritması kullanılmıştır. Bahsedilene ek olarak bu yöntemde Şekil 3.3’de görülen iki katmanlı inanç yayılımı gerçekleştirilmektedir [5, 28]. BP gerçeğe iki katmanın birisi akış vektörünün \mathbf{x} doğrultusundaki bileşenini diğeri ise \mathbf{y} doğrultusundaki bileşenini iteratif olarak hesaplamak için kullanılmaktadır. Bölüm 3.2 de bahsedildiği gibi düzgünleştirme teriminin ayrıştırılmış olarak kullanılmasının nedeni iki katmanlı inanç yayılım algoritmasının kullanılmasını kolaylaştırmaktır.



Şekil 3.3: İki katmanlı inanç yayılım algoritması.

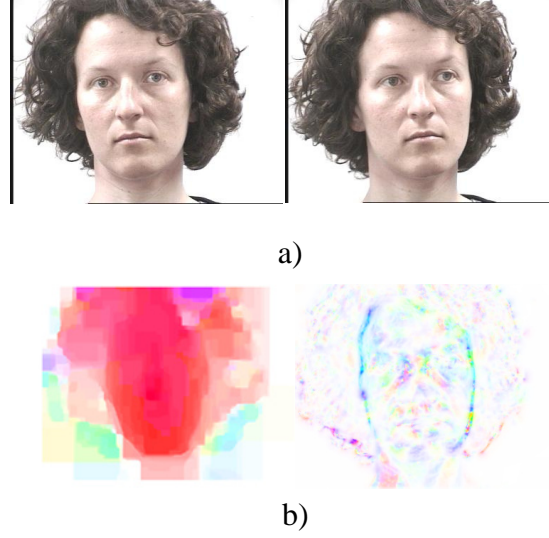
SIFT akış algoritmasında herhangi bir piksel bir diğer piksel ile eşleştirilebilir. Bu nedenle olası tüm durumların analizi işlemsel karmaşıklığı çok arttırır. Karmaşıklığı

azaltmak amacıyla kabadan-ayrıntılıya (coarse-to-fine) işleme teknikleri kullanılır [5]. Amaçlanan kaba işleme adımında akış vektörlerini kabaca tahmin etmek ve her seviyede daha da iyileştirerek ayrıntılı akış vektörlerine ulaşmaktır. Gerçeklenen SIFT akış yöntemi 4 işleme seviyesi kullanmaktadır. Konu işlemsel karmaşıklık açısından ele alınırsa, video çerçevesinin $k \times k$ piksellik olduğunu varsayarsak kabadan-ayrıntılıya işleme gerçekleştirildiğinde işlemsel karmaşıklık $O(k^2 \log k)$ düzeyinde iken tek seviyeli yöntemde bu karmaşıklık $O(k^4)$ olarak hesaplanır [5]. Bu oldukça büyük bir kazançtır, örneğin 384×288 'lik bir video çerçevesi çifti için bir seviyeli SIFT akış hesabı yaklaşık 40 saniye sürerken 4 seviye ile bu sürenin 10 saniyeye düştüğü görülmüştür.

Yapılan çalışmalarda bu yapıyı kullanmanın hızı arttırmanın yanı sıra enerji enküçüklemeye tek seviyeli yöntemle göre daha doğru sonuçlar verdiği gözlemlenmiştir.

3.4 SIFT Akış Algoritması Parametreleri Ve Etkileri

Bu bölümde Bölüm 3.1, 3.2 ve 3.3 te tanımlanan SIFT akış algoritmasında kullanılan parametreler ve bu parametlerin akış kestirim başarımına etkileri irdelenmektedir. SIFT akış kestirim performansı irdelenmeden önce kullanılan olasılıksal modellemenin gradyan tabanlı yöntemlere üstünlüğünü görebilmek açısından Şekil 3.4'te karşılaştırmalı sonuçlar verilmektedir. Olasılıksal bir yöntem olan SIFT akış yönteminin gradyan tabanlı bir yöntem olan Horn and Schunk Optik akış [13] algoritmasından daha iyi sonuç verdiği görsel olarak Şekil 3.4'de görülebilmektedir. Şekil 3.4.a'da kullanılan video çerçeveleri, Şekil 3.4.b'de elde edilen optik akış vektörlerinin renk kodlu gösterimleri verilmektedir. SIFT akış için Çizelge 3.1'de listelenen parametreler kullanılmıştır. Horn ve Schunk algoritmasında ise α parametresi 2, iterasyon sayısı 60 olarak alınmıştır. SIFT akış algoritmasının akış vektörlerini doğrulukla bulmasına karşın Horn ve Schunk sonuçlarında bölütsel bütünlük sağlanamadığı görülmektedir.

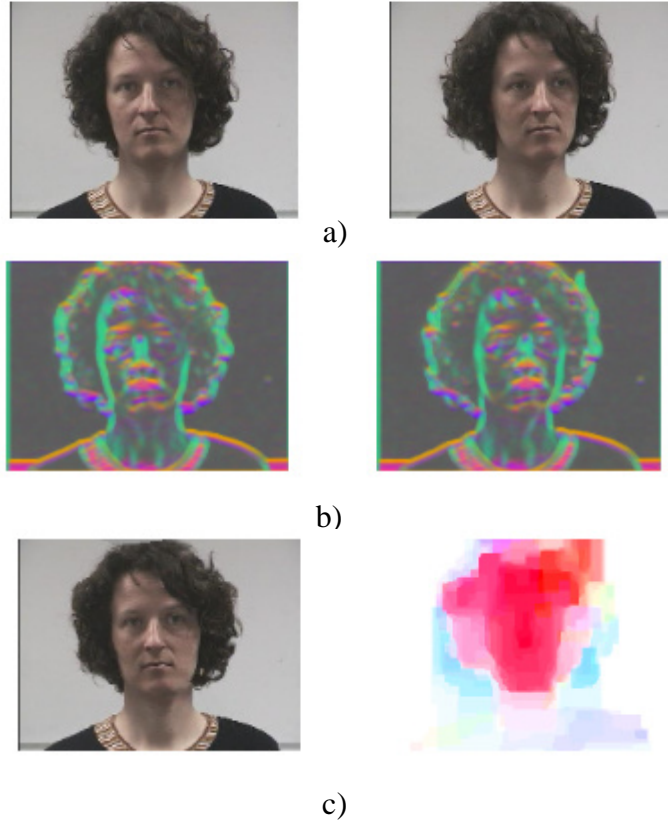


Şekil 3.4: a)Akış vektörleri hesaplanan video çerçeveleri b)SIFT akış yöntemi ile elde edilen akış vektörlerinin renk kodlu gösterimi (solda). Optik akış yöntemi ile elde edilen akış vektörlerinin renk kodlu gösterimi (sağda).

SIFT akış kestirim algoritmasının parametreleri ve kullanılan değerler Çizelge 3.1 de listelenmektedir [5]. Testlerde etkisi gözlemlenmek istenen parametre değiştirilirken diğer tüm parametreler sabit tutulmuştur. Çizelge 3.1’de gösterilen ve testlerde de kullanılan parametrelerle yapılan test sonuçları Şekil 3.5’de gösterilmiştir.

Çizelge 3.1:SIFT akış algoritmasının parametreleri.

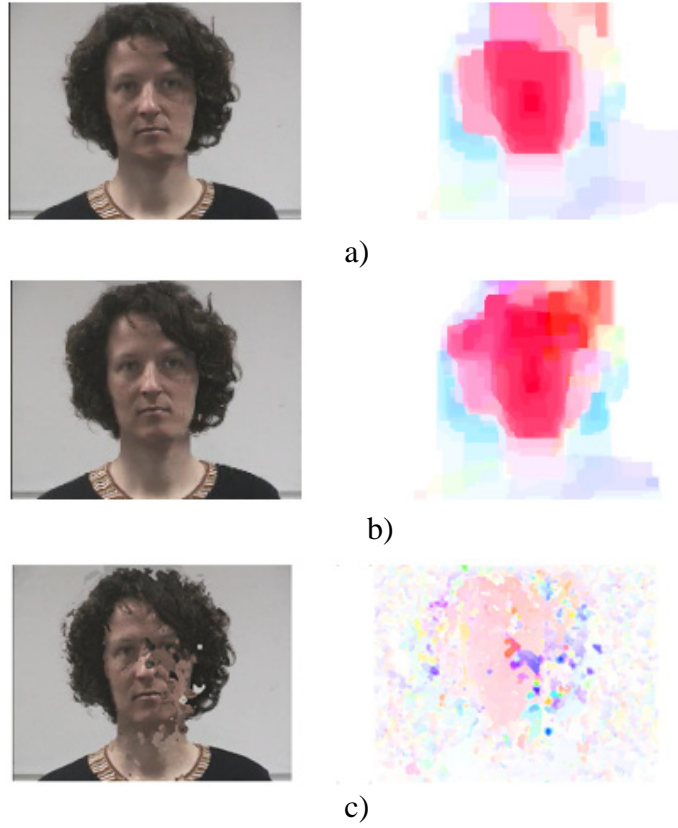
Parametre Açıklaması	Parametre Simgesi	Değeri
Düzgünlük teriminin katsayısı	α	2
Küçük yer değiştirme teriminin katsayısı	η	0.005
Düzgünlük terimi için eşik değeri	d	40
Seviye sayısı	nlevels	4
En üst seviyedeki döngü sayısı	TopIteration	100
Genel döngü sayısı	nIteration	60
SIFT imgesi elde edilirken kullanılan pencere	patchsize	8
Örnekleme hassasiyeti	gridspacing	1



Şekil 3.5: Varsayılan parametre değerleri ile elde edilen SIFT Akış kestirim sonuçları. a) 0 derecelik (solda) ve 15 derecelik (sağda) video çerçeveleri. b) Karşı düşen SIFT imgeler. c) Akış kestirimi sonucunda yeniden oluşturulan 15 derecelik video çerçevesi (solda) ve kestirilen SIFT akış vektörlerinin renk kodlu gösterimi (sağda).

α , Düzgünlük terimi yumuşatma katsayısının etkisi :

Eşitlik 3.3 de görülebilen bu katsayı maliyet fonksiyonunda düzgünlük teriminin ağırlığını kontrol eder. Farklı α değerleri ile yapılan testler için renk kodlu SIFT akış vektörleri ve yeniden oluşturulan imge şekil 3.6' da gözlemlenebilmektedir.

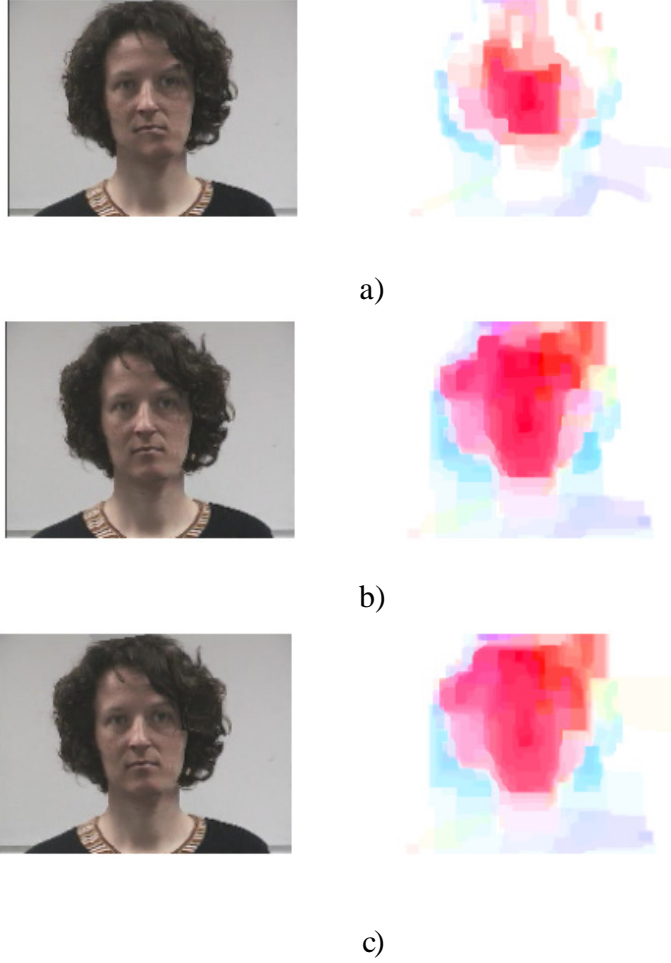


Şekil 3.6: α parametresinin etkisi a) α : 4 için b) α : 2 için (testlerde kullanılan değer) c) α : 0 için, terim yok sayılmıştır.

$\alpha = 4$ durumunda SIFT akış vektör renk diyagramlarını karşılaştırdığımızda saç bölgesinde doğru yoğunlukta vektörlerin bulunamadığı gözlemlenmiştir. $\alpha = 0$ durumunda ise ilgili terimin etkisi tamamen ortadan kaldırılmıştır. Renk diyagramından görüldüğü gibi yoğunluk tamamen ortadan kalkmış belirli aralıklarda farklı yoğunlukta vektörlerle düzgün olmayan bir geri çatılmış video çerçevesi elde edilmiştir. Yapılan testlerdebu parametre için en uygun değer 2 (şekil 3.5.b) olarak belirlenmiştir.

η , Küçük yer değiştirme terimi yumuşatma katsayısının etkisi:

Eşitlik 3.3 de görülebilen bu katsayı maliyet fonksiyonunda küçük yer değiştirme teriminin bu fonksiyonundaki ağırlığını kontrol eder. Farklı η değerleri ile yapılan testler için SIFT akış vektörü renk diyagramları ve yeniden oluşturulan imge şekil 3.7'de gözlemlenebilmektedir.

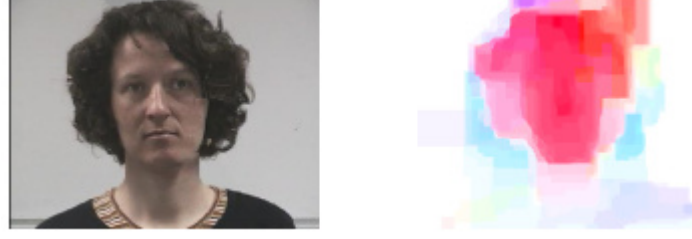


Şekil 3.7: η parametresinin etkisi a) η : 0.15 için b) η : 0.005 için (testlerde kullanılan değer c) η : 0 için , terim yok sayılmıştır.

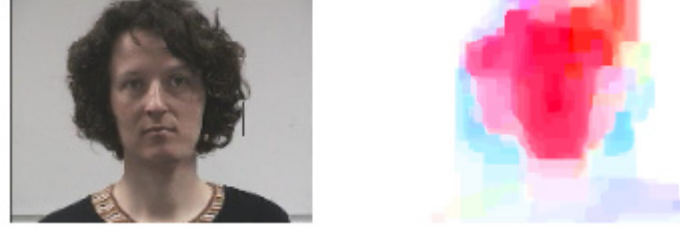
$\eta = 0.15$ durumunda SIFT akış vektör renk diyagramlarını karşılaştırdığımızda saç ve alın bölgesinde kestirilen akış vektörlerinin doğru bulunamadığı ve bulunanların yüz bölgesinden farklı özellikte vektörler olduğu gözlemlenmiştir. $\eta = 0$ durumunda ise ilgili terimin etkisi tamamen ortadan kaldırılmıştır. Renk diyagramından görüldüğü gibi yoğunluk şekil 3.7.b'ye göre azalmış ve yanak bölgesinde düzgün olmayan bir geri çatılmış video çerçevesi elde edilmiştir. Testlerde bu parametre için en uygun değer 0.005 olarak belirlenmiştir.

nLevels, Seviye sayısının etkisi:

Bu parametre Bölüm 3 de bahsedilen kabadan-ayrıntılıya piramit yapısını oluşturan seviye sayısıdır. Farklı nLevels değerleri ile yapılan testler için SIFT akış vektörü renk diyagramları ve geri çatılan video çerçevesi şekil 3.8'de gözlemlenebilmektedir.



a) nLevels: 6 için



b) nLevels:4 için (testlerde kullanılan değer)



c) nLevels: 1 için

Şekil 3.8: nLevels parametresinin etkisi.

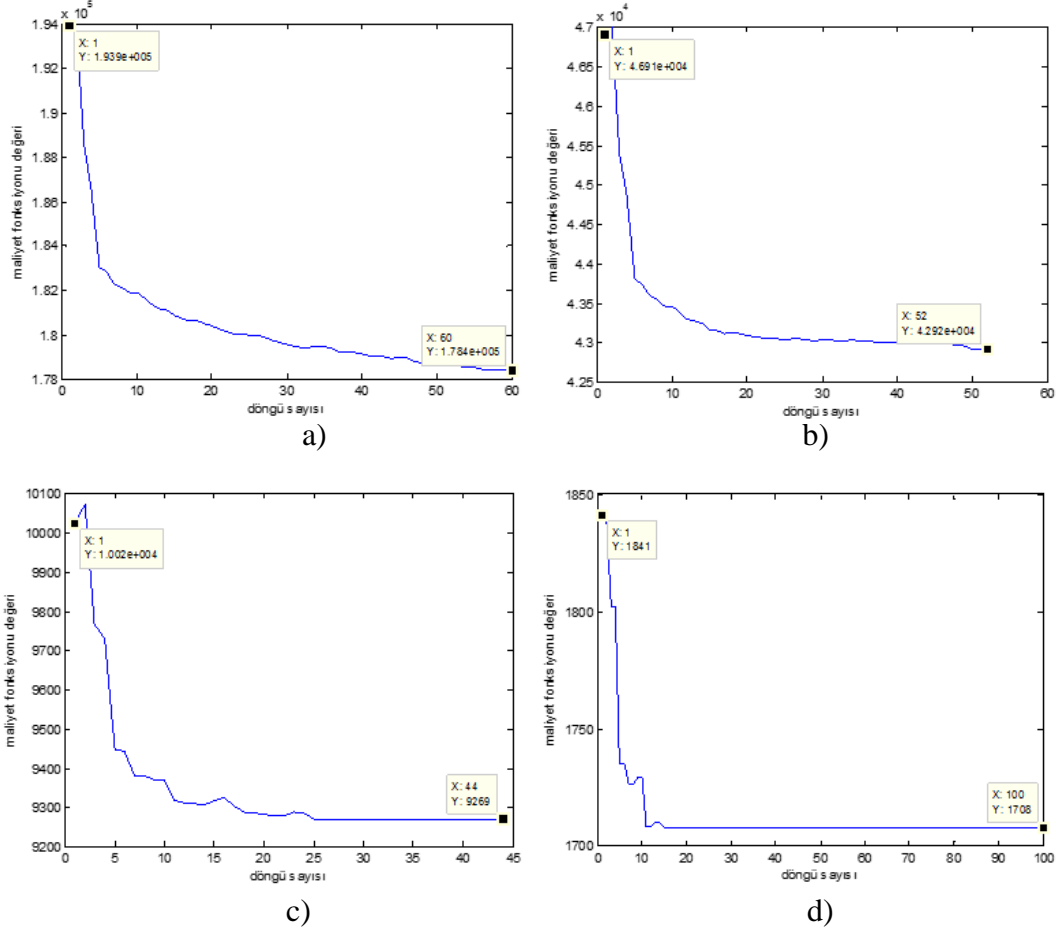
nLevels = 1 durumunda baş bölgesinde renk yoğunluğunun yüz bölgesine göre farklılık göstererek yanlış hesaplandığı gözlemlenmektedir. Bölüm 3’de bahsedildiği gibi kabadan-ayrıntılıya işlemenin temelinde her seviyede farklı çözünürlüklerde vektör hesaplaması yaparak en doğru akışı kestirebilmektir. Tek bir seviyeyle işlem yapıldığında ise kestirimtek döngü içerisinde yapılmaya çalışılır. Bu da gözlemlendiği gibi hareket vektörlerinin doğru olmayan sonuçlara yakınsamasına neden olur. nLevels = 6 için elde edilen geri çatılan video çerçevesi ve SIFT akış vektörü renk diyagramında nLevels=4 durumuna göre belirgin bir bozulma gözlemlenmemiştir. Ancak seviye sayısını arttırmak işlem yükünü arttıracığından nLevels sayısı için en uygun değer 4 olarak belirlenmiştir.

TopIteration,nIteration, Döngü sayısı etkisi:

Döngü sayısı algoritmada kullanılan nIterations ve nLevels değişkenlerine göre eşitlik 3.6 kullanılarak hesaplanır [5]. Bu denklemde d_n n. seviyedeki döngü sayısını göstermektedir. En üst seviye döngü sayısı (d_1) formüle göre belirlenmez.

$$d_n = nIterations - \frac{(nIterations - nIterations * 0.6)}{nLevels} * (nLevel - n) \quad (3.6)$$

Algoritmada bu değer özel olarak girilir ve tez kapsamında bu sayı 100 olarak belirlenmiştir. Farklı seviyeler için maliyet fonksiyonunun döngü sayısına bağlı değişimi Şekil 3.9'da gösterilmiştir.



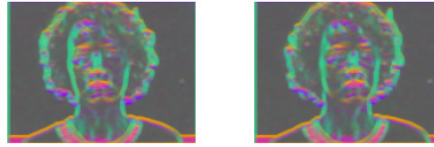
Şekil 3.9: Seviyelerde döngü sayısına göre maliyet fonksiyonu değişim grafikleri
a) 4.seviye 60 döngü için maliyet değişimi b) 3. Seviyede 52 döngü için maliyet c) 2.seviye 44 döngü için maliyet değişimi d) 1.seviye 100 döngü için maliyet.

Görüldüğü üzere her bir üst seviyede döngü bir öncekinden daha düşük maliyet değerinde başlamış ve en üst seviyede en düşük başlangıç maliyeti değerine ulaşılmıştır. Bir başka deyişle kullanılan hiyerarşik yapı tek bir seviye kullanılarak elde edilebilecekten daha düşük bir maliyet ile eniyileme yapılabilmesine olanak tanımaktadır. Çünkü sistem döngüye başlarken her vektör için bir başlangıç değeri atar. Tek seviyede işlem gerçekleşirse bu atanan başlangıç değerlerinin çok uygun olmamasından dolayı en düşük maliyetli sonuca yakınsanamayabilir. Ancak her bir

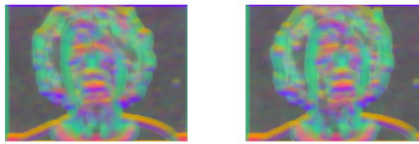
seviyede kestirilen akış vektörleri üst seviyenin başlangıç değerini oluşturursa kestirim daha doğru yapılabilir.

patchSize parametresinin etkisi:

Bu parametre Bölüm 3.1’de bahsedilen SIFT dönüşümü yapılırken kullanılan pencere boyutunu belirlemeye yarayan parametredir. Bu yüzden asıl etkisini SIFT imgelerin oluşturulduğu algortmada gösterir. Şekil 3.10’da farklı patchSize da elde edilen SIFT imgeleri ve kestirilen akış vektörlerinin renk kodlu gösterimleri gözlenmektedir. Görüldüğü gibi patchSize parametresi ile SIFT görüntüsünün keskinliğini kontrol edebiliriz. Şekil 3.10’dan anlaşılacağı üzere patchSize arttırıldığında SIFT görüntüsünde kenar ve köşelerin keskinliği ortadan kalkarak daha yumuşak geçişlere sahip görüntüler oluşmuştur. Bu görüntüler kullanılarak bulunan vektörlerde ise kenarlardaki yoğunlukları azalmış geçişler belirginsizleşmiştir.

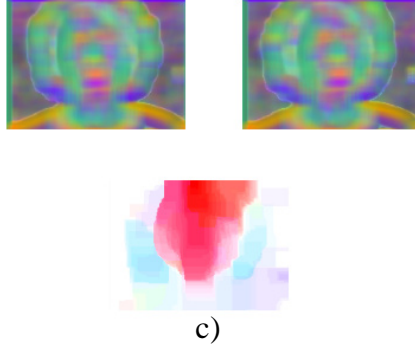


a)



b)

Şekil 3.10: patchSize parametresinin etkisi a) patchSize: 8 için b) patchSize: 16 için c) patchSize: 32 için.



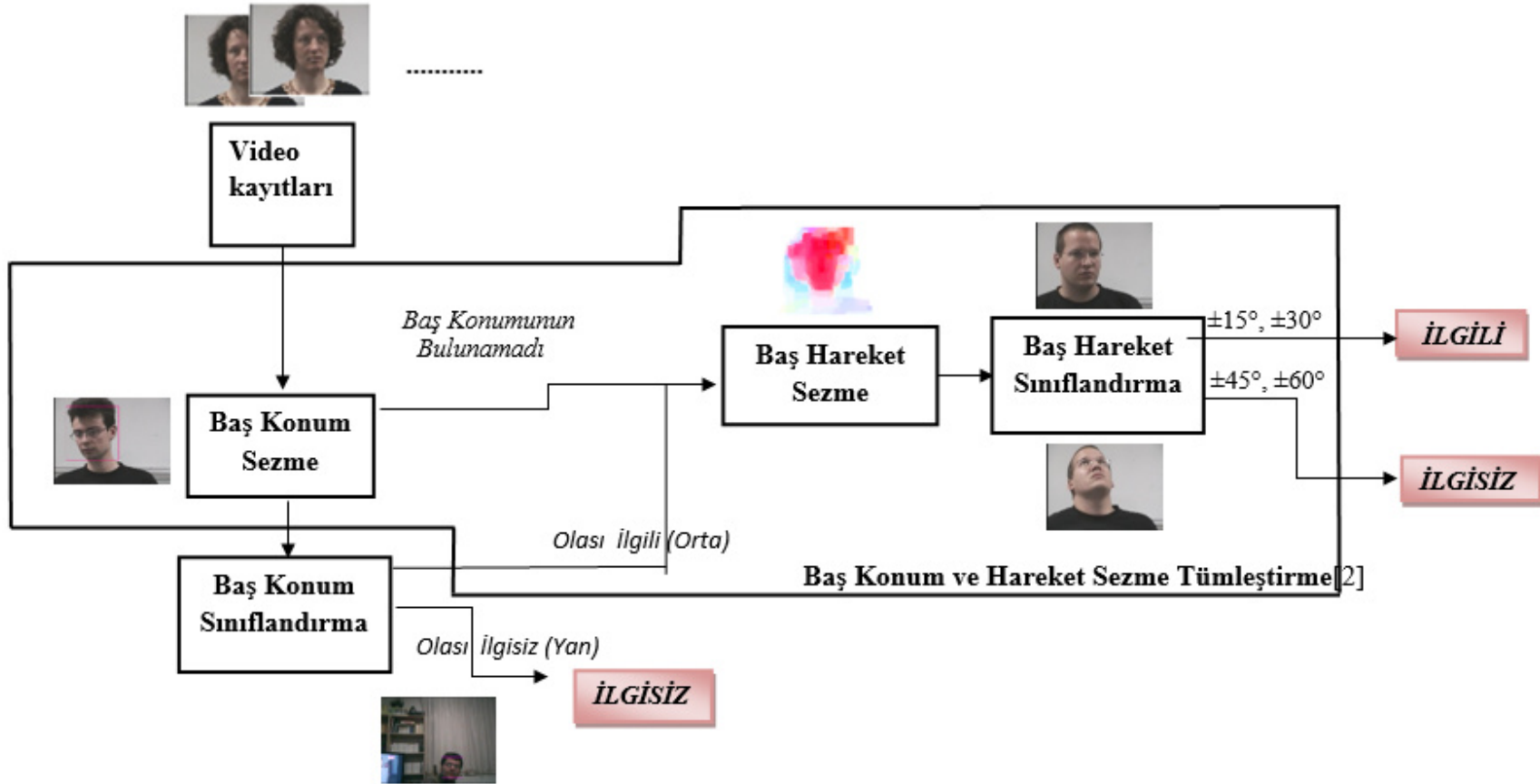
Şekil 3.10 (devam): patchSize parametresinin etkisi a) patchSize: 8 için
b) patchSize: 16 için c) patchSize: 32 için.

Gridspacing parametresinin etkisi:

Bu parametre Bölüm 3.1 bahsedilen SIFT öznitelik çıkarımı yapılırken kullanılan örnekleme hassasiyetini belirleyen parametredir. Bu yüzden asıl etkisini SIFT imgelerin oluşturulduğu algorithmada gösterir. Örnekleme değıştiđi için SIFT imgelerinin boyutları (çözünürlükleri değışir). Gridspacing parametresini 2 katına çıkarmak oluşan akış vektörü sayısını yarı yarıya azaltır. 4 katına çıkarmak ise vektör sayısını 4 te 1 ine düşürmektedir.

4. BAŞ HAREKET VE KONUM BİLGİSİ TÜMLEŞTİRİLEREK İLGİLİLİK SINIFLANDIRMA

Tez kapsamında, baş hareket kestirim öznitelikleri ile gerçekleştirilen ekrana ilgililik, ilgisizlik sınıflandırma başarımının baş konum sezme ve baş hareket kestirimi algoritmalarını tümleştirerek artırılması amaçlanmıştır. Bu amaçla video çerçevelerinden öncelikle Viola Jones yüz sezme algoritması kullanılarak baş konum bilgisi elde edilir. Baş hareketinin örtüşme yarattığı durumlarda veya baş konumunun video çerçevesinin dışına çıktığı durumlarda baş konum sezme algoritması sonuç bulamamaktadır. Bu durumda doğrudan SIFT akış algoritması ile hareket kestirimi yapılmaktadır. Baş konum sezme algoritması ile konumun sezilebildiği durumlarda ise elde edilen öznitelikler ile sınıflandırılan ve baş konumu “yanda” olarak belirlenen video çerçeveleri, erkana ilgisiz olduğu düşünülerek “ilgisiz” olarak etiketlenir. Baş konum belirleme öznitelikleri ile sınıflandırılmada “ortada” yani “olası ilgili” olarak etiketlenmiş video çerçevelerinde karar tümleştirme gerçekleşir. Bu video çerçeveleri ve baş konum sezici ile konum sezilemeyen video çerçeveleri için SIFT akış yöntemi ile hareket öznitelikleri bulunur ve ikili sınıflandırma yapılır. Sınıflandırma sonucu ± 30 derece ve altı baş dönme hareketi saptanan video çerçeveleri “ilgili”, ± 45 derece ve üstü baş dönüş hareketi saptanan video çerçeveleri “ilgisiz” sınıfına atanır. Sınıflandırma ve karar tümleştirme aşağıdaki bölümlerde ayrıntılı anlatılmaktadır, Şekil 4.1’de gerçekleştirilen yöntemin blok diyagramı görülmektedir.



Şekil 4.1: Kullanılan yöntemin blok diyagramı .

4.1 Baş Dönüş Hareketi Sınıflandırma

Bu çalışmada baş dönüş hareketi bilgisi ile kişilerin ekranda gösterilen içeriğe ilgili veya ilgisiz olma durumlarının otomatik sınıflandırılması amaçlanmıştır. Bu kapsamda tez çalışmasında baş hareketlerinin sınıflandırılmasında Bölüm 2.2' de tanımlanan yeniden ağırlıklandırılmış ağaç kullanan inanç yayılım algoritması (TRW BP) ile kestirilen SIFT akış vektörlerinin kullanımı önerilmektedir. SIFT akış özniteliklerinin inanç yayılımı ile hesaplanması MATLAB ortamına gömülmüş C++ yazılımı ile gerçekleştirilmiştir. Elde edilen akış vektörlerinden eşitlik 4.1 de görüldüğü şekilde i . piksel için kestirilen $v(x(i),y(i))$ akış vektörlerinden elde edilen açı ve genlik değerleri 2 boyutlu öznitelik vektörlerini oluşturmaktadır.

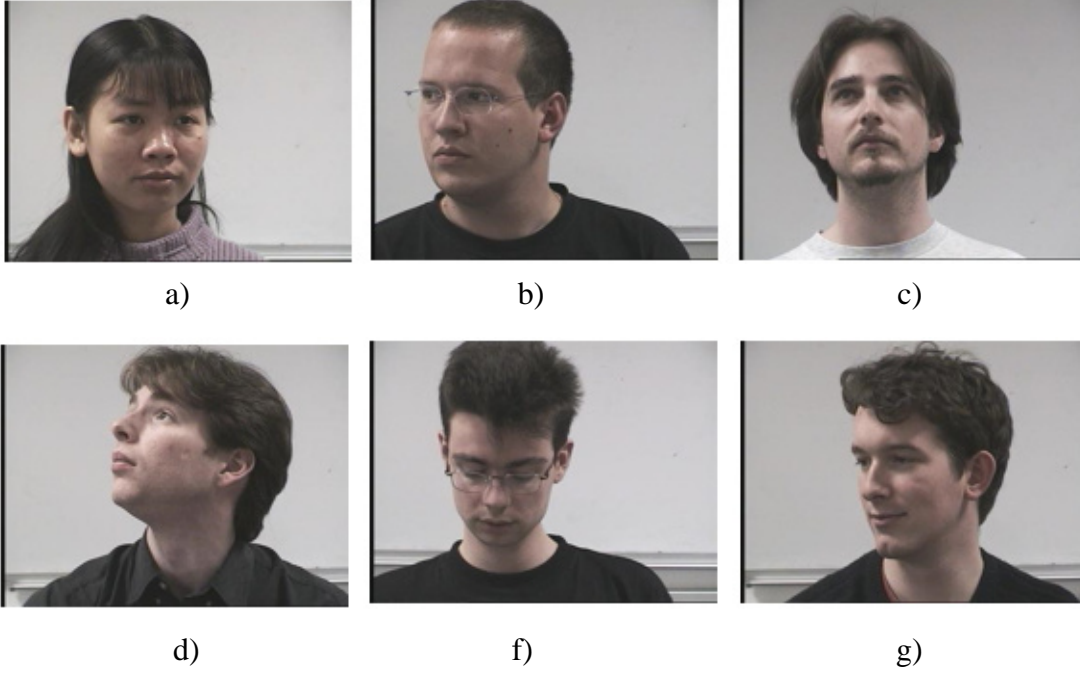
$$genlik(i) = \sqrt{x(i)^2 + y(i)^2}$$

$$açı(i) = \tan^{-1} \frac{y(i)}{x(i)} \quad (4.1)$$

Sınıflandırma yapılırken eğitim seti olarak Pointing04 veri tabanından yararlanılmıştır [11]. Bu veri tabanında şekil 4.2 de görülen farklı baş duruşuna sahip 15 kişiye ait kayıt bulunmaktadır. Eğitim setini oluşturan özniteliklerin çıkarılmasında, Çizelge 4.1'de listelenen, x doğrultusunda (pan) 9 açı, y doğrultusunda (tilt) 7 açı ve bu açılardan oluşan çapraz açılar için yapılmış kayıtlar kullanılmıştır.

Çizelge 4.1: Eğitim setinde yer alan açılar.

X doğrultusu-Pan	Y doğrultusu-Tilt
-60°	-60°
-45°	-30°
-30°	-15°
-15°	0°
0°	15°
15°	30°
30°	60°
45°	
60°	



Şekil 4.2: Pointing04 veri tabanı örnekleri a) $x=-45$ derece, $y=0$ derece , b) $x=15$ derece , c) $y=30$ derece, d) $y=-60$ derece, e) $x=-60$ derece, $y=30$ derece, f) $x=-45$ derece, $y=-15$ derece dönmüş baş hareketi.

x doğrultusunda hareket için eğitim seti, izleyicinin ekrana 0° açı ile baktığı video çerçevesi ile diğer 8 farklı açıda bakışın kaydedildiği video çerçeveleri arasında hesaplanan akış vektörlerinin açı ve genlikleri kullanılarak oluşturulmuştur. y doğrultusunda hareket için eğitim seti, izleyicinin ekrana 0° açı ile baktığı video çerçevesi ile diğer 6 farklı açıda bakışın kaydedildiği video çerçeveleri arasında hesaplanan akış vektörlerinin açı ve genlikleri kullanılarak oluşturulmuştur. Bir çerçeve çifti için yaklaşık olarak 33000 adet öznitelik vektörü yer almaktadır.

Çapraz hareket durumları için x ve y doğrultusundaki her bir açı sabit tutularak , kalan açılar ekrana 0° açılı bakış kayıtları referans alınarak taranmıştır. Örneğin x doğrultusunda baş hareketinin ekrana 15° lik açı ile baktığı kayıt sabit tutularak bütün y doğrultusundaki farklı bakış açılarına göre akış vektörlerinin açı ve genlikleri hesaplanmıştır.

Kullanılan 8 farklı baş hareketi sınıf etiketleri Çizelge 4.2 de listelenmektedir.

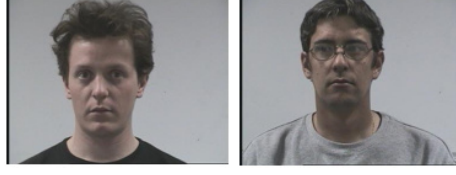
Çizelge 4.2: Eğitim ve test seti sınıf etiketleri.

	Pan	Tilt	Pan açıları sabit tutularak Tilt								Tilt açıları sabit tutularak Pan					
			-60	-45	-30	-15	15	30	-45	60	-60	-30	-15	15	45	60
-60°	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-45°	2										2	2	2	2	2	2
-30°	3	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
-15°	4	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4
15°	5	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5
30°	6	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6
45°	7										7	7	7	7	7	7
60°	8	6	6	6	6	6	6	6	6	6	8	8	8	8	8	8

Sınıfları 0° lik video çerçevesi referans alınarak oluşturmak yerine 15° aralıklı, yani ardışık olarak 0°-15° , 15°-30° vb dönüşler şeklinde etiketlemek daha doğru görünebilir. Ancak bir sonraki bölümde x yönü için bu hesaplamalar yapılmış ve ayırdediciliğin 0° referanslı sonuçlara göre oldukça düşük olduğu görülmüştür. Öte yandan ekrandaki içerik ile ilgililik ölçümünde işlemsel karmaşıklığı da azaltacak şekilde problem ekran ile ilgili ya da ilgisiz şeklinde ikili sınıflandırma problemi olarak ele alınmıştır. Bu amaçla hareket sınıflandırma Çizelge 4.2’de belirlenen sınıflar üzerinden yapıldıktan sonra, 30° ve altı baş dönüş hareketi içeren video çerçeveleri ilgili diğerleri ilgisiz olarak sınıflandırılmıştır.

Test seti olarak Şekil 4.3’de görülen eğitim setiyle aynı kişilerin farklı kıyafet ya da aksesuarlara sahip kayıtlarının olduğu test video çerçeveleri kullanılmıştır. Test setinde de sınıflar ve veri setleri eğitim setiyle aynı şekilde hazırlanmıştır.

Yukarıda bahsedilen SIFT Akış algoritması ile kestirilen genlik ve açı özniteliklerinden elde edilen iki boyutlu vektörler ile oluşturulan eğitim ve test verilerinin sınıflandırmasında, Rastgele Orman (Random Forest) algoritması kullanılmıştır. Bu algoritma karar ağacı (decision tree) olarak bilinen yapının birden fazla kere kullanılması ile oluşur [29] . Bu sınıflandırıcı için WEKA paket programında yer alan RandomForest algoritması kullanılmıştır.



a)



b)

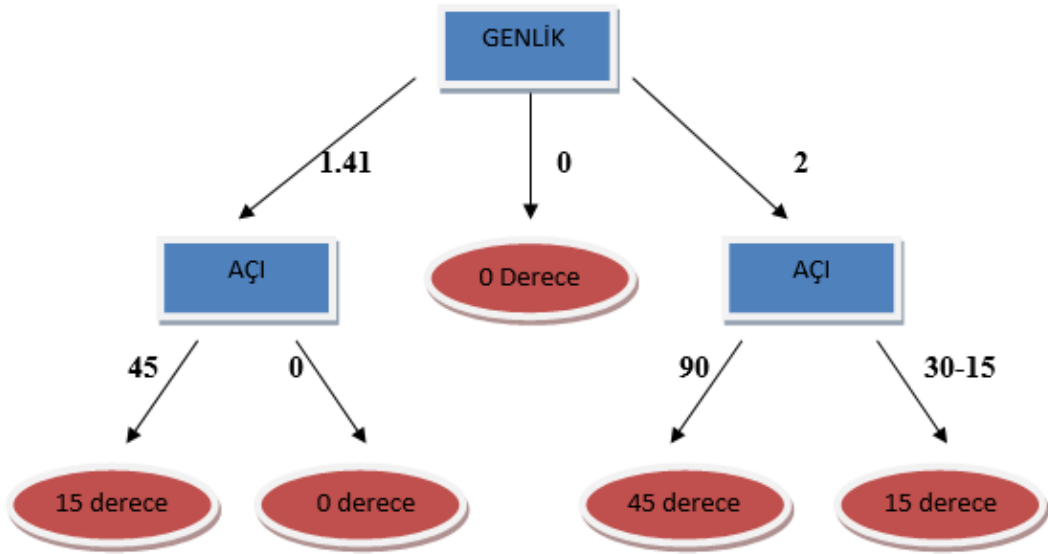
Şekil 4.3: a) Eğitim setine ait video çerçeveleri. b) Test setine ait video çerçeveleri.

4.1.1 Karar ağacı

Karar ağaçları sınıfları bilinen eğitim seti verisinden oluşturulur. Bu ağaç yapısında büyük veri grupları küçük alt veri gruplarına bölünür. Her bölme adımından sonra elde edilen alt grupların üyeleri bir diğeriyle çok daha benzer hale gelmektedir. Bölünen her grubun tüm elemanları aynı sınıf etiketine sahip olana kadar kümeleme işlemine devam edilir [29]. Tez kapsamında ilgilenilen baş dönüş açısı sınıflandırma probleminde olduğu gibi yüksek sayıda verilerin kullanıldığı pek çok sınıflama probleminde karar ağaçları karmaşıklık azaltma ve hızlandırma açısından yararlı bir çözüm olmaktadır. Bu yöntemde oluşturulan ağaç yapıları düğüm noktaları, dallanmalar ve yapraklardan oluşmaktadır. Düğüm noktaları seçilen öznitelikler, dallanmalar ilgili özniteliğin ait olduğu sınıflar ve yapraklar ise sınıflandırma sonucunu temsil eder. Örnek vermek için Çizelge 4.3'te listelenen öznitelikleri ve sınıf etiketlerini göz önüne alalım. Bu eğitim setinden 'genlik' özniteliği başlangıç düğümü olarak seçilirse oluşturulan karar ağacı Şekil 4.4'te görüldüğü gibi elde edilecektir. Karar ağacı kullanmakta amaçlanan benzer veya aynı sınıf değerine sahip olma eğiliminde alt grupların elde edildiği görülmektedir.

Çizelge 4.3: Karar ağacı sınıflandırma eğitim veri seti.

Genlik	Açı	Sınıf
0	0	0 derece
1.41	45	15 derece
2	90	45 derece
1.41	0	0 derece
2	30	15 derece
0	0	0 derece
2	15	15 derece



Şekil 4.4: Eğitim setine uygun olarak oluşturulmuş karar ağacı.

4.1.2 Rastgele orman algoritması

Leo Breiman tarafından 2001 yılında önerilen Rasgele Orman (RF) algoritmasında amaç tek bir karar ağacı yerine ana eğitim kümesinden seçilen farklı alt kümeler ile eğitilmiş çok sayıda ağaç oluşturmaktır [10]. Test aşamasında ise karar verme, oluşturulan ağaçların çoğunluk oylamalı seçimine göre yapılır.

Breiman'in bu algoritmada getirdiği en büyük yenilik rastgelelik özelliğidir. Bu özelliğe göre her bir düğüm, rastgele olarak seçilen özniteliklere göre dallara ayrılır [10].

x, y, z öznitelikler, c atama yapılan sınıf ve n gözlenen öznitelik vektörü sayısı olmak üzere eğitim seti $D = \{(x_1, y_1, z_1, c_1); (x_2, y_2, z_2, c_2); \dots; (x_n, y_n, z_n, c_n)\}$ şeklinde gösterilsin. RT algoritmasında izlenmesi gereken adımlar aşağıdaki gibi özetlenebilir [10].

- 1) Oluşturulacak T_i ağacı için D eğitim veri setinden D_i alt kümesi seçilir.
- 2) Oluşturulacak ağaçta her bir düğüm noktası için kullanılacak m adet öznitelik vektörü rastgele olarak seçilir, dallanmalar yapılarak alt ağaçlar oluşturulur.
- 3) Oluşturulmak istenilen N adet ağaç 1 ve 2. aşamalar tekrar edilerek elde edilir.

Bu algoritmada kullanıcı tarafından belirlenen değişkenler oluşturulacak ağaç sayısı N ve her bir düğümde kullanılacak öznitelik sayısı m 'dir.

2. adımda bahsedilen her bir düğüm için oluşturulan dallar CART (Classification and Regression Tree) algoritmasının kriterine göre elde edilir [10].

CART algoritmasına göre veriler her bir düğümde ikili alt kümelere ayrılır ve her bir alt küme bir önceki alt kümeden daha homojen bir yapıdadır. Bu ayırma işlemi saf homojenlik sağlanana kadar ya da belirlenen durma kriterleri sağlanana kadar devam eder [29].

Burada dikkat edilmesi gereken durum, düğümlerdeki alt kümeler oluşturulurken, her bir alt kümedeki sınıf etiketlerinin mümkün olduğunca homojen olmasını sağlamaktır.

CART modelinde sınıf homojenliğini hesaplamak için GINI ölçütü kullanılır [29,34]

GINI ölçütü 4.2'de verilmiştir.

$$gini = 1 - \sum_{j=1}^n p_j^2 \quad (4.2)$$

Eşitlik 4.2'de n sınıf sayısı, j sınıf numarası ve p_j ise j sınıfına ait gözlem sayısının toplam gözlem sayısı T 'ye oranını belirtmektedir. Burada bir düğümdeki GINI değeri hesaplanırken T gözlem sayısı o düğümdeki gözlem sayısı olarak alınır. Yani

T değeri eğitim setinin tümündeki gözlem sayısı değildir. Çünkü alt düğümlere inildikçe ilgili düğümü ilgilendiren gözlem sayısı azalır.

Bir düğümü alt gruba böldüğümüzde oluşan iki alt grup için toplam GINI değeri eşitlik 4.3’de belirtilmiştir.

$$gini(toplam) = \frac{N_1}{N} x gini(düğüm1) + \frac{N_2}{N} x gini(düğüm2) \quad (4.3)$$

Bu eşitlikte N_1 ve N_2 her bir alt gruptaki gözlem sayısı, N ise bir üst düğümden gelen toplam gözlem sayısı N_1+N_2 ‘yi ifade etmektedir.

Toplam GINI indeksi azaldıkça sınıf homojenliği artar. Bir alt düğümün GINI indeksi bir üst düğümün GINI indeksinden daha az olduğunda o dal başarılıdır. GINI indeksi sıfıra ulaşıncı o grupta tek bir sınıf kalmış olur ve dallanma işlemi sona erer [34].

4.2 Baş Konum Sınıflandırma

Bu çalışmada konum bilgisi ile kişilerin ekranda gösterilen içeriğe olası ilgili veya olası ilgisiz olma durumlarının sınıflandırılması amaçlanmıştır. Baş konum öznitelikleri, yüz konumu başlangıç pikselleri (x,y) ve baş bölgesi alanı A olmak üzere 3 boyutludur. Kullanılan sınıf etiketleri çizelge 4.4 de gösterilmiştir.

Çizelge 4.4: Violen-Jones sınıf etiketleri.

Ekranı uzaklık	Sınıf Kriteri	
	Orta (Olası İlgili)	Yan (Olası İlgilisiz)
Uzak($A < 20000$)	$70 < x < 441$ $70 < y < 281$	$x < 71$ veya $x > 440$ $y < 71$ veya $y > 280$
Yakın($A > 55000$)	$50 < x < 251$ $80 < y < 161$	$x < 51$ veya $x > 250$ $y < 81$ veya $y > 160$
Orta ($19999 < A < 55001$)	$149 < x < 351$ $50 < y < 231$	$x < 150$ veya $x > 350$ $y < 51$ veya $y > 230$

Orta olarak sınıflandırılan video çerçeveleri ilgililik problemi kapsamında “olası ilgili”, yanda olarak sınıflandırılan video çerçeveleri ise “olası ilgisiz” olarak etiketlenir.

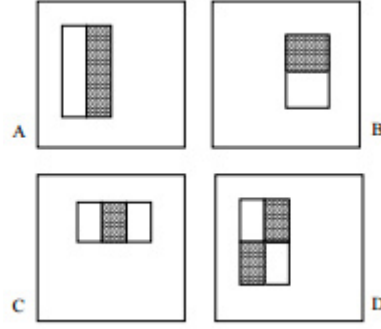
Tez kapsamında baş konum sınıflandırma algoritması olarak Viola Jones yüz sezme algoritması kullanılmıştır.

Viola Jones algoritması hızlı ve yüksek doğrulukla yüz bulma gerçekleyen ve bu nedenle yaygınlıkla kullanılan bir algoritmadır. 2001 yılında Paul Viola ve Micheal Jones tarafından önerilmiştir [4]. Bu algoritma ile SIFT akış algoritmasını uygulamadan önce yüzün yeri saptanarak, arkaplanın olabildiğince atılması ve sadece ilgilenilen kısımların sınıflandırmaya sokulması amaçlanmıştır.

Bu algoritmada Haar öznitelikleri denilen basit dörtgen şekilleri kullanır. Bu amaçla video karesi integral resim olarak adlandırılan bir gösterime çevrilir [4,30]. Bu gösterim Haar özniteliklerin farklı ölçeklerde elde edilmesinde kolaylık ve hız artışı sağlar.

Sınıflandırıcı elde edilme aşamasında ise algoritmadaki hızı sağlamak amacıyla çok sayıda olan Haar benzeri özniteliklerden önemli olan daha az sayıdaki öznitelige odaklanılır. Bu işlem için ise AdaBoost yöntemi kullanılmaktadır [4,31].

Viola Jones algoritmasında yüz sezme işlemi, doğrudan pikseller ve sahip oldukları yoğunluk değerleri ile değil belirlenen öznitelikler ile yapılır. Bahsedilen öznitelikler [32]’de de kullanılan Haar basis fonksiyonuna göre oluşturulur. Virole Jones algoritmasında kullanılan, 3 farklı Haar özniteliği vardır. İlki iki-dikdörtgen öznitelik olarak adlandırılır. Bu öznitelik iki dikdörtgensel bölgenin içinde yer alan piksel değerlerinin toplamının farkı ile ifade edilir. Bu dikdörtgenler aynı boyutta ve şekilde olmalıdır. İkincisi üç-dikdörtgen öznitelik olarak adlandırılır ve ortada yer alan bir dikdörtgensel bölgenin piksel toplamı ile bu bölgenin yanlarında yer alan iki dikdörtgensel bölgenin piksel toplamının farkını verir. Son olarak ise dört-dikdörtgen öznitelikten bahsedilebilir. Burada ise diagonallerde yer alan dikdörtgensel bölgelerin piksel farklarından sözedilir [4]. Bu dikdörtgensel bölgeler şekil 4.5 de resmedilmiştir. Bahsedilen Haar özniteliklerinin varlığı eşitlik 4.4 de gösterildiği şekilde koyu renkli piksel değerlerinin açık renkli piksel değerlerinden çıkarılmasıyla belirlenir. Fark belli bir eşik değerinin üstünyse bu özniteliğin var olduğu kabul edilir.



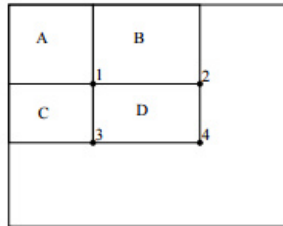
Şekil 4.5: Viola-Jones öz nitelikleri A ve B iki-dikdörtgen öz nitelik. C üç-dikdörtgen öz nitelik. D dört dikdörtgen öz nitelik.

$$\text{Dikdörtgensel Öz nitelikler} = \sum_{i \in \text{beyaz alan}} p_i - \sum_{j \in \text{gri alan}} p_j \quad (4.4)$$

Bu öz nitelikler belirli bir eşik değerinin üstünde yada altında bulunma durumuna göre yüzdür yada değildir şekilde sınıflandırılacaktır.

Öz niteliklerin hesaplanması işlemi integral resim adı verilen yapı ile büyük bir hız kazanır. (x,y) noktası için integral resim bu noktaların üst ve solunda yer alan pikseller toplamına eşittir .

Şekil 4.6 yı göz önüne alırsak, 1 noktası için integral resim A, 2 noktası için integral resim A+B , 3 noktası için integral resim A+C , 4 noktası için ise A+B+C+D olur. Haar öz nitelikleri elde etme aşamasında D bölgesinde yer alan piksel değerleri toplamı hesaplanılmak istenirse, $4 + 1 - (2 + 3)$ şeklinde yazılabilir. Görüldüğü gibi tek bir dikdörtgensel bölge 4 değer ve iki kare içeren Haar öz nitelikleri 8 değer ile hesaplanabilir [4].



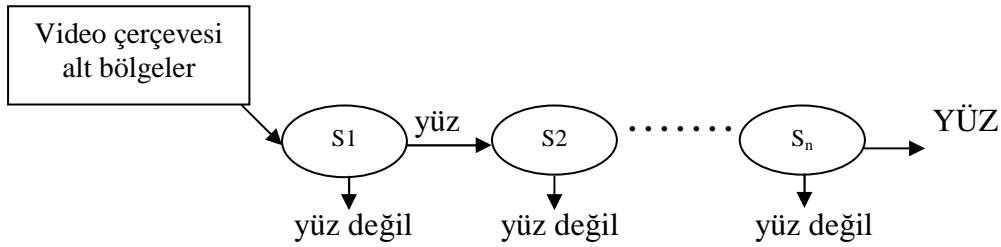
Şekil 4.6 : İntegral resmin öz nitelik hesaplamadaki kolaylığı.

İntegral resim kullanımı Haar benzeri öz niteliklerin hesaplanmasını hızlandırır da tüm video çerçevesi şekil 4.4 deki öz nitelikler ile tarandığında çok fazla öz nitelik ortaya çıkmaktadır. Bu öz niteliklerin hepsinin kullanılması yerine işe yarayacak

olanların seçilmesi hız açısından önemlidir. Bunun için ise Boosting algoritmasından yararlanır.

Boosting algoritması temel olarak uygun olan özneliklerin seçilip ağırlıklandırılarak birleştirilmesidir. Eşitlik 4.4 de hesaplanan bu öznelikle zayıf sınıflandırıcı olarak kabul edilerek, değeri belirli bir eşik değerinden büyük ise sınıflandırmanın sonucu pozitif, değilse negatif olarak belirlenir. Daha sonra hataya göre ağırlıklandırmalar yapılarak güçlü bir sınıflandırıcı elde edilir [4].

Özneliklerinin sayısını bu şekilde düşürülmesine rağmen yeterli hıza ulaşamadığı için kademeli sınıflandırıcı adı verilen yapılar devreye girmiştir. Burada boosting ile azaltılan öznelik sayısı n adet alt gruba bölünerek farklı sınıflandırıcılar oluşturulur. Bir sınıflandırıcılarda, yüz değil olarak etiketlenen resim alt bölgeleri bir sonraki sınıflandırıcılar ile test edilmez (şekil 4.7). Tüm sınıflandırıcıların yüz olarak etiketlediği kısımlar yüz olarak etiketlenir [4]. Böylece çerçevenin tamamı bütün sınıflandırıcılara sokulmamış olur. Buda hız açısından büyük bir katkı sağlar.



Şekil 4.7 Violen-Jones kademeli sınıflandırıcı.

Violen-Jones algoritmasında kullanılan integral resim, boosting, kademeli sınıflandırıcılar yapıları sayesinde yüz bulma işlemi anlık olarak gerçekleştirilebilmektedir. Tezimizde OpenCV'nin Violen-Jones kütüphanesini kullanarak video üzerinde yüz sezme gerçekleştirildi. Violen-Jones algoritması ile sezilen yüz görüntüleri şekil 4.8 da görülebilmektedir.



Şekil 4.8 Violen-Jones yüz sezme algoritması çıktı örnekleri.

5. BAŞARIM TESTLERİ

Tez kapsamında hareket sezme, konum sezme ve tümleştirme başarımlarının hesaplanması için çeşitli testler yapılmıştır. Baş hareketini açısız düzeyde göz ile saptamak zor bir problemidir. Eğitim seti olarak gerçek videolar kullanılmak istenirse etiketlenmenin doğru bir şekilde yapılması oldukça zordur. Bu nedenle tüm testlerde eğitim verisi olarak baş duruş açılarının özel olarak ölçülmüş ve etiketlenmiş olduğu Pointing'04 veri tabanı kullanılmıştır [11]. Pointing04 eğitim ve test veritabanında, 15'er farklı kişi ve her kişinininde 63 farklı baş duruşunu içeren 1890 video çerçevesi vardır. Pointing04 veritabanında, her kişi için yüz konumlarındaki değişiklik yaklaşık sıfırdır, sadece açısız değişiklikten kaynaklı bir değişiklik vardır.

Tez kapsamında yapılan testler 3 ana başlıkta toplanabilir:

1. Baş dönüş hareketi sınıflandırma başarımları testleri.
2. Baş konum sınıflandırma başarımları testleri .
3. Karar tümleştirme ile baş dönüş hareketi sınıflandırma başarımları testleri.

Aşağıdaki paragraflarda yapılan testler sırasıyla açıklanmakta ve sonuçları raporlanmaktadır.

Baş dönüş ve konum öznitelikleri kullanılarak elde edilen sınıflandırma başarımları, literatüre uygun olarak, doğru bulma ve hatalı bulma yüzdelerini birlikte yansıtan ölçütlerle raporlanmıştır. Eşitlik 5.1'de belirtilen, bir sınıfa hatasız atanan gözlem sayısının o sınıfa ait toplam gözlem sayısına oranını veren geri çağırma (recall (R)) ve bir sınıfa hatasız atanan gözlem sayısının o sınıfa atanan toplam gözlem sayısına oranını veren duyarlılık (presicion(P)) yüzdeleri raporlamada kullanılan ölçütlerdir [33]. Bunlara ek olarak eşitlik 5.2'de verilen F1 ölçütü hesaplanarak R ve P ölçütlerinin birarada değerlendirilmesi sağlanmaktadır.

$$P_i = \frac{DS_i}{TS_i} \quad i = \{1,2, \dots\}$$

$$R_i = \frac{DS_i}{GS_i} \quad i = \{1,2, \dots\} \quad (5.1)$$

$$F_i \text{ ölçütü} = 2 \times \frac{P_i \times R_i}{P_i + R_i} \quad (5.2)$$

Eşitlik 5.1 ve 5.2’de görülen DS_i , TS_i ve GS_i sırasıyla i . sınıf için doğru tahmin edilen gözlem sayısını, i . sınıf olarak tahmin edilen toplam gözlem sayısını ve gerçek sınıfı i olan toplam gözlem sayısını göstermektedir.

5.1 Baş Dönüş Hareketi Sınıflandırma Başarım Testleri

Baş dönüş hareketi sınıflandırma testleri için eğitim seti olarak Pointing’04 veri tabanından elde edilen ve ikili sınıflı olarak eğitilmiş sınıflandırıcılar kullanılmıştır. Test seti olarak ise Pointing’04 veri tabanının yanı sıra, gerçek videolardan da yararlanılmıştır.

5.1.1 Pointing’04 veri tabanında başarımlar

Yapılan ilk testte amaçlanan, ikili eğitilmiş sınıflandırıcılar yardımıyla baş dönüş hareketinin belirlenen iki sınıftan birisine ait olup olmadığının sınılanmasıdır (verification). Bu nedenle sınıflandırıcıya eğitilmiş olduğu ikili sınıftan test verileri verilmiş ve başarımlar video çerçevesi bazında verilen karar açısından değil, hareket vektörleri bazında sınıflandırma başarımları olarak raporlanmıştır. Bu test kapsamında sınıflandırıcılar veri tabanında bulunan ve Çizelge 4.2’de listelenen 8 farklı dönüş açısı ve bunların kombinasyonları için ikili eğitilmiştir. Testlerde de ikili sınıflandırma yapılmış her iki sınıfta elde edilen başarımların ortalaması ortalama F ölçütü cinsinden raporlanmıştır (Test 1). Bir diğer test grubunda eğitim ve test veri setleri birleştirilerek 10 çapraz doğrulama testleri yapılmıştır. Test 1 deki gibi sınıflandırıcılar ikili sınıflandırıcı olarak eğitilmiş, başarımlar ortalama F ölçütü cinsinden raporlanmıştır (Test 2). Bunların yanısıra, belirtilen her iki test sonucu, baş dönüş açıları ilgili-İlgisiz olarak iki sınıfta gruplanarak ortalama F ölçütü cinsinden raporlanmıştır (sırasıyla Test 3 ve Test 4).

Çizelge 5.1, 5.2 ve 5.3 incelendiğinde sınıflandırma başarımlarının “İlgili”, “İlgisiz” şeklinde raporlandığı testlerde (Test 3 ve Test 4) daha yüksek doğruluk elde edildiği

görülmektedir. Bu durum baş dönüş hareketinin yakın dönüş açılı kayıtlarda daha yüksek oranda yanlış sınıflandırılmasından kaynaklanmaktadır. $\pm 15^\circ$ ve $\pm 30^\circ$, $\pm 45^\circ$ ve $\pm 60^\circ$ sınıfları birbirleriyle karışmaktadır. Yine aynı tablolara bakıldığında 10 çapraz doğrulama testleri ile başarımın arttığı gözlemlenmektedir (Test 2 ve Test 4). Bunun nedeni veri tabanında yer alan bazı kişilerin kayıtlarında başarımın çok düşük olması ve bu durumun sınıflandırma başarımına etkisinin 10 çapraz doğrulama ile azaltılmasıdır.

Çizelge 5.1: “Pan” ve “tilt” hareketleri için elde edilen ortalama başarım (F ölçütü).

Hareket	Test 1	Test 2	Test 3	Test 4
Pan ($\pm 60^\circ, \pm 45^\circ, \pm 30^\circ \pm 15^\circ$)	0,66	0,75	0,72	0,79
Tilt ($\pm 60^\circ, \pm 30^\circ \pm 15^\circ$)	0,60	0,77	0,62	0,79

Çizelge 5.2 : “Tilt” açısı sabit tutularak “pan” ” hareketi için elde edilen ortalama başarım (F ölçütü).

Hareket	Test 1	Test 2	Test 3	Test 4
Tilt= -15° , pan ($\pm 60^\circ, \pm 45^\circ, \pm 30^\circ \pm 15^\circ$)	0.62	0,75	0.66	0.77
Tilt = -30° , pan ($\pm 60^\circ, \pm 45^\circ, \pm 30^\circ \pm 15^\circ$)	0.62	0,75	0.68	0.77
Tilt= -60° , pan ($\pm 60^\circ, \pm 45^\circ, \pm 30^\circ \pm 15^\circ$)	0.59	0,72		
Tilt = 15° , pan ($\pm 60^\circ, \pm 45^\circ, \pm 30^\circ \pm 15^\circ$)	0.64	0,74	0.69	0.77
Tilt = 30° , pan ($\pm 60^\circ, \pm 45^\circ, \pm 30^\circ \pm 15^\circ$)	0.59	0,73	0.60	0.75
Tilt = 60° , pan ($\pm 60^\circ, \pm 45^\circ, \pm 30^\circ \pm 15^\circ$)	0.55	0,70		
Ortalama	0,60	0,73	0,66	0,765

Çizelge 5.3 : “Pan” açısı sabit tutularak “tilt” hareketi için elde edilen ortalama başarımlar (F ölçütü).

Hareket Türü	Test 1	Test 2	Test 3	Test 4
Pan = -15° , Tilt ($\pm 60^\circ, \pm 30^\circ \pm 15^\circ$)	0.57	0,72	0.57	0.74
Pan = -30° , tilt ($\pm 60^\circ, \pm 30^\circ \pm 15^\circ$)	0.58	0,72	0.58	0.73
Pan = -45° , tilt ($\pm 60^\circ, \pm 30^\circ \pm 15^\circ$)	0.57	0,71		
Pan = -60° , tilt ($\pm 60^\circ, \pm 30^\circ \pm 15^\circ$)	0.57	0,71		
Pan = 15° , tilt ($\pm 60^\circ, \pm 30^\circ \pm 15^\circ$)	0.57	0,72	0.57	0.73
Pan = 30° , tilt ($\pm 60^\circ, \pm 30^\circ \pm 15^\circ$)	0.57	0,72	0.57	0.73
Pan = 45° , tilt ($\pm 60^\circ, \pm 30^\circ \pm 15^\circ$)	0.59	0,73		
Pan = 60° , tilt ($\pm 60^\circ, \pm 30^\circ \pm 15^\circ$)	0.58	0,73		
Ortalama	0,575	0,72	0,57	0,73

Çizelge 5.4 ve 5.5’de dönüş hareketi sınıflandırma başarımları çerçeve bazında raporlanmaktadır. Amaçlanan her video çerçevesi için hesaplanan onlarca akış vektörü için karar vermek yerine çoğunluklu oylama ile video çerçevesine karşı düşen tek bir baş dönüş hareketi etiketinin belirlenmesidir. Başarımlar önceki teste benzer şekilde farklı ikili sınıflar için ve ilgili-ilgisiz sınıflandırma için R, P ve F-ölçütü cinsinden raporlanmaktadır. Önceki testten farklı olarak ikili sınıflandırmada alınan test video çerçevesinin tüm sınıflardan hangisine ait olabileceği şeklinde sorgulama yapılmaktadır (authentication).

Çizelge 5.4 : Pointing'04 eğitim ve test veri seti ile “pan” hareketi için çerçeve bazında elde edilen baş hareketi ikili sınıflandırma başarımı (Test A).

8 sınıf başarımı	Sınıf	P	R	F	İlgili-İlgisiz başarımı	Sınıf	P	R	F
	1	0,75	0,2	0,32		İlgisiz	0,98	0,8	0,88
	2	0,35	0,53	0,42					
	7	0,4	0,4	0,4					
	8	0,58	0,27	0,36		İlgili	0,83	0,98	0,90
	3	0,67	0,67	0,67					
	4	0,39	0,8	0,52					
	5	0,71	0,67	0,69					
	6	0,78	0,47	0,58					
	Ort.	0,58	0,5	0,5		Ort.	0,91	0,89	0,89

Çizelge 5.5 : Pointing'04 eğitim ve test veri seti ile “tilt” hareketi için çerçeve bazında elde edilen baş hareketi ikili sınıflandırma başarımı (Test B).

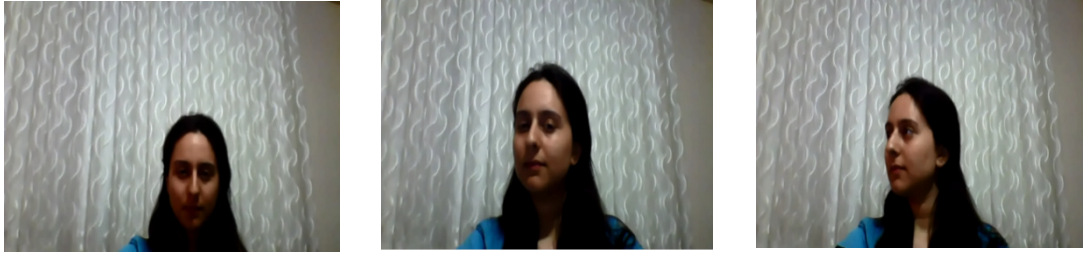
6 sınıf başarımı	Sınıf	P	R	F	İlgili-İlgisiz başarımı	Sınıf	P	R	F
	1	0,66	0,53	0,59		İlgili	0,83	0,87	0,85
	2	0,6	0,2	0,3					
	3	0,5	0,34	0,41					
	4	0,24	0,4	0,3		İlgisiz	0,7	0,63	0,66
	5	0,24	0,27	0,25					
	6	0,5	0,47	0,49					
	Ort.	0,46	0,37	0,41					

Çizelge 5.4’de görüldüğü gibi açısal sınıflardaki başarımlar düşüktür. Ancak ilgililik – ilgisizlik şeklindeki ikili sınıflandırma başarımı arttırmaktadır. Buradan da anlaşılmaktadır ki açısal düzeyde sınıfların ayırımı başarılı olmamakla beraber karıştırılan sınıflar aynı ilgililik sınıfına ait olduğu için ilgililik başarımı oldukça yüksektir.

Çizelge 5.5’de veri tabanı eğitim ve test seti ile “tilt” hareketi için elde edilen başarımlar raporlanmaktadır. Görüldüğü gibi ikili dönüş açıları cinsinden elde edilen başarımlar “pan” hareketinde olduğu gibi düşüktür. Ayrıca ilgililik sınıflandırma başarımları da özellikle ilgisiz sınıfı için pan hareketinden düşüktür. Bunun nedeni “tilt” hareketi için özellikle yüksek dönüş açılarında hareket kestirim doğruluğunun düşmesidir.

5.1.2 Test video kayıtlarında başarımlar

İkinci grup testlerde amaçlanan, Pointing’04 veri tabanında yapılan sınıflandırıcı eğitimi kullanılarak, kaydedilen yeni test videolarında baş dönüş hareketi sınıflandırma başarımlarını ölçmektir. Bu amaçla bir önceki testte Pointing’04 eğitim veri tabanı ile ikili olarak eğitilen sınıflandırıcılar kullanılmıştır. Test seti olarak ise yapılan yeni test video kayıtlarından yararlanılmıştır (Şekil 5.1). Videolarda arka planın etkisini yok etmek amacıyla yüz Viola-Jones ile bulunmuş ve sadece ilgili bölgelerde hareket kestirimi yapılmıştır. Çizelge 5.8’de “pan” hareketi içeren test video kayıtlarında elde edilen başarımlar raporlanmaktadır.



Şekil 5.1 : Gerçek video çerçeveleri örnekleri.

Çizelge 5.6: Pointing’04 veri tabanında eğitilen sınıflandırıcılar ile test video kayıtlarında dönüş sınıflandırma başarımları (Test C).

İlgili-İlgisiz Sınıfları	Sınıf	P	R	F
	İlgili	0,90	0,66	0,76
	İlgisiz	0,36	0,72	0,48
	Ort.	0,63	0,69	0,62

Çizelge 5.4’de ve çizelge 5.6’da raporlanan test başarımları karşılaştırıldığında, Pointing’04 veri tabanı ile yapılan testlerde gerçek videolarda yapılan testlere göre %30’a yakın daha yüksek başarımlar gözlemlenmektedir. Bunun nedeni, test seti olarak veri tabanı kullanıldığında, eğitim ve test setinin aynı kişilerden oluşmasının başarımları arttırmasıdır. Ayrıca gerçek videolarda ışıklılık değişimleri yüksektir ve etiketleme göz ile yapılmıştır bu nedenleyanlışlık yapılmış olması olasıdır.

5.2 Baş Konum Sınıflandırma Başarım Testleri

Yüz konum sezme için Violen-Jones algoritması ile yüz konumu saptanmakta ve OpenCV ortamında oluşturulan algoritma ile bulunan yüz konumu 3 boyutlu öznitelikler ile temsil edilmektedir. Elde edilen 3 boyutlu öznitelikler ile sınıflandırma yapılmış ve başarımlar raporlanmıştır. Tez çalışmasında, yüzün video çerçevesinin yan bölgesinde olma durumu “olası ilgisiz”, orta bölgesinde olma durumu “olası ilgili” olarak etiketlenmiştir. Yüz konum sezme sınıflandırmasında, eğitim seti olarak sınıf etiketleri tarafımızca belirlenen video kayıtları, test seti olarak ise e “pan” hareketi içeren video kaydı kullanılmıştır. Viola-Jones algoritması, baş dönüş hareketinin çok fazla olduğu ve video çerçevesinin dışına çıktığı durumlarda yüz yakalayamamıştır.. Bu video çerçeveleri test setine katılmamıştır.. Yüz konumunun sezildiği video çerçeveleri için elde edilen baş konum sınıflandırma sonuçları çizelge 5.7’de görülmektedir.

Çizelge 5.7: Yüz konum sınıflandırma başarımları (Test D).

İlgili-İlgisiz Sınıfları	Sınıf	P	R	F
	Olası İlgili (Orta)	0,89	0,93	0,91
	Olası İlgisiz (Yan)	0,88	0,82	0,85
	Ort.	0,89	0,88	0,88

Viola Jones algoritması ile yüz yakalanmayan video çerçevelerinde baş dönüş hareketi sınıflandırma %74 doğrulukla gerçekleşmiştir.. Ayrıca yüz konum özniteliklerine göre yanlış sınıflandırılan video çerçevelerinde baş dönüş hareketi sınıflandırmada doğru bulma başarımları %62 olarak hesaplanmıştır.

5.3 Karar Tümeştirme ile Baş Dönüş Hareketi Sınıflandırma Başarımı

İlgililik sınıflandırma doğruluğunu arttırmak amacıyla hareket ve konum sınıflandırıcı çıkışlarında karar tümeştirme gerçekleştirilmiştir. Baş dönüş hareketi ve yüz konum sınıflandırıcı kararları tümeştirilirken sınıf etiketleri çizelge 5.8’de görüldüğü şekilde belirlenmiştir. Tümeştirmede hareket sınıflandırma eğitim seti olarak Pointing’04 veri tabanı ve test seti olarak gerçek videolar kullanılmıştır. Yüz konum sınıflandırma için ise eğitim ve test seti olarak gerçek videolar kullanılmıştır.

Çizelge 5.8 Viola-Jones ve SIFT akış öznelikleri, tümeşik sınıf etiketleri.

Viola Jones Algoritması	SIFT Akış Algoritması	Sınıf
Yan (Olası İlgisiz)	-	İlgisiz
Orta (Olası İlgili)	30 derece üstü	
	30 derece ve altı	İlgili

Çizelge 5.9: Viola-Jones ve SIFT akış tümeşik başarımları (Test E).

İlgili-İlgisiz Sınıfları	Sınıf	P	R	F
	İlgili	0,78	0,59	0,67
	İlgisiz	0,67	0,83	0,74
	Ort.	0,73	0,76	0,71

Tümeştirme etkisi Çizelge 5.9’den gözlemlenebilir. Sadece optik akış ile baş dönüş sınıflandırma yapıldığında %62 olan ortalama F ölçütü konum sınıflandırma ile tümeştirdiğinde %9 oranında artmaktadır. Ayrıca her video karesi için SIFT akış algoritması uygulanmadığından test süresi azalmaktadır. Çizelge 5.10’da karşılaştırma açısından Test A, Test C ve Test E sonuçları birlikte raporlanmaktadır.

Çizelge 5.10: Karar tümleştirme ve baş dönüş sınıflandırma başarımı (ortalama F ölçütü).

	Pointing'04Veri TabanındaBaş Dönüş Sınıflandırma (Test A)	Gerçek Test Videolarında Baş Dönüş Sınıflandırma (Test C)	Karar Tümleştirme (Test E)
P	0,91	0.69	0.76
R	0,89	0.63	0.73
F	0,89	0.62	0.71

6. SONUÇ VE ÖNERİLER

Tez kapsamında, ilgili video çerçevelerinden elde edilen öznitelikler ile sınıflandırıcıların eğitilmesi ve test çerçevelerinin bu sınıflandırıcılar ile sınıflandırılması sonucu elde edilen başarımın uygunluğu tartışılmıştır. Eğitim veri tabanı olarak baş duruş açıları incelikle hesaplanmış ve etiketlenmiş video çerçevelerinin bulunduğu Pointing'04 eğitim veri tabanı ve gerçek videolar, test seti olarak Pointing'04 test veri tabanı ve gerçek videolar kullanılmıştır.

Bu çalışmada amaç, yüz konum sezme ve baş dönüş hareket kestirimi özniteliklerinin tümleştirilerek ilgililik sınıflandırma başarımının hesaplanmasıdır. Tümleştirme işlemi ile ilgililik sınıflandırma başarı, yalnız baş hareket kestirimi öznitelikleri ile elde edilen başarım ile karşılaştırılarak, tümleştirme yönteminin üstünlüğü gösterilmiştir. Baş hareket kestirimi için Pointing'04 veri tabanı ile yapılan testlerde, ilk olarak amaçlanan, test veri setinin belirlenen iki sınıftan birisine ait olup olmadığının sınılanmasıdır. Bu testte başarımın ortalama F ölçütü cinsinden öznitelik vektörü bazında hesaplanmıştır.

- Bütün sınıflar ayrı ayrı kabul edilmiş ve tüm ikili sınıflandırma başarımın ortalaması alınarak başarım hesaplanmıştır. Bu durumda başarım pan için %66 ve tilt için %60 olarak raporlanmıştır.
- Eğitim ve test verileri birleştirilerek 10 çapraz doğrulama testleri yapılmış ve tüm ikili sınıflandırma başarımın ortalaması alınarak başarım hesaplanmıştır. Bu durumda başarım pan için, %75 ve tilt için %77 olarak raporlanmıştır.
- Bütün sınıflar ayrı ayrı kabul edilmiş ve sadece farklı ilgililik sınıfına ait ikili sınıfların başarımın ortalaması alınarak başarım hesaplanmıştır (ilgililik sınıflandırma) Bu durumda başarım pan hareketi için %72 ve tilt hareketi için %62 olarak raporlanmıştır .
- Eğitim ve test verileri birleştirilerek 10 çapraz doğrulama testleri yapılmış ve sadece farklı ilgililik sınıfına ait ikili sınıfların başarımın ortalaması

alınarak başarıml hesaplanmıştır (İlgililik sınıflandırma). Bu durumda başarıml pan için, %79 ve tilt için %79 olarak raporlanmıştır.

Görüldüğü gibi 10 çapraz doğrulama testleri ile başarıml artmıştır. Bunun nedeni kullanılan eğitim setlerinde bazı kişilere ait verilerin başarımlının düşük olmasıdır. Ayrıca ilgililik sınıflandırma yapıldığında da başarımlın arttığı gözlemlenmektedir. Bu durumda söylenebilirki, aynı üst sınıfa ait sınıfların sınıflandırılma başarımlı oldukça düşüktür. İlgililik sınıflandırmada amaç ilgili-İlgisiz sınıflandırma olduğu için her ikisinde aynı ilgililik sınıfına ait açların sınıflandırma başarımlarının yüksek olması gerekli değildir. Çünkü bu sınıflandırmalarda sonuç ne olursa olsun atanan ilgililik sınıfı aynı olacaktır.

Tezde asıl yapılmak istenilen yüz konum sezme ve hareket kestirimi öznitelikleri tümleştirme işlemi, gerçek test videoları ile yapılmıştır. Veri tabanı ile yapılan eğitim-test testleri başarımlında en yüksek başarıml pan hareketinde gözlemlendiğinden ilk olarak gerçek videolar ile başarıml testleri pan hareketi için uygulanmıştır.

Karşılaştırma yapabilmek ve tümleştirme işlemi ile daha başarılı sonuçlar elde edilip edilmediğini saptayabilmek için öncelikli olarak sadece hareket kestirimi öznitelikleri ile başarıml hesaplanmıştır ve başarıml %62 olmuştur. Tümleştirme işlemi ile öncelikle yüz konum sezimi yapılp bu sezim ile 'ortada' olarak etiketlenen video çerçevelerinde hareket kestirimi yaptığımızda sonuç %71'e yükselmiştir.

Gerçek videolar ile yapılan testler şuan için sadece pan hareketi için yapılmıştır. Diğer yönler içinde başarımların hesaplanarak raporlanması gerekmektedir. Ayrıca şuan için tek bir test videosu ile yapılan testlerin, başarımlı daha kuvvetli olarak doğrulamak amacıyla, amaca yönelik başka videolarda da yapılması gerekmektedir. Daha uzun test videolarında başarımlın raporlanması ile güvenilirlik sınırlarının belirlenmesi sağlanacaktır.

Bunların yanı sıra ek öznitelikler eklenerek doğruluğun artırılması ve zamansal bölütlemenin sisteme eklenmesi ile anlık değişimlerden etkilenmenin giderilmesi sağlanmalıdır.

SIFT akış kestirimde, akış kestirim yoğunluğunu düşürerek karmaşıklık azaltılması sağlanabilir. Yani şuan her piksel için hesaplanan akış vektörleri sayısı azaltılarak karmaşıklık azaltılabilir.

Karmaşıklık azaltılması için yapılabilecek bir diğel alıřma ise zamansal rnekleme sıklıđını azaltarak daha az video erevesinde hareket kestirimi ve konum kestirimi yapmaktır.

Yapılacak tm bu alıřmalar ile bařarımın daha da arttırılması ve daha hızlı alıřan bir sistem oluřturulması amalanmaktadır.

KAYNAKLAR

- [1] **Ho, H. T. ve Chellappa, R.** (2012) Automatic Head Pose Estimation Using Randomly Projected Dense SIFT Descriptors. Department of Electrical & Computer Engineering Center for Automation Research , UMIACS , University of Maryland , College Park , MD 20742 (sf: 153–156).
- [2] **Hernandez, J., Liu, Z., Hulten, G., Debarr, D., Krum, K., ve Zhang, Z.** (2013). Measuring the engagement level of TV viewers. *10th IEEE Int. Conf. Work. Autom. Face Gesture Recognition*.
- [3] **Jo, J., Lee, S. J., Jung, H. G., Park, K. R., ve Kim, J.** (2011). Vision-based method for detecting driver drowsiness and distraction in driver monitoring system. *Opt. Eng.*, vol. 50, no. 12, p. 127202.
- [4] **Viola P. ve Jones, M.** (2001). Rapid object detection using a boosted cascade of simple features. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. CVPR*, (cilt 1).
- [5] **Liu, C., Yuen, J., Torralba, A., Sivic, J. ve Freeman, W. T.** (2008). SIFT Flow : Dense Correspondence across Different Scenes. *IEEE Transactions on Pattern Analysis and Mavhine Intelligence* .(cilt: 1, sayı: 1, sf: 28–42).
- [6] **Lowe, D. G.** (2004). Distinctive Image Features from Scale-Inveriant Keypoints.
- [7] **Doria, D.** (2011). Belief Propagation on MRFs. *The Insight Journal*. (sayı: 1, sf: 1–6).
- [8] **Simoncelli, E. P., Adelson, E. H. ve Heeger, D. J.** (1991). Probability Distributions of Optical Flow. *IEEE Conference on Computer Vision and Pattern Recognition, Maui, Hawaii*. (sf: 310-315).
- [9] **Sun, J., Zheng, N. ve Shum, Y.** (2003). Stereo matching using belief propagation. *IEEE Trans. Pattern Anal. Mach. Intell.* (cilt:25 sayı:7 sf: 787–800).
- [10] **Breiman, L.** (2001). Random Forrest. *Mach. Learn.* (sf: 1–33).
- [11] **Url-1** <<http://www-prima.inrialpes.fr/Pointing04/data-face.html>>, alındığı tarih: 2013.
- [12] **Lucas B. D. ve Kanade, T.** (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th international joint conference on Artificial intelligence (IJCAI'81)* , (cilt: 2. sf: 674–679).

- [13] **Horn B. K. P. ve Schunck, B. G.** (1993). Determining optical flow. *Artif. Intell.* (cilt: 59, sayı: 1–2, sf: 81–87).
- [14] **Yedidia, J. S., Freeman, W. T. ve Weiss, Y.** (2001). Understanding Belief Propagation and its Generalizations. *Exploring Artificial Intelligence in the New Millennium.* (sf: 239-269).
- [15] **Weiss Y. ve Freeman, W. T.** (2001). On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *IEEE Trans. Inf. Theory* (cilt: 47, sayı: 2, sf: 736–744).
- [16] **Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Tappen, M. ve Rother, C.** (2006). A comparative study of energy minimization methods for markov random fields. *Proc. Eur. Conf. Comp. Vis.* (sf: 1–17).
- [17] **Felzenszwalb, P., Huttenlocher, D.** (2004). Efficient Belief Propagation for Early Vision. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference* (cilt :1, sf: 261-268).
- [18] **Boccignone, G., Marcelli, A., Napoletano, P., Ponte, V., Melillo, D., Sperimentale, F. ve Giuria, V. P.** (2007). Motion Estimation via Belief Propagation,” *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on.* (sf: 55-60)
- [19] **Boykov, Y., Veksler, O. ve Zabih, R.** (2001). Minimization via Graph Cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* (cilt: 23, sayı: 11, sf: 1–18).
- [20] **Kschischang, F. R., Frey, B. J. ve Loeliger, H.** (2001). Factor graphs and the sum-product algorithm. *IEEE Trans. Inf. Theory.* (cilt: 47, sayı: 2, sayfa: 498–519).
- [21] **Kolmogorov, V.** (2006). Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.* (cilt: 28, sayı:10, sayfa: 1568–1583).
- [22] **Kolmogorov, V.** (2005). On the optimality of tree-reweighted max-product message-passing. *Artificial Intelligence.*
- [23] **Coughlan, J.** (2009). A Tutorial Introduction to Belief Propagation. *The Smith-Kettlewell Eye Research Institute.*
- [24] **Kolmogorov, V** (2007) . Message-passing algorithms for energy minimization. *ICCV07_tutorial_Part 3* .
- [25] **Lowe, D. G.** (1999). Object Recognition from Local Scale-Invariant Features. *IEEE Int. Conf. Comput. Vis.*(cilt: 2, sf:1150-1157).

- [26] **Kocadere, G.** (2010). Video Görüntülerinin Sabitlenmesi (Yüksek Lisans Tezi).Adres: <http://acikarsiv.ankara.edu.tr/browse/24466/>
- [27] **Bruhn, A., Weickert, J. ve Schnörr, C.** (2005). Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *Int. J. Comput. Vis.* (cilt: 61, sayı: 3, sf: 1–21).
- [28] **Shekhovtsov, A.** (2007). Efficient MRF Deformation Model for Non-Rigid Image Matching. *Computer Vision and Pattern Recognition, 2007. CVPR '07* . (sf: 1–6).
- [29] **Speybroeck, N.** (2012). Classification and regression trees. *Int. J. Public Health.* (cilt: 57, sayı: 1, sf: 243–246).
- [30] **Crow, F. C.** (1984). Summed-area tables for texture mapping. *Comput. Graph.* (cilt: 18, sayı: 3, sf: 207–212).
- [31] **Freund Y. ve Schapire, R.** (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. *Comput. Learn. theory.* (cilt: 55, sayı: 1, sf: 119–139).
- [32] **Papageorgiou, C. P., Oren, M. ve Poggio, T.** (1998). A general framework for object detection. *Sixth Int. Conf. Comput. Vis. IEEE.* (cilt: 6, sf: 555–562).
- [33] **Baytaş, İ. M., Günsel, B.** (2014). Head Motion Classification with 2D Motion Estimation. *Signal Processing and Communications Applications Conference (SIU)*. (sf: 325-328).
- [34] **Url-2** <<http://people.revoledu.com/kardi/tutorial/DecisionTree/how-to-measure-impurity.htm>> alındığı tarih: 2014.

ÖZGEÇMİŞ



Ad Soyad : Filiz Gürkan
Doğum Yeri ve Tarihi : Eskişehir-11.07.1990
E-Posta : fgurkan90@gmail.com

ÖĞRENİM DURUMU:

- **Lisans** : 2012, Ege Üniversitesi, Mühendislik Fakültesi Elektrik Elektronik Mühendisliği,
- **Yükseklisans** : 2015, İstanbul Teknik Üniversitesi, Elektronik ve Haberleşme Mühendisliği, Telekomünikasyon Mühendisliği

MESLEKİ DENEYİM VE ÖDÜLLER:

(2012-2013) Vestel Elektronik A.Ş.- Elektrik Test Mühendisi
(2015-Halen) Medeniyet Üniversitesi - Araştırma Görevlisi

YAYIN VE PATENT LİSTESİ

Gürkan F., Günsel B., Başyurt E. (2015, May) . Hareket ve Yerleşim Öznitelikleri Tümlleştirilerek Baş Hareketi Sınıflandırma. In *Signal Processing and Communications Applications Conference (SIU)*.