SURFACE REFLECTANCE ESTIMATION FROM SPATIO-TEMPORAL SUBBAND STATISTICS OF MOVING OBJECT VIDEOS

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND

ELECTRONICS ENGINEERING

AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCES

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Onur Külçe

August 2012

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Levent Onural (Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Katja Doerschner(Co-supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Ergin Atalar

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Hüseyin Boyacı

Approved for the Graduate School of Engineering and Sciences:

Prof. Dr. Levent Onural
Director of Graduate School of Engineering and Sciences

ABSTRACT

SURFACE REFLECTANCE ESTIMATION FROM SPATIO-TEMPORAL SUBBAND STATISTICS OF MOVING OBJECT VIDEOS

Onur Külçe
M.S. in Electrical and Electronics Engineering
Supervisor: Prof. Dr. Levent Onural
Co-Supervisor: Assist. Prof. Dr. Katja Doerschner
August 2012

Image motion can convey a broad range of object properties including 3D structure (structure from motion, SfM), animacy (biological motion), and its material. Our understanding of how the visual system may estimate complex properties such as surface reflectance or object rigidity from image motion is still limited. In order to reveal the neural mechanisms underlying surface material understanding, a natural point to begin with is to study the output of filters that mimic response properties of low level visual neurons to different classes of moving textures, such as patches of shiny and matte surfaces. To this end we designed spatio-temporal bandpass filters whose frequency response is the second order derivative of the Gaussian function. Those filters are generated towards eight orientations in three scales in the frequency domain. We computed responses of these filters to dynamic specular and matte textures. Specifically, we assessed the statistics of the resultant filter output histograms and calculated the mean, standard deviation, skewness and kurtosis of those histograms. We found that there were substantial differences in standard deviation and skewness of specular

and matte texture subband histograms. To formally test whether these simple measurements can in fact predict surface material from image motion we developed a computer-assisted classifier based on these statistics. The results of the classification showed that, 75% of all movies are classified correctly, where the correct classification rate of shiny object movies is around 77% and the correct classification rate of matte object movies is around 71%. Next, we synthesized dynamic textures which resembled the subband statistics of videos of moving shiny and matte objects. Interestingly the appearance of these synthesized textures were neither shiny nor matte. Taken together our results indicate that there are differences in the spatio-temporal subband statistics of image motion generated by rotating matte and specular objects. While these differences may be utilized by the human brain during the perceptual process, our results on the synthesized textures suggest that the statistics may not be sufficient to judge the material qualities of an object.

Keywords: The Human Visual System, Surface Reflectance, Movie Subband Statistics, Three-Dimensional Second Order Derivative of Gaussian Filter, Texture Synthesis, Steerable Pyramid

ÖZET

HAREKET EDEN NESNE VİDEOLARININ ALTBAND İSTATİSTİKLERİ KULLANILARAK YÜZEY YANSITMA ÖZELLİĞİNİN BELİRLENMESİ

Onur Külçe
Elektrik ve Elektronik Mühendisliği Bölümü Yüksek Lisans
Tez Yöneticisi: Prof. Dr. Levent Onural
Yardımcı Tez Yöneticisi: Assist. Prof. Dr. Katja Doerschner
Ağustos 2012

Nesnelerin hareket görüntüleri, 3 boyutlu (3B) yapı, biyolojik hareket, maddenin yapısı gibi nesne özelliklerini kapsayan geniş bir alanda bilgi verebilir. Görsel sistemin, yüzey yansıması veya nesnenin katılığı gibi karmaşık özellikleri nasıl algıladığı ile ilgili bilgimiz ise sınırlıdır. Maddenin yüzey yapısının algılanmasını sağlayan sinirsel mekanizmanın ortaya çıkarılması için doğal başlangıç noktası, mat ve parlak yüzeyleri içeren farklı sınıflardaki hareketli dokulara uygulanan, düşük seviye görsel sinir hücrelerini taklit eden süzgeçlerin çıkışlarını incelemektir. Bu amaçla, frekans tepkisi Gauss'un ikinci dereceden türevi olan mekan-zamansal band-geçirgen süzgeçler tasarladık. Bu süzgeçler, frekans alanında sekiz yönde ve üç ölçekte tasarlandılar. Bu süzgeçlerin hareketli parlak ve mat dokulara olan çıktılarını bulduk. Özellikle, süzgeçlerin çıkışlarındaki *histogramların* istatistiklerini değerlendirdik ve histogramların ortalama, standart sapma, yamukluk ve sivrilik (basıklık) ölçülerini hesapladık. Parlak ve mat doku altband histogramları arasında standart sapma ve yamukluk açısından belirgin farklar olduğunu gözlemledik. Bu basit ölçümlerin gerçekte parlak veya mat nesnelerin

tanınması için kullanılıp kullanılamayacağını anlamak için, bu istatistikleri kullanan bilgisayar tabanlı bir sınıflandırıcı geliştirdik. Sonuçlar kullanılan filmlerin %75'inin doğru olarak sınıflandırılabildiğini gösterdi. Mat nesnelerin doğru olarak sınıflandırma oranı %71 iken parlak nesnelerin doğru sınıflandırılma oranı %77 olarak bulundu. Daha sonra, altband istatistikleri açısından parlak ve mat nesnelere benzeyen hareketli dokular üretildi. İlginç bir şekilde, üretilmiş filmler ne mat ne de parlak bir görünüme sahipti. Bu sonuçlar tümden ele alındığında, dönen parlak ve mat nesnelerin mekan-zamansal altband istatistikleri arasında fark olduğu gözlemlendi. Beyin bu farkları, algılama sürecinde kullanabileceği halde, üretilen dokular bize, istatistikler nesnenin madde yapısının kalitesini anlamak için yeterli olmayabileceğini gösterdi.

Anahtar Kelimeler: İnsan Görsel Sistemi, Yüzey Yansıması, Film Altband İstatistikleri, Üç boyutlu Gauss'un İkinci Dereceden Türevi, Doku Sentezi, Döndürülebilir Piramit.

ACKNOWLEDGMENTS

First of all, I would like to thank to my supervisors, Prof. Levent Onural and Assist. Prof. Katja Doerschner for their invaluable support in every stage of this thesis work. I learned a lot from them, but, maybe the most important one is, they showed that how a scientific problem should be approached.

Secondly, I would like to thank to Prof. Ergin Atalar and Assist. Prof. Hüseyin Boyacı for accepting to read this thesis and their involvements of my thesis committee.

Furthermore, I thank to Dr. Özgür Yılmaz for his contributions in this thesis. He helped me a lot especially in the technical parts of this thesis.

I also would like to thank to my family for their endless support throughout my life. I wish I am going to be a parent like them.

I also would like to thank to all my friends who are in Ankara and made me fun even if I had been worried about my future. I also thank to my old friends who are in İstanbul or in abroad. Although we are not in the same town, it is a great happiness for me to be friend with them. Moreover, I thank to my officemates and all other people who make UMRAM a very nice place to work and enjoy.

Finally, I want to thank to TUBITAK as well for supporting me financially through my MS degree program.

Contents

List of Figures

List of Tables

Dedicated to My Family

Chapter 1

INTRODUCTION

Understanding the visual system and visual perception has been the interest of many scientists and philosophers, such as Aristotle, Plato, Leonardo DaVinci or Hermann von Helmholtz. In 1981, the Nobel Prize in Physiology or Medicine was given to David H. Hubel and Torsten Wiesel for their contributions to understanding the visual system. They showed that neurons in the visual cortex respond to very specific stimuli, such as bars at different orientations or moving edges, in particular locations of the visual field called the receptive field. They also proposed that the visual system is organized hierarchically, where *simple* cells respond to basic stimulus properties, such as brightness and orientation of edges, and *complex* cells that receive feedforward input from these simple cells respond to motion of oriented edges [1]. More recently many higher level neurons with more complex properties of the receptive field have been discovered, such as those respondent exclusively to faces [2]. One aim of research in vision is to understand how complex visual phenomena such as color, object or surface material are estimated from the visual input, i.e., to understand the specific processing hierarchy. In this thesis, we examine whether simple brightness intensity statistics, such as those that may be computed by simple cortical cells, of moving objects can account for the appearance of surface reflectance.

We next give explanations of the terms *surface reflectance*, *statistics of visual stimuli*, *motion* and *subband*.

Surface Reflectance: When we see the surface of an object, we can easily judge the material of that object based on its surface reflectance properties and we can tell, for example, whether the object is made up shiny, matte or velvet material. Surface material provides important information about how to interact or evaluate a given object. For example, stucco which has bad quality is often appears less shiny than stucco which is of good quality. In trying to understand whether a car had crashed before one often examines discontinuities in shiness on its surface.

In electromagnetics terminology, reflectivity of a surface can be defined as the ratio between the energy of the reflected light to the incident light [3]. In computer vision, surface reflectance is often modeled as a *bidirectional reflectance distribution function* (BRDF). The BRDF includes four parameters, azimuth and zenith angles of the light source and azimuth and zenith angles of the observers. These angles are defined with respect to surface normal, where the reflection occurs.

Shininess or opaqueness of a surface is determined by its reflectance properties. That is, if all radiated energy is contained in reflected light beam which has the same angle with incoming light, then, the surface seems perfectly shiny. This type of reflectance is called *specular reflectance*. On the other hand, if reflected light beams carry equal amount of energy in every direction, then, the surface seems perfectly opaque. This type of reflectance is called *diffuse reflectance*. A surface may have both of those reflectance types at the same time. That is, if the degree of the specular reflectance increases, it seems shinier and if the degree of

the diffuse reflectance increases, it seems more matte. For example, if most of radiated energy is contained in the light beams whose reflection angles are close to angle of incoming light, the surface can still seem shiny. In addition, there can be some situations that cannot be decided whether the object is matte or specular. In such a situation, it can be said that the surface has nearly same specular and diffuse reflectance amounts and it seems semi-matte. These reflectance models are included in the BRDF and more details about the BRDF as well as specular and diffuse reflectance types can be found in [4].

Surface reflectance is a crucial piece of information for many computer vision algorithms. For example, shape from shading is a method that provides recovery of shape of an object from its reflectance properties, assuming diffuse reflectance (Lambertian) [5, 6, 7]. Other algorithms have assumed known specular reflectance for estimating shape [8]. However, how to estimate surface reflectance from movies directly is largely unstudied. This thesis is one of those attempts.

Statistics of Visual Stimuli: It is reasonable to assume that the visual system is adapted to the environment that we live in. Therefore, exploring the properties of the visual input can help to understand the working mechanism of the human Visual system (HVS). The distribution of brightness intensities of the visual input is, for example, a simple property that has been examined by several researchers.

In 1954, Attneave indicated that the visual information in a natural scene is highly redundant. The word *redundant* means that natural scenes have recurrent characteristics in terms of color, brightness and shape. For example, the sky in a sunny day is blue or a road is gray color. In order to increase the information gathered from a visual stimuli and storing the maximum information, the HVS should have an efficient coding algorithm to reduce that redundancy [9]. In 1961, Barlow formulized the suggestion of Attneave in a mathematical manner and

proposed the concept of *redundancy reduction* [10]. According to the principle, instead of responding and storing each part of a natural scene, the HVS shows reaction to the probabilistic characteristics of the visual stimuli. In this way, a scene which contains redundancies is abstracted according to its statistical characteristics and efficiently processed and perceived by the brain.

Previous research on visual scene statistics mainly focused on the first order (extracted from image histogram) and second order (extracted from correlation properties) statistics. More details on redundancy reduction, image statistics and coding/storing visual information can be found in [11, 12, 13, 14, 15].

Motion: The physical world around us and/or we are constantly in motion and so, motion perception is extremely crucial for many aspects in our life. For example, while we walk across the street, if we see that there is an oncoming fast moving car, we decide to walk faster or wait until the car passes by. So, how is motion computed by the brain? One possibility is that the HVS processes only instantaneous two-dimensional visual stimuli, and in this case, the perception of motion could be explained by simply investigating the perceptual mechansims of instaneous frames (or instances in time) independently. However, let us consider a disease called *akinetopsia*. People with this disease lack the perception of motion, although they can see standing objects. For example, they can see their hands when they put them onto a table. However, they cannot see the same hands while they are washing or waving them. Or, they cannot judge the speed of oncoming cars and they cannot see people moving in a room. Their perception is that of *snapshots*. Researches have shown that akinetopsia is caused by a lesion found in a certain part of the visual system in the brain, that is called as middle temporal region (MT) [16]. This region of the brain is in fact responsible for motion perception. Therefore, motion perception is more than independent perception of motion in two-dimensional consecutive frames.

What other kinds of information does image motion convey? Research has shown that motion is extremely crucial for estimating three-dimensional shape in structure from motion (SfM) [17, 18, 19, 20, 21, 22, 23, 24], perception of biological entities [25, 26, 27] and more recently also in material perception [28, 29, 30, 31]. What these works have in common is that they demonstrate that motion provides information, that cannot be extracted from still images.

Subband: In 1959, the Nobel laureates David H. Hubel and Torsten Wiesel conducted an experiment on the early stages of the visual system of a cat [1]. In the experiment, the cat was shown some lines at different angles and the activity of neurons in primary visual cortex and lateral geniculate nucleus (LGN) were recorded. The experimental results suggested that each different set of neurons responds to lines at different angles. They called these neurons as *simple neurons*. In addition, they also found that a different set of neuron responded to the lines which have different lengths. After that, they showed the cat moving lines which have different motion direction. As being in the case of different angled and length of lines, different sets of neurons responded the lines which have different motion direction. They called these neurons as *complex neurons*. At the end of their work, they reached the conclusion that the visual system decomposes a visual stimuli into its features which have different angles, lengths and motion directions, then process each decomposed part separately.

In the discipline of image processing and related fields, the behaviour of the single and complex neuron sets can be interpreted under the concepts of *orientation* and *scale*. In other words, it can be said that the single and complex neurons have orientation and scale selectivities. As we explain in Chapter 2, orientation and scale selectivity characteristics of the neurons can be modeled in the frequency domain. That is to say, a particular orientation and scale correspond to a particular location in the frequency plane. This location in the

frequency plane can be called as *subband*. In this thesis, we model the selectivities of the neuron sets by the digital filters designed in the frequency plane. Detailed information about the behaviour of the simple and complex cells as well as the details of the HVS can be found in [32] and [33].

Compilation of the Stated Concepts and Outline of this Thesis: The image that a moving object generates can convey important information about the appearance of that object, including its surface reflectance characteristic. For example, if an object is painted as though it reflects its environment, it can be perceived as shiny when it is stationary. However, when it moves, it rapidly becomes apparent that it does not reflect the environment specularly but that is it just painted. On the other hand, if the object surface was specular, reflections on the surface had specific motion characteristics and differed from the motion of the object. Example movies that show motion of painted objects and shiny objects can be found on the website `http://www.umram.bilkent.edu.tr/~kulce/`.

In this thesis, we try to understand the fairly complex perceptual attribute of surface material by examining the responses of the early visual system to moving matte and specular surfaces. We first mimic the response of the early stages of the HVS. To accomplish this, we design an image processing tool which consists of subband filters which have different orientation and scale selectivities similar to simple and complex visual neurons. In Chapter 2 we give the technical details of the filters. In Chapter 3, we test the hyphothesis that the statistics play role on the recognition of surface reflectance, by examining similarities and differences between subband statistics of the motions generated by the matte and shiny rotating objects.

The chapters are organized as follows:

- **Chapter 2:** This chapter includes the technical details about the image processing tool that we develop to model the early stages of the HVS. Since we design oriented filters in different scales, we explain what the orientation and scale is and their interpretation in the frequency domain. Secondly, we give the filter properties of three-dimensional derivative of Gaussian filters, such as orientation and scale bandwidth and associated design considerations. Then, since we used second derivative of Gaussian filters to model the HVS, we give explicit design steps of that filter. Next, since one of the aims of this thesis is to check whether the subband statistics are sufficient cues on the surface reflectance recognition, we need to synthesize new movies which have the same subband statistics with a selected shiny or matte object movie. The algorithm for synthesis requires the steerable pyramid, which is a subband decomposition/reconstruction filterbank. Therefore, we give technical details on the steerable pyramid and the analytical constraints that the steerable pyramid filters should have. Finally, we provide explicit design steps of adopted three-dimensional steerable filter design methods.

- **Chapter 3:** This chapter shows experimental results on the subband statistics of matte and shiny moving objects. We first describe our dataset and the statistical parameters that we use. Secondly, since our dataset consists of matte and shiny versions of an object, we give statistical differences of the motions of such pairs. After that, we examine the differences between the averages of the statistical parameters of matte object motions and the averages of the statistical parameters of shiny object motions. Then, we attempt to classify surface reflectance using motion statistics. Finally, we propose a motion syntesis algorithm based on subband histogram matching and discuss classification results on the synthesized textures.

- **Chapter 4:** We end with a brief summary, discussions of the results and possible future work.

The contribution of this thesis can be summarized as follows:

- We examine the role of subband motion statistics on surface reflectance appearance and develop a successful classifier based on the subband statistics. We show the limits of this approach using synthesized reflectance textures.
- We examine the filter properties of three-dimensional derivative of Gaussian filters in any derivative order and provide explicit design steps of first and second derivative of Gaussian fitlers.
- We develop two different design methods for three-dimensional steerable pyramid filters.

1.1 Previous Works on Surface Reflectance Recognition

We end this chapter by briefly reviewing previous work on human surface reflectance recognition.

Still Images: In [34], Sharan et al. examine the importance of diffuse reflectance through an image matching experiment and find that diffuse reflectance parameters affect surface reflectance recognition. Moreover, they work on the statistics extracted from histograms of subbands of complex photographs and develop a machine learning algorithm based on those statistics for separation of shiny objects from matte ones. In [35], Flemming et al. reveal that stored assumptions and previous knowledge of humans about real world illumination

statistics also affect the decision of whether an object is shiny or matte. In [36], Dror et al. propose another machine vision algorithm to differentiate matte and shiny surfaces. In their work, the surfaces that they use have arbitrary shapes and the illumination type is unknown. The pattern recognition algorithm that they develop also uses statistics of the images. In [37], Motoyoshi et al. find that perceived glossiness is highly correlated with the skewness of both image luminance histogram and image subband histograms. If the skewness of the histograms is high, then, the object seems more glossy and has less albedo. They also propose that there are mechanisms on the human visual system that are sensitive to skewness of the luminance histograms. However, in [38], Anderson and Kim have shown that unique statistics are not sufficient to explain the perception of surface gloss but that other higher level properties such as the alignment of highlights with the shading gradient are important. In [39], Adelson summarizes the recent works on the topic and emphasizes the importance of the subband statistics on the perception of glossiness in still images.

Motion: In [29], Doerschner et al. introduce three motion cues to identify surface material. These cues are extracted from optic flow characteristics of the objects and named as *coverage*, *divergence* and *3D shape reliability*. By using those characteristics of the optic flow, they developed a classifier algorithm and they can predict the shininess of an object. In [30], the authors classify materials as shiny or matte by examining their dominant direction of motion and motion velocities. They assume specular features as sliding onto the surface of the object while it is moving. A specular feature move faster than the object in flat regions and slower than object in convexly curved regions. In [31], Zang et al. work on perception of motion of nonrigid shiny and matte objects. They also use optic flow characteristics related to motion. In [28], Hartung et al. investigates surface reflectance perception according to the parameters; naturalness of the

illumination environment, consistency between background and reflection, and optic flow.

Chapter 2

FILTERS AND VIDEO PYRAMID

Images and videos are filtered for different purposes. In this thesis, since we desire to investigate statistical differences between subbands of motion of shiny and matte objects, Gaussian derivative filters are designed in different orientations and scales. Their advantageous properties are separability, steerability and having short lengths. Moreover, in order to see to what extent the subband statistics are important on the surface reflectance perception, a kind of hierarchical video decomposition architecture, *steerable pyramid* for three-dimensional signals, is designed.

In this chapter, after giving the preliminaries on the concept, we state the mathematical expressions and design steps of the Gaussian derivative filters. Then, we proceed with the general properties, usage and explanation of our adopted filter design techniques for the steerable pyramid. Detailed explanations and theorems on steerability of such filters can be found in [40], and the extension of the steerability property to three-dimensional separable filters is explained in [41]. The steerable pyramid is explained in details in [42].

2.1 Orientation and Scale

2.1.1 Orientation in 2-D Images

In this thesis, we make an orientation analysis on motion. However, before we proceed to the orientation concept in three-dimensional signals, we find it beneficial to mention the orientation in the two-dimensional case, in other words, in *still* images.

Many images, especially ones which are captured from the nature, have many oriented structures. The orientation of a feature is characterized by the direction of the edges of that feature. In Figure 2.1, we define a coordinate system for a still image and an orientation angle α.

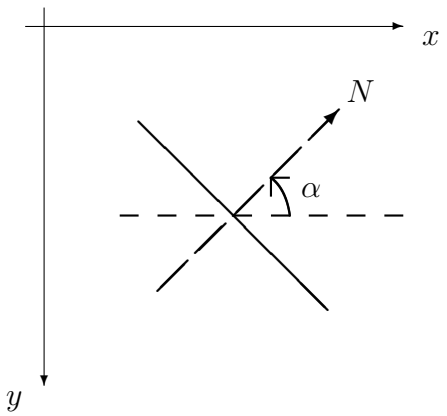


Figure 2.1: The diagonal solid line shows an edge of a feature and the dashed arrow indicated with N shows the line which is normal to the edge. The orientation of the edge is defined by the angle α. If $\alpha = 0$ rad, the edge is called vertically oriented and if $\alpha = \pi/2$ rad, the edge is called horizontally oriented.

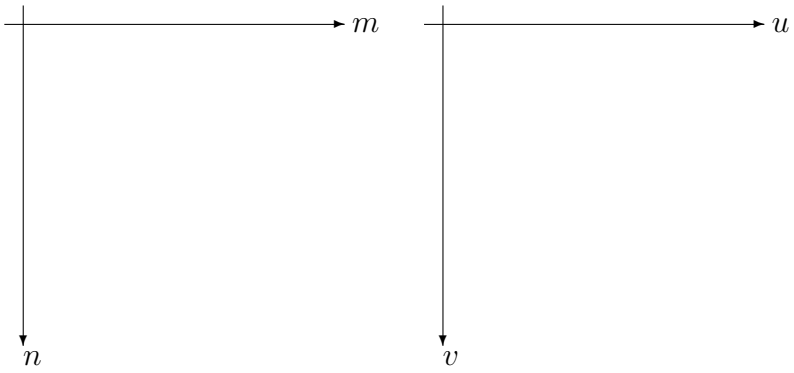
In this thesis, we work on discrete images. Let a finite image be defined by the function $I(m, n)$, where m and n are integers such that $m \in [0, M-1]$, $n \in [0, N-1]$. Thus m and n identify the pixel locations along the x and y axes, respectively. The two-dimensional discrete Fourier transform (DFT), $F(u, v)$, of

this image is calculated by;

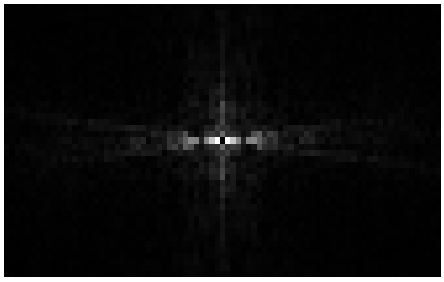
$$F(u,v) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} I(m,n) e^{-j\frac{2\pi um}{M}} e^{-j\frac{2\pi vn}{N}}, \tag{2.1}$$

where $j = \sqrt{-1}$, u and v are integers such that $u \in [0, M-1]$, $v \in [0, N-1]$ [43]. The plane that these coefficients lie is called as *Fourier plane* or *frequency plane*. Moreover, the magnitude spectrum coefficients are the absolute values of the DFT coefficients. In the coordinate system defined for the DFT, u and v correspond to discrete frequencies along the x and y axes shown in Figure 2.1, respectively.

The DFT of an image which contains oriented structures shows a characteristic pattern. That is, in the Fourier plane, existence of a nonzero coefficient in an arbitrary (u, v) point implies existence of a feature which have the orientation angle $\alpha = \arctan(-v/u)$ in the image. Also, the magnitude of that coefficient determines the dominancy of the oriented feature and the phase determines its location in the image [43]. In Figure 2.2, three images which have oriented structures and their magnitude spectrum are shown as examples. The first image is the jersey of the football team *Barcelona* and the second image is a zebra, more complicated image since it has features in multiple orientations. The jersey has vertically oriented stripes along y axis ($\alpha = 0$ rad) and the stripes of the zebra elongates mainly in the diagonal direction. The last image is an artifical image which also has oriented structures but it contain higher frequency with respect to other images. In accordance with the previous statement, the large coefficients in the magnitude spectrum of the images lie along the direction of dominant spatial orientation of the images.

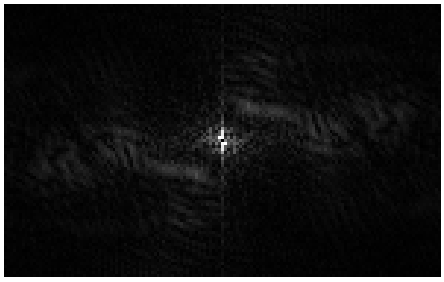


(a) A gray-scale image of the jersey of Barcelona

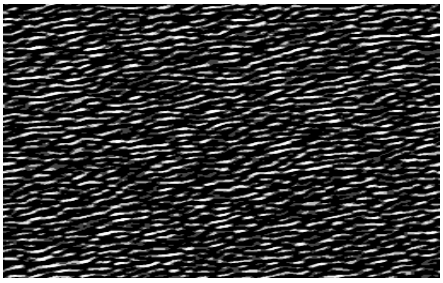


(b) The magnitude spectrum of the image of the jersey

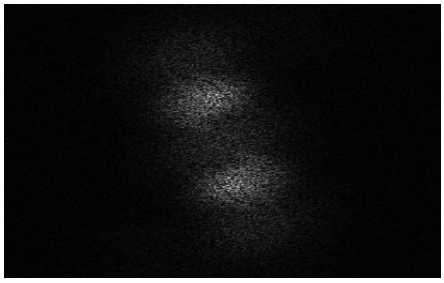
(c) A gray-scale image of a zebra



(d) The magnitude spectrum of the image of the zebra



(e) An artifical image



(f) The magnitude spectrum of the artificial image.

Figure 2.2: Three images which contain oriented structures together with the defined coordinate axes are shown. The origin of the magnitude spectrums are brought to the origin of the images.

Therefore, a filter may detect a feature which has a certain orientation, if its magnitude response has large coefficients along the same orientation angle of the feature. These type of filters, which have orientation selectivity, are called *oriented filters*. Readers can find detailed explanations about orientation selectivity and oriented edge detection in [43] and [44].

2.1.2 Orientation in 3-D Images

The orientation concept can be extended to the spatio-temporal domain. In a still image, a variaton on the signal occurs as a result of the intensity change in nearby pixels. Since a video can be assumed as a rectangular prism whose volume is filled with rows, columns and frames, a variation in time corresponds to the intensity change of nearby pixels in successive frames.

Three-dimensional DFT of a video, $I(m, n, t)$, is computed by just adding the new frequency variable p to the two-dimensional DFT. That is;

$$F(u, v, p) = \sum_{t=0}^{T-1} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} I(m, n, t) e^{-j\frac{2\pi um}{M}} e^{-j\frac{2\pi vn}{N}} e^{-j\frac{2\pi pt}{T}}, \tag{2.2}$$

where t is an integer in $[0, T-1]$. Inverse discrete Fourier transform (IDFT) is also defined by;

$$I(m, n, t) = \frac{1}{MNT} \sum_{p=0}^{T-1} \sum_{v=0}^{N-1} \sum_{u=0}^{M-1} F(u, v, p) e^{j\frac{2\pi um}{M}} e^{j\frac{2\pi vn}{N}} e^{j\frac{2\pi pt}{T}}, \tag{2.3}$$

where p is an integer in $[0, T-1]$.

Equation 2.3 states that a video in the discrete domain can be written as a linear combination of three-dimensional complex exponential functions. Since we work on real signals, complex exponential functions reduce to discrete cosines, which are in the form of $g(m, n, t) = cos(um + vn + pt + \varphi)$, where φ represents the phase. The perceived motion direction of this cosine is always the same

as the spatial orientation, which is computed as $\arctan(-v/u)$, due to *aperture problem* [45]. The cosine function $cos(um + vn + pt)$ can be written in the form of $g(m + V_x t, n + V_y t) = cos\big[u(m + V_x t) + v(n + V_y t)\big]$, where V_x and V_y are the components of the velocity vector $\vec{V} = [V_x \; V_y]^T$ along x and y directions. For a three-dimensional cosine signal, these components are $V_x = up/(u^2 + v^2)$ and $V_y = vp/(u^2 + v^2)$. The speed is also calculated by $|\vec{V}| = \sqrt{V_x^2 + V_y^2} = |p|/\sqrt{u^2 + v^2}$. Therefore, in the three-dimensional magnitude plane, existence of a nonzero coefficient in an arbitrary (u, v, p) point implies the presence of a moving feature whose spatial domain orientation and speed are $\arctan(-v/u)$ and $|p|/\sqrt{u^2 + v^2}$, respectively. Three frequency component, u, v and p constitute an orientation vector in the three-dimensional space, which is $[u \; v \; p]^T$.

As an example, in Figure 2.3, the large coefficients of the DFT of a moving vertically oriented edge is shown. As it can be seen from the figure, according to the $\arctan(-v/u)$ formula, spatial domain orientation angle of each cosine is zero. In addition, their speed, which is $|p|/\sqrt{u^2 + v^2}$ is equal to the speed of the moving edge.

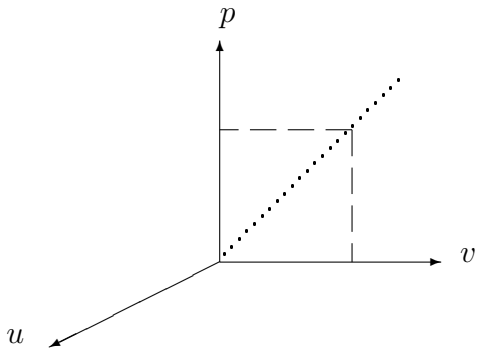


Figure 2.3: The DFT of a motion of a vertical edge is shown. The large nonzero coefficients are found where the dots are located. If video of the moving edge consists of real pixel values, the DFT is symmetric with respect to origin.

Therefore, in order to extract the motion characteristics of a feature, spatio-temporally oriented filters can be used. Further explanations about the spatio-temporal orientation can be found in [46] and [47].

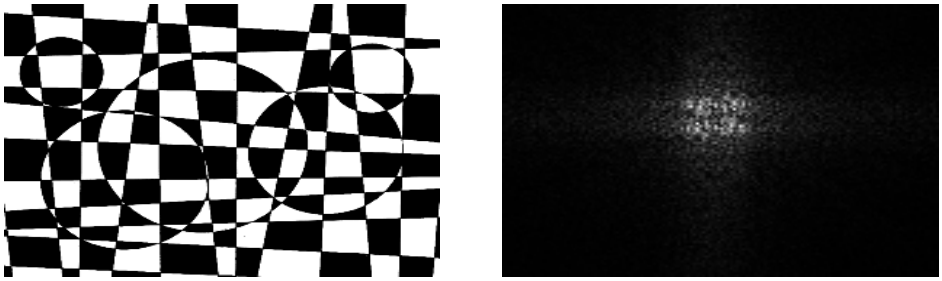
In this thesis, since shiny features have their own motion characteristics, orientation analysis in the spatio-temporal domain would be helpful to extract information about shininess. In addition, as we mentioned in the previous chapter, in the early stages of the HVS, orientation selective cells are found [32]. Therefore, oriented filters are also useful to model the HVS.

2.1.3 Scale

In images and videos, correlation of pixel intensities both in small and large regions can give information about the image features. The term *scale* in image processing is used to point out the size of the region that is being examined. Coarser scale is associated with large regions and finer scale is associated with small regions. In the three-dimensional images, $I(m, n, t)$, analysis in the different scales extracts information about motion continuity and duration of different sized features.

In the Fourier spectrum, the *scale* concept can be interpreted as follows. If there is a variation at a finer scale, that means there is a fast change on that local region of the signal. A fast change in a signal occurs if the coefficients of high frequency components in the Fourier spectrum are large. On the contrary, a slow change in a signal can be noticed in coarser scales. And a slow change in a signal occurs as a result of the dominant low frequency cosine components. Therefore, in order to extract features on the finer scales, filters tuned to high frequencies and in order to extract features on the coarse scales, filters tuned to low frequencies should be used. An example showing the scale concept is given

in Figure 2.4. In this figure, variations in both coarse and fine scales can be seen. The magnitude spectrum of this image is also shown.



(a) An image which contain features at different scales.

(b) The magnitude spectrum of the image.

Figure 2.4: First image has circles at coarser scales and rectangle-like shapes at finer scales. Second image shows the magnitude spectrum of the first image.

Readers can find detailed explanations about the *scale* concept in [48].

In this thesis, the motion of both small and large shiny features relative to object surface are required to be analyzed. (This would give information about coherent motion duration of shiny and matte features). Therefore, scale analysis might provide information about shininess.

2.1.4 Filter Design for Selectivity in Specific Orientation and Scale

Before we proceed with the design steps of our filters, in this section, we first mention a general technique that can be applied for the design of filters which have arbitrary scale and orientation selective characteristics.

In this thesis, we start to design the filters by specifying their discrete time Fourier transforms (DTFT). Let a three-dimensional discrete filter be the function $f(x, y, t)$, where $x,\ y,\ t$ are integers such that $x,\ y,\ t \in (-\infty, \infty)$. The

three-dimensional DTFT, $F(\omega_x, \omega_y, \omega_t)$, of this filter is calculated by;

$$F(\omega_x, \omega_y, \omega_t) = \sum_{t=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} \sum_{x=-\infty}^{\infty} f(x, y, t) e^{-j(\omega_x x + \omega_y y + \omega_t t)}, \tag{2.4}$$

where $\omega_x, \omega_y, \omega_t$ are continuous real numbers such that $\omega_x,\ \omega_y,\ \omega_t \in (-\infty, \infty)$.

In order to design the orientation and scale selective filters, it is good to use the spherical coordinates for three-dimensional signals. This coordinate system allows direct identification of the orientation and scale selectivities of the filters. In the frequency domain and according to Figure 2.5, the spherical coordinates are defined as

$$\omega_r = \sqrt{\omega_x^2 + \omega_y^2 + \omega_t^2} \tag{2.5}$$

$$cos(\phi) = \frac{\omega_t}{\omega_r},\ \text{where } 0 \leqslant \phi \leqslant \pi \tag{2.6}$$

$$cos(\theta) = \frac{\omega_x}{\sqrt{\omega_x^2 + \omega_y^2}},\ \text{where } 0 \leqslant \theta < 2\pi \tag{2.7}$$

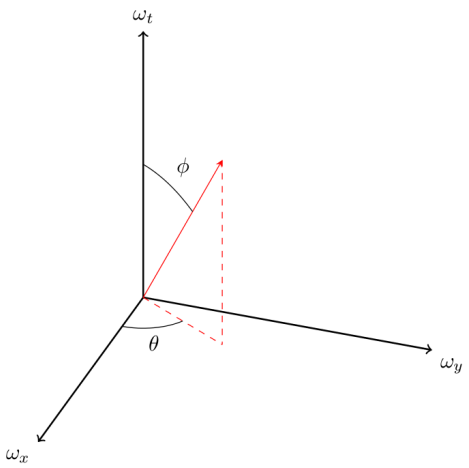


Figure 2.5: Spherical coordinate angles

The radial and the angular part of the filter can be designed separately. Let the function $F_s(\omega_r, \theta, \phi)$ be the representation of the filter $F(\omega_x, \omega_y, \omega_t)$ in the spherical coordinates. Let us assume that the filter has the property,

$$F_s(\omega_r, \theta, \phi) = W(\omega_r)G(\theta, \phi), \tag{2.8}$$

where, $W(\omega_r)$ is the radial part and $G(\theta, \phi)$ is the angular part of the filter.

This kind of separation allows separate treatment of radial and angular parts in terms of center frequency and bandwidth. The term *center frequency* refers to the frequency that the amplitude of the filter is at its maximum and *bandwidth* is the term used for the quantity of the frequency interval of the passband of the filter. There are different bandwidth definitions in the literature; in this thesis we use 3-dB bandwidth definiton [49]. Some important notes on the filter design are the following;

- These filters are assumed to be zero-phase filters. Therefore, specifying the magnitude spectrum is enough to design the filters.

- It is crucial to remember that the DTFT is periodic with 2π, therefore we specify the DTFT coefficients in the interval $[-\pi, \pi)$ and assume that DTFT has periodic extensions. Therefore, according to Equation 2.5, the maximum value of ω_r can be $\pi\sqrt{3}$. If filters which have same radial but different angular functions are to be designed, in order to keep the characteristics of the specified radial function, $W(\omega_r)$ should be zero when ω_r is greater than π. The reason is that, if $W(\omega_r)$ is designed such a way that it is nonzero when ω_r is greater than π, multiplication of $W(\omega_r)$ by a certain angular function, $G(\theta, \phi)$, may lead to nonzero coefficients outside the interval of $[-\pi, \pi)$ at ω_x, ω_y or ω_t axes. This situation disobeys 2π periodicity rule and it becomes impossible to design the filters with the specified radial function.

- If the spatio-temporal domain filter coefficients are desired to be real zero-phase filters, then the frequency domain coefficients should be chosen to be symmetric with respect to the origin.

- After specifying the filter characteristics in the DTFT domain, they should be converted to the discrete spatio-temporal domain for digital processing. In the literature, there are different techniques for this conversion. Some methods are explained in detail in [50]. In this thesis, we give details of our adopted technique in Section 2.3.3.

2.2 Steerable Filters

Various kinds of filters may be needed for different applications. For example, for smoothing, low-pass filters; for edge detection, high-pass or band-pass filters are used. In each of those three types of filtering operations, as in our case, orientation selective filters may be needed. When an image/video is analyzed in different orientations, after specifying the center frequencies of the angular parts of the filters, designing all of them in the discrete spatial/spatio-temporal domain separately can cause a huge workload. Each filter also occupies memory on its operation platform (a computer, an embedded system etc). Moreover, for an input signal, the output of each of those oriented filters are needed to be computed separately either with the convolutions in the spatial/spatio-temporal or with the multiplications in the Fourier domain. Those large number of convolutions or multiplications also require higher memory sizes and more advanced processors.

Steerable filter concept, on the other hand, introduces the fact that an arbitrarily oriented filter output of a multi-dimensional signal can be found by linear combination of a number of basis filter outputs. Those basis filters are the same

except the center frequencies of their angular functions. In other words, they are rotated copies of a prototype filter.

In the three-dimensional space, let $I(\omega_x, \omega_y, \omega_t)$ be the three-dimensional DTFT of a video and $F_{\theta,\phi}(\omega_x, \omega_y, \omega_t)$ be the three-dimensional DTFT of a filter whose center frequency in the orientation space is tuned to the spherical angles θ and ϕ, as shown in Figure 2.5.

The *steerability property* can be written as,

$$I(\omega_x, \omega_y, \omega_t) F_{\theta_c,\phi_c}(\omega_x, \omega_y, \omega_t) = \sum_{i=1}^{N} k_i(\theta_c, \phi_c) I(\omega_x, \omega_y, \omega_t) F_{\theta_i,\phi_i}(\omega_x, \omega_y, \omega_t). \quad (2.9)$$

As we mentioned in the introduction of this chapter, in [40], there are theorems about computing minimum number and orientations of basis filters for two and three dimensional filters. In [41], a method of three dimensional separable steerable filter design is presented. In those articles, theorems about the steerability property are presented for functions in the form of $F_{\theta,\phi}(\omega_x, \omega_y, \omega_t) = W(\omega_r) P_N(\omega_{x'})$ where $W(\omega_r)$ is a spherically symmetric windowing function, that is $\omega_r = \sqrt{\omega_x^2 + \omega_y^2 + \omega_t^2}$, and $P_N(\omega_{x'})$ is an N^{th} order polynomial in $\omega_{x'} = \alpha\omega_x + \beta\omega_y + \gamma\omega_t$. α, β and γ are defined as the directional cosines, which are the functions of θ and ϕ. Another approach for steerable filters is the following. Let $\omega_{x'}$, $\omega_{y'}$, $\omega_{t'}$ axes be the rotated versions of standard ω_x, ω_y, ω_t axes. In the cartesian coordinates, rotation is represented by (according to the angles specified in Figure 2.5) [51],

$$\begin{bmatrix} \sin(\phi_c) & 0 & \cos(\phi_c) \\ 0 & 1 & 0 \\ -\cos(\phi_c) & 0 & \sin(\phi_c) \end{bmatrix} \left(\begin{bmatrix} \cos(\theta_c) & -\sin(\theta_c) & 0 \\ \sin(\theta_c) & \cos(\theta_c) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_t \end{bmatrix} \right) = \begin{bmatrix} \omega_{x'} \\ \omega_{y'} \\ \omega_{t'} \end{bmatrix}. \quad (2.10)$$

In the above equation, the product inside the parantheses first rotates the coordinate system around ω_t axis and the second product rotates around $\omega_{y'}$

axis. The confusion about the signs of the matrix elements can be handled by assuming that the rotation of a plane is counterclockwise when looked topview from the positive side of the rotation axes. From Equation 2.10, by writing $\omega_{x'}$ as,

$$\omega_{x'} = \omega_x \sin(\phi_c)\cos(\theta_c) + \omega_y \sin(\phi)\sin(\theta_c) + \omega_t \cos(\phi_c), \tag{2.11}$$

basis filters and their corresponding coefficients in Equation 2.9 can be found.

In this thesis, we use this rotation concept for steerable filters in the design of the derivative of Gaussian filters. We give the details in the next section.

2.3 Derivative of Gaussian Filters

In this thesis, since we need oriented band-pass filters for extracting subband characteristics of shiny and matte object motions, we decided to use derivatives of Gaussian function. The motivation behind this choice is their separability, steerability property and smoothness (smoothness here is used as a term to indicate the smooth transition from zero to its maximum value in the magnitude spectrum). Separability of a filter reduces the computational complexity by converting multi-dimensional convolutions to convolutions in smaller dimensions. A linear time invariant filter is separable if its impulse response, $f(x, y, t)$, can be written as;

$$f(x, y, t) = g(x)h(y)k(t).$$

The Fourier transform of a separable function is also a separable function. Steerability, as explained in the above section, provides computational efficiency and smoothness provides short filter length in the spatio-temporal domain without aliasing. Moreover, as we explain in the next section, usability of 1^{st} derivative of Gaussian filter in the steerable pyramid is also important.

In the discrete time Fourier domain, n^{th} derivative of Gaussian filter with respect to x' in the rotated coordinate axes x', y', t' can be written as a multiplication of the DTFT of the Gaussian filter and $(j\omega_{x'})^n$ [52]. That is,

$$G^{(n)}(\omega_{x'}, \omega_{y'}, \omega_{t'}) = C\,(j\omega_{x'})^n\, e^{-\frac{\sigma^2}{2}\left(\omega_{x'}^2+\omega_{y'}^2+\omega_{t'}^2\right)}, \tag{2.12}$$

where σ is a parameter that influences the bandwidth and the center frequency of the radial frequency and $C = \sigma^n \left(\frac{e}{n}\right)^{\frac{n}{2}}$ is a normalization constant such that the maximum value of $G^{(n)}(\omega_{x'}, \omega_{y'}, \omega_{z'})$ is equal to one.

2.3.1 Steerability

As in Equation 2.11, $\omega_{x'}$ can be written in terms of ω_x, ω_y, ω_t. Since the term inside the exponent in Equation 2.12 is spherically symmetric, it can be directly replaced by $\omega_x^2+\omega_y^2+\omega_t^2$. Therefore, an oriented n^{th} derivative of Gaussian filter along the angles specified by the spherical coordinates θ_c and ϕ_c is, [41],

$$\begin{aligned} G^{(n)}_{\theta_c,\phi_c}(\omega_x, \omega_y, \omega_t) =& C\jmath^n e^{-\frac{\sigma^2}{2}\left(\omega_x^2+\omega_y^2+\omega_t^2\right)} \\ & \Big[\cos(\theta_c)\sin(\phi_c)\omega_x + \sin(\theta_c)\sin(\phi_c)\omega_y + \cos(\phi_c)\omega_t\Big]^n \end{aligned} \tag{2.13}$$

Equation 2.13 can be written as a sum of n^{th} order polynomials times an exponential function. That is;

$$G^{(n)}_{\theta_c,\phi_c}(\omega_x, \omega_y, \omega_t) = \sum_{k=0}^{n}\sum_{l=0}^{k} \overbrace{C\binom{n}{k}\binom{k}{l}\left(\cos(\theta_c)^l \sin(\theta_c)^{k-l} \sin(\phi_c)^k \cos(\phi_c)^{n-k}\right)}^{\text{interpolation coefficients}} \underbrace{\left(j^n \omega_x^l \omega_y^{k-l} \omega_t^{n-k} e^{-\frac{\sigma^2}{2}\left(\omega_x^2+\omega_y^2+\omega_t^2\right)}\right)}_{\text{basis filters}}. \tag{2.14}$$

The interpolation coefficients and the basis filters in Equation 2.14 are, thus, found as in Equation 2.9 for derivatives of Gaussian filters [41]. The basis functions are in the frequency domain and they should be transformed to the spatio-temporal domain to be used in the convolutions. The basis functions are all

separable functions and therefore, three-dimensional convolutions can be accomplished with one-dimensional convolutions.

2.3.2 Orientation and Scale Characteristics

Although specifying the derivative of Gaussian filters in the cartesian coordinates reveals the separability and steerability properties, the orientation and scale characteristics can be computed if the filters are written in the spherical coordinates. The filters can be written in the spherical coordinates as;

$$G^{(n)}_{\theta_c,\phi_c}(\omega_r,\alpha) = Cj^n e^{-\frac{\sigma^2}{2}{\omega_r}^2} \left(\omega_r \cos_{\theta_c,\phi_c}(\alpha)\right)^n, \tag{2.15}$$

where $\cos_{\theta_c,\phi_c}(\alpha)$ is equal to the inner product of the unit vectors whose elements are the cartesian coordinates, ω_x, ω_y and ω_t, of the filter whose orientation angle is towards (θ_c, ϕ_c) and the cartesian coordinates of an arbitrarily oriented signal towards (θ, ϕ) [53]. In addition, since the function in Equation 2.15 is given in the spherical coordinates, we use the spherical coordinate representations. That is,

$$\begin{aligned}
\cos_{\theta_c,\phi_c}(\alpha) &= < v_{\theta_c,\phi_c}, v_{\theta,\phi} > \\
&= \begin{bmatrix} \omega_{x\theta,\phi} & \omega_{y\theta,\phi} & \omega_{z\theta,\phi} \end{bmatrix} \begin{bmatrix} \omega_{x\theta_c,\phi_c} & \omega_{y\theta_c,\phi_c} & \omega_{z\theta_c,\phi_c} \end{bmatrix}^T \\
&= \begin{bmatrix} \cos(\theta)\sin(\phi) & \sin(\theta)\sin(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} \cos(\theta_c)\sin(\phi_c) \\ \sin(\theta_c)\sin(\phi_c) \\ \cos(\phi_c) \end{bmatrix} \\
&= \sin(\phi_c)\sin(\phi)\cos(\theta_c)\cos(\theta) + \sin(\phi_c)\sin(\phi)\sin(\theta_c)\sin(\theta) + \cos(\phi_c)\cos(\phi)
\end{aligned} \tag{2.16}$$

The equality between Equation 2.13 and Equation 2.15 can be seen through substituting $\cos_{\theta_c,\phi_c}(\alpha)$ term in Equation 2.15 by the result obtained in Equation

2.16. That is,

$$G^{(n)}_{\theta_c,\phi_c}(\omega_r,\theta,\phi) = Cj^n e^{-\frac{\sigma^2}{2}{\omega_r}^2}$$
$$\left[\omega_r \sin(\phi_c)\sin(\phi)\cos(\theta_c)\cos(\theta)+\omega_r \sin(\phi_c)\sin(\phi)\sin(\theta_c)\sin(\theta)+\omega_r \cos(\phi_c)\cos(\phi)\right]^n \tag{2.17}$$

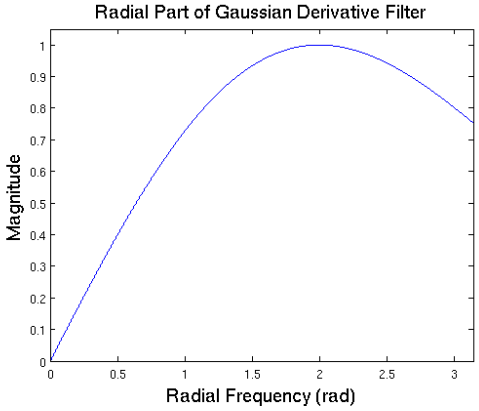
and if the equalities

$$\omega_x = \omega_r \sin(\phi)\cos(\theta), \quad \omega_y = \omega_r \sin(\phi)\sin(\theta), \text{ and } \omega_t = \omega_r \cos(\phi)$$

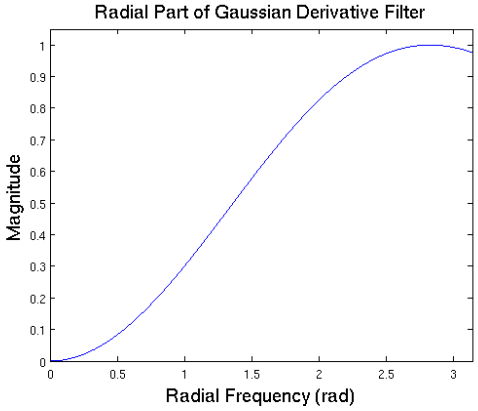
are replaced in Equation 2.17, the equality between Equation 2.13 and Equation 2.15 is provided.

First of all, the DTFT of a signal should be rectangularly periodic with 2π [43]. However, it can be seen from Equation 2.15 that, the derivatives of Gaussian filters have infinite support without periodicity. In order to solve this problem in the design of the discrete derivative of the Gaussian filters, we need to compute $e^{-\frac{\sigma^2}{2}{\omega_r}^2}\omega_r^n$ for $\omega_r \in (0\ \pi]$ and then, assume that the filters have rectangularly periodic extensions in the Fourier domain.

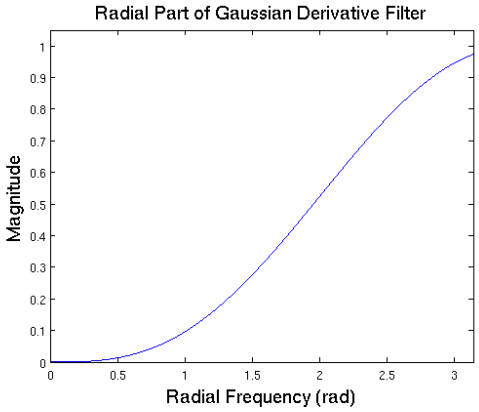
It is also worth to note that, in order not to lose the band-pass characteristics of the filters, their magnitudes should converge to zero around the radial frequency π. As an example, in Figure 2.6, the radial parts of Gaussian derivative filters for different derivative orders, n, and the parameter σ is shown (That is $Ce^{-\frac{\sigma^2}{2}{\omega_r}^2}\omega_r^n$). It can be understood from these figures that if σ is 0.5 or lower, the filters do not show band-pass characteristics whatever the order of the derivative is. If σ equals to 1, 1^{st} and 2^{nd} order derivatives can be used, on the other hand, 3^{rd} order derivative may not be appropriate depending on the application type. If σ equals to 1.5, 1^{st}, 2^{nd} and 3^{rd} order derivative of Gaussian filters are appropriate to use.



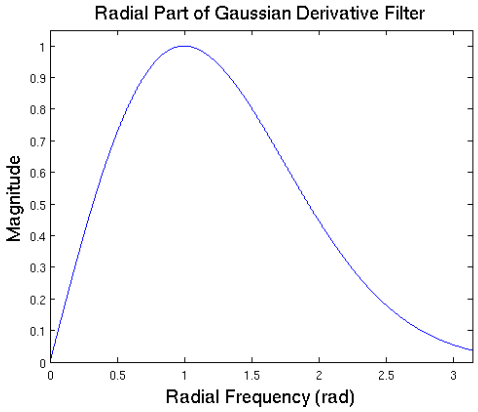
(a) n=1, $\sigma = 0.5$



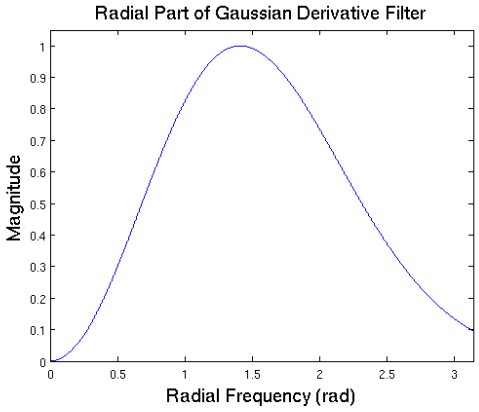
(b) n=2, $\sigma = 0.5$



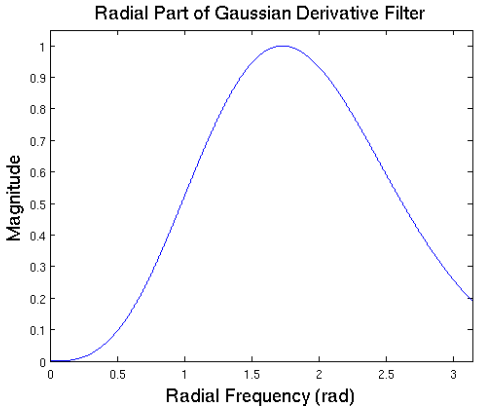
(c) n=3, $\sigma = 0.5$



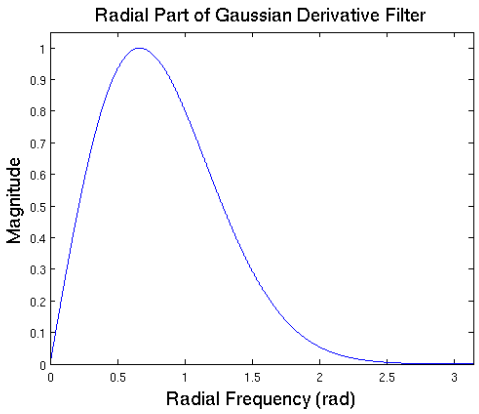
(d) n=1, $\sigma = 1$



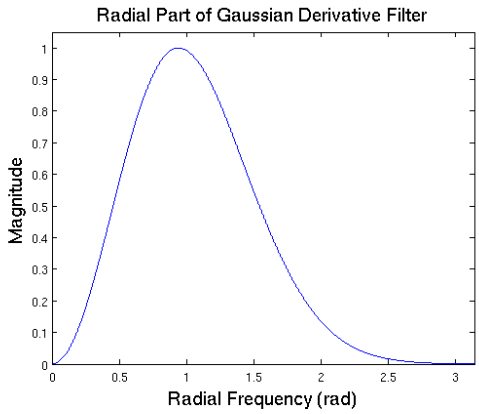
(e) n=2, $\sigma = 1$



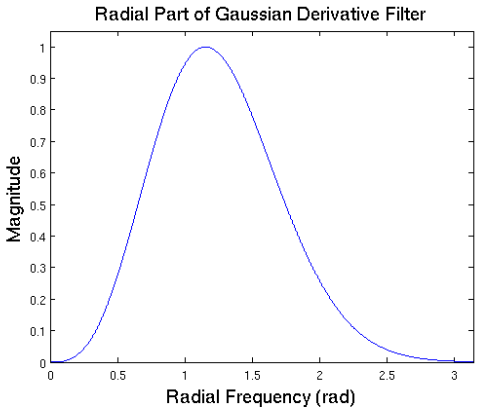
(f) n=3, $\sigma = 1$



(g) n=1, $\sigma = 1.5$



(h) n=2, $\sigma = 1.5$



(i) n=3, $\sigma = 1.5$

Figure 2.6: The radial functions of the derivatives of Gaussian filters for different derivative orders, n, and σ.

The scale selectivity of the derivatives of Gaussian filters is determined by the function $Ce^{-\frac{\sigma^2}{2}{\omega_r}^2}\omega_r^n$. The center frequency, ω_r^c, is determined by,

$$\omega_r^c = \pm\frac{\sqrt{n}}{\sigma}.$$

This equality can be reached by taking the first derivative of $e^{-\frac{\sigma^2}{2}{\omega_r}^2}\omega_r^n$ with respect to ω_r and then equating it to zero.

As we stated before, we adopted the 3-dB bandwidth definition. That is the quantity of the interval between the cutoff frequencies at which the magnitude of the filter reduces up to $1/\sqrt{2}$ of its maximum. The 3 dB bandwidth of the

filter is,

$$\omega_r^{BW} = |\omega_r^{cutoff1} - \omega_r^{cutoff2}|, \tag{2.18}$$

where $\omega_r^{cutoff1}$ and $\omega_r^{cutoff2}$ are the solutions to

$$e^{-\frac{\sigma^2}{2}{\omega_r}^2}\omega_r^n = \left(\frac{n}{e}\right)^{\frac{n}{2}} \frac{1}{\sqrt{2}\sigma^n}. \tag{2.19}$$

Since we could not find an analytical solution to this equation, we find the bandwidth of the filters by numerical methods. As it can be seen from the above equations, the center frequency and the bandwidth is determined by the order of the derivative and the parameter σ.

The orientation selectivity of a derivative of Gaussian filter is determined by the angular function $\left[\cos_{\theta_c,\phi_c}(\alpha)\right]^n$. The center frequencies in the orientation space are θ_c and ϕ_c. For example, a filter which is oriented towards the angles $\theta_c = 0$ rad and $\phi_c = \pi/2$ rad (it is actually ω_x axis), has no response to the signals whose nonzero coefficients lie only along ω_y or ω_z axes. Since the angle between those axes and ω_x axis is $\pi/2$. The 3-dB bandwidth, α_{BW}, of the angular part of an n^{th} derivative of Gaussian filter is,

$$\alpha_{BW} = \arccos(2^{-\frac{1}{2n}}). \tag{2.20}$$

This bandwidth is calculated by finding the angle that the gain of the filter decreases to $1/\sqrt{2}$ of its value at that angle.

2.3.3 Spatio-temporal Domain Discrete Filter Design

In order to use the filters in discrete applications and in the spatio-temporal domain, we used *frequency sampling* algorithm to design the filters [54]. Here we give the filter design steps for one dimensional filters. The reason is that the

derivative of Gaussian functions are separable and the filters can be operated in one-dimensional convolutions.

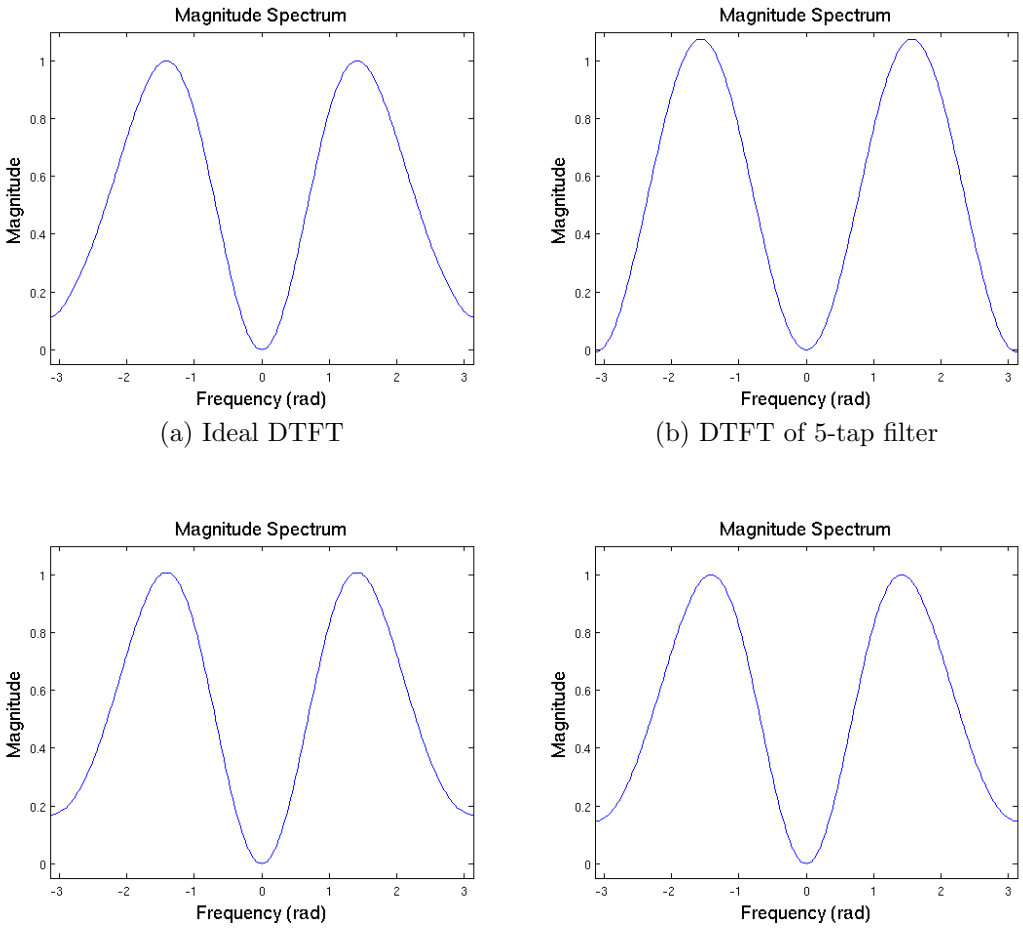
Let $F(\omega)$ be the DTFT of a one-dimensional discrete domain filter, where $\omega \in (-\infty, \infty)$. The spatio temporal domain discrete filter coefficients, $f(t)$, can be computed according to the equation,

$$f(t) = \frac{1}{M} \sum_{k=-(\lfloor \frac{M-1}{2} \rfloor)}^{\lceil \frac{M-1}{2} \rceil} F(k\frac{2\pi}{M}) e^{j\frac{2\pi kt}{M}}, \tag{2.21}$$

where, M is a positive integer, k is an integer such that $-(\lfloor \frac{M-1}{2} \rfloor) \leqslant k \leqslant (\lceil \frac{M-1}{2} \rceil)$ and t is an integer such that $-(\lfloor \frac{M-1}{2} \rfloor) \leqslant t \leqslant (\lceil \frac{M-1}{2} \rceil)$.

The number of the coefficients in the filter is determined by the number M. It is important to make a smart choice for this number. First of all, for the derivatives of Gaussian filters, M should be an odd number to have real coefficients and to have zero phase filters for even numbered derivative order. Second, we know that while the bandwidth of a signal gets narrower, interval of the nonzero coefficients in the time domain gets larger ([33]). Therefore, the number of the samples calculated from the magnitude spectrum should be large enough. In order to decide for an appropriate value, we first design a filter and we recompute DTFT from those coefficients. Then we compare the ideal DTFT and the recomputed DTFT by looking at their plots. (In order to compute an approximate DTFT by a computer, we pad the filter coefficients with a large number of zeros, and then compute the DFT from such zero padded filter coefficients.) In Figure 2.7, we give samples of the DTFT of the filters which are computed by taking different number of samples from 2^{nd} derivative of the Gaussian filter. From the figures, it can be seen that, if one needs exactly the same DTFT as the ideal one, 5-tap filter is not appropriate, because the maximum of its DTFT coefficients is greater than one and its bandwidth is narrower than the ideal one. Moreover, although 7-tap and 9-tap filters look

quite similar, if they are zoomed, it can be seen that the bandwidth of 7-tap filter also differentiates from that of the ideal one.



(a) Ideal DTFT

(b) DTFT of 5-tap filter

(c) DTFT of 7-tap filter

(d) DTFT of 9-tap filter

Figure 2.7: The magnitude spectrum of 2^{nd} derivative of Gaussian filters for different length filters.

2.3.4 Explicit Design Steps of 2^{nd} Derivative of Gaussian Filter

As it can be seen from Equation 2.12, 2^{nd} derivative of Gaussian filter is a zero phase filter (actually phase of the filter is π rad because $j^2 = -1$, however, a multiplication by -1 makes it zero phase). Its radial center frequency and bandwidth are $\sqrt{2}/\sigma$ rad and 0.31π rad, respectively. The angular frequency bandwidth is 0.18π rad and the center frequency can be decided for a particular

orientation selection. This filter is narrow enough in terms of both orientation and scale selectivity for our purpose. Therefore, we used it as the band-pass filters for the analysis of the movies. A filter response of a video, $I(\omega_x, \omega_y, \omega_t)$, along the center frequencies for the orientation, (θ_c, ϕ_c), can be computed as,

$$
\begin{aligned}
I(\omega_x,\omega_y,\omega_t)\big[- G^{(2)}_{\theta_c,\phi_c}(\omega_x,\omega_y,\omega_t)\big] = \\
& Ck_1(\theta_c,\phi_c)\big[I(\omega_x,\omega_y,\omega_t)\omega_x^2 e^{-\frac{\sigma^2}{2}\omega_x^2} e^{-\frac{\sigma^2}{2}\omega_y^2} e^{-\frac{\sigma^2}{2}\omega_t^2}\big] \\
+ & Ck_2(\theta_c,\phi_c)\big[I(\omega_x,\omega_y,\omega_t)\omega_y^2 e^{-\frac{\sigma^2}{2}\omega_y^2} e^{-\frac{\sigma^2}{2}\omega_x^2} e^{-\frac{\sigma^2}{2}\omega_t^2}\big] \\
+ & Ck_3(\theta_c,\phi_c)\big[I(\omega_x,\omega_y,\omega_t)\omega_t^2 e^{-\frac{\sigma^2}{2}\omega_t^2} e^{-\frac{\sigma^2}{2}\omega_x^2} e^{-\frac{\sigma^2}{2}\omega_y^2}\big] \\
- & Ck_4(\theta_c,\phi_c)\big[I(\omega_x,\omega_y,\omega_t)\cdot j\omega_x e^{-\frac{\sigma^2}{2}\omega_x^2} j\omega_y e^{-\frac{\sigma^2}{2}\omega_y^2} e^{-\frac{\sigma^2}{2}\omega_t^2}\big] \\
- & Ck_5(\theta_c,\phi_c)\big[I(\omega_x,\omega_y,\omega_t) j\omega_x e^{-\frac{\sigma^2}{2}\omega_x^2} j\omega_t e^{-\frac{\sigma^2}{2}\omega_t^2} e^{-\frac{\sigma^2}{2}\omega_y^2}\big] \\
- & Ck_6(\theta_c,\phi_c)\big[I(\omega_x,\omega_y,\omega_t) j\omega_y e^{-\frac{\sigma^2}{2}\omega_y^2} j\omega_t e^{-\frac{\sigma^2}{2}\omega_t^2} e^{-\frac{\sigma^2}{2}\omega_x^2}\big] \qquad (2.22)
\end{aligned}
$$

where,

$$
\begin{aligned}
k_1(\theta_c,\phi_c) &= \sin^2(\phi_c)\cos^2(\theta_c), & k_2(\theta_c,\phi_c) &= \sin^2(\phi_c)\sin^2(\theta_c), \\
k_3(\theta_c,\phi_c) &= \cos^2(\phi_c), & k_4(\theta_c,\phi_c) &= 2\sin^2(\phi_c)\cos(\theta_c)\sin(\theta_c), \\
k_5(\theta_c,\phi_c) &= 2\sin(\phi_c)\cos(\phi_c)\cos(\theta_c), & k_6(\theta_c,\phi_c) &= 2\sin(\phi_c)\cos(\phi_c)\sin(\theta_c).
\end{aligned}
$$

Equation 2.22 is reached from Equation 2.14. As it can be seen from Equation 2.22, there are three different filters which are in the form of $\omega^2 e^{-\frac{\sigma^2}{2}\omega^2}$, $j\omega e^{-\frac{\sigma^2}{2}\omega^2}$ and $e^{-\frac{\sigma^2}{2}\omega^2}$. The term $\omega e^{-\frac{\sigma^2}{2}\omega^2}$ is multiplied by j to have real time and space domain coefficients. We designed these filters for three different σ values. These values are chosen such that the band-pass filters cover all the frequency plane. The selected σ values are 1, 1.8 and 3.2. The spatio-temporal domain coefficients of the filters mentioned in this section are computed as explained in section 2.3.3 and they are given in Appendix B.

We designed filters whose frequency responses are tuned to eight different orientation angles for all of the σ values. These angle pairs are given in Table 2.1.

	1	2	3	4	5	6	7	8
θ_c	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π	$\frac{5\pi}{4}$	$\frac{3\pi}{2}$	$\frac{7\pi}{4}$
ϕ_c	$\frac{\pi}{4}$	$\frac{\pi}{4}$	$\frac{\pi}{4}$	$\frac{\pi}{4}$	$\frac{\pi}{4}$	$\frac{\pi}{4}$	$\frac{\pi}{4}$	$\frac{\pi}{4}$

Table 2.1: Selected center frequencies in the orientation space for the analysis of the videos are shown. The angles are in radians.

The number of the oriented filters and the specified angles are selected such that they cover all the orientation space. As a result, the total number of the subbands that we investigate the movies is 24 (three scales and eight orientations for each scale).

It can be understood from Equation 2.15 that, 2^{nd} derivative of Gaussian filters are symmetric with respect to the origin in terms of both radial and angular parts. Therefore, a filter that is oriented towards a certain angle in the Fourier domain is also oriented towards the symmetric angles. The center orientation frequencies of the filters are expressed in Figure 2.8 and we give examples about the orientation selectivity of some filters in Figure 2.9.

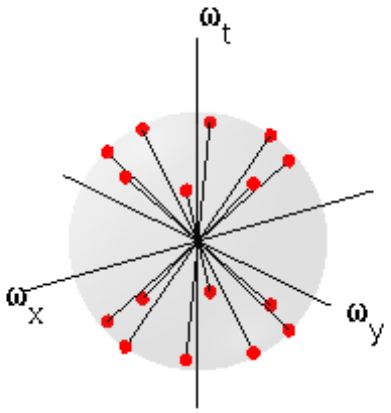


Figure 2.8: The center frequencies of the oriented filters are pointed out with the dots.

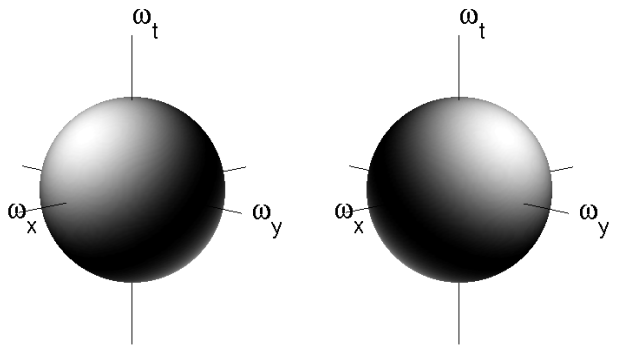


Figure 2.9: Orientation selectivities of two filters on the unit sphere are shown. The magnitude of the filter is proportional to color brightness.

2.4 Video Pyramid

In this thesis, one of the aims is to check whether the first order subband statistics are the sufficient cues to the surface reflactance recognition. For that purpose, we need to synthesize a movie which has the same first order subband statistics of the motions of shiny and matte objects. We synthesized the movies by applying the algorithm proposed in [55]. We express the details about the algorithm in the next chapter. In that algorithm, *steerable pyramid* is used as the video decomposition and reconstruction tool. Since it includes a number of filtering operations and since the filters are based on the steerable filters, we mention the steerable pyramid in this section.

Some filter outputs for different videos are demonstrated in the companion website of this thesis. The URL of the website is `http://www.umram.bilkent.edu.tr/~kulce/`.

2.4.1 Steerable Pyramid for Videos

Steerable pyramid is one of the wavelet decomposition techniques in the literature. Readers can find detailed information on other wavelet decomposition

techniques in [56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69]. The steerable pyramid basically decomposes a multidimensional signal into its subbands and then provides reconstruction from those subbands. Each subband allows analysis of a particular scale and orientation. The filterbank architecture of the steerable pyramid is shown in Figure 2.10.

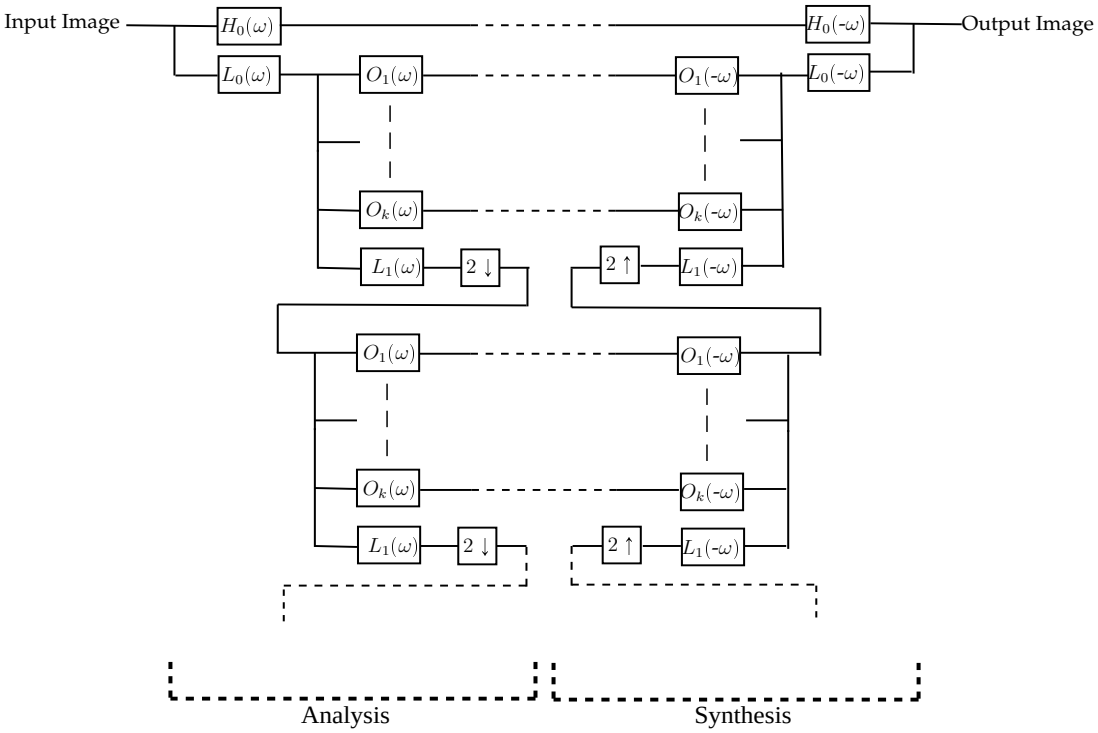


Figure 2.10: Steerable pyramid analysis/synthesis scheme is shown. $H_0(\omega)$, $L_0(\omega)$, $L_1(\omega)$, $O_i(\omega)$ are high-pass, low-pass, low-pass and oriented filters (band-pass or high-pass) respectively. The index, k, represents the number of the orientation bands. Dashed lines under downsampling and upsampling symbols represent that the pyramid decomposition/reconsruction scheme continues until a desired number of scale bands are reached.

As it can be seen from Figure 2.10, an input video is first decomposed into its high and low frequency components by applying filters $L_0(\omega)$ and $H_0(\omega)$. Then, filtering the video with the oriented filters, $O_i(\omega)$, produces oriented band-pass parts of the image. The filter, $O_i(\omega)$, can be either band pass or high pass filters, but, during the design of these filters, it should be noted that the multiplication of these filters with $L_0(\omega)$ should produce a band pass filter. Finally, a coarser scale

is obtained by applying another low pass filter $L_1(\omega)$ and then downsampling the video. The same procedure, that is filtering the video with oriented and low pass filters and downsampling continues until a desired number of scale band is reached. At the coarsest scale, only the low-pass part of the video remains.

The ideal filter characteristics are given in [42]. In Figure 2.11, for visual purposes, we give a sample frequency partition of two-dimensional coordinate system by the high pass, low pass and four oriented filters. The extension to three dimensional space is straightforward.

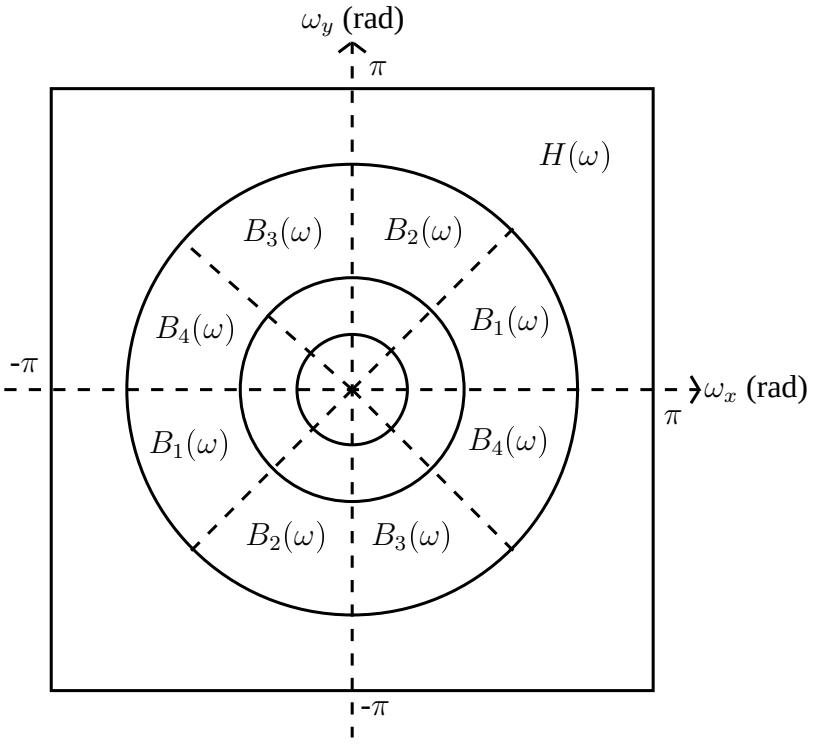


Figure 2.11: Two dimensional frequency domain partition by the steerable pyramid decomposition. $H(\omega)$ represents high-pass content and $B_i(\omega)$ represents the filter output of $O_i(\omega)$. The coarser scales are represented by the concentric circles and they are also partioned to their oriented components by $O_i(\omega)$.

The pyramid decomposition is self inverting. That is, in order to reconstruct the video, the same filters with a small modification are used. This modification is that, if the filters in the decomposition part do not have zero phase, the filters

in the reconstruction part should have reverse of the phase of the decomposition filters. This is accomplished in the reconstruction part by using the symmetrics of the decomposition filters with respect to origin. In this way, it is assured that there is no phase difference between the input video and the reconstructed video (The negative sign in Figure 2.10 next to ω symbols comes from this requirement). In addition, upsampling is applied in the reconstruction part instead of downsampling.

The main drawback of this filter bank is that it is *overcomplete*. Since the input signal is not downsampled after oriented filters, a video which has p pixels is represented in the subbands by $(\frac{8k}{7}+1)p$ pixels.

Some filter design techniques for the steerable pyramid for two-dimensional signals are given in [70] and [71]. Here we give the details on three-dimensional filter design.

2.4.2 Requirements of Three-Dimensional Filters for Steerable Pyramid

The steerable pyramid filters should satisfy three conditions for the perfect reconstruction [42];

$$Flat\ Sytem\ Response => |H_0(\omega)|^2 + |L_0(\omega)|^2 \left(|L_1(\omega)|^2 + \sum_{i=1}^{k} |O_i(\omega)|^2 \right) = 1, \tag{2.23}$$

$$Recursion => |L_1\left(\frac{\omega}{2}\right)|^2 \left(|L_1(\omega)|^2 + \sum_{i=1}^{k} |O_i(\omega)|^2 \right) = |L_1\left(\frac{\omega}{2}\right)|^2, \tag{2.24}$$

$$Anti\text{-}aliasing => |L_1(\omega)| = 0 \quad for \quad \omega > \frac{\pi}{2}. \tag{2.25}$$

While we design the filters, we assume that the oriented filters are in the form of

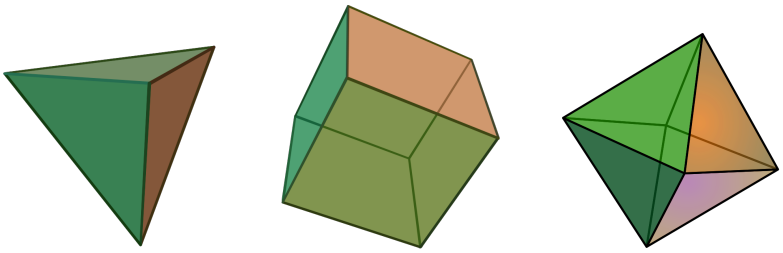
$$O_i(\omega) = O_{(\theta_i,\phi_i)}(\omega) = W(\omega_r) \cdot G_i(\theta, \phi),$$

where (θ_i, ϕ_i) represents the center frequencies in the orientation space. We also assume that the low-pass and the high-pass filters are spherically symmetric filters. Therefore, Equation 2.23 and Equation 2.24 can be satisfied, if the sum of the square of the oriented filters are independent from the angular variables θ and ϕ. In other words, that sum should be a spherically symmetric function. In order to solve this problem, a filter design technique for the angular parts is given in [72]. In this thesis, we adopted that technique.

Orientation Selectivity Characteristics of Filters: Let the angular parts of the oriented filters equal to $\left[\cos_{\theta_i,\phi_i}(\alpha)\right]^2$, $j\cos_{\theta_i,\phi_i}(\alpha)$, $-j\cos_{\theta_i,\phi_i}(\alpha)$ or $|\cos_{\theta_i,\phi_i}(\alpha)|$. That is;

$$O_{\theta_i,\phi_i}(\omega_r, \theta, \phi) = \begin{cases} W(\omega_r)\left[\cos_{\theta_i,\phi_i}(\alpha)\right]^2, \\ W(\omega_r) j \cos_{\theta_i,\phi_i}(\alpha), \\ W(\omega_r)(-j)\cos_{\theta_i,\phi_i}(\alpha), \\ W(\omega_r)|\cos_{\theta_i,\phi_i}(\alpha)|. \end{cases} \tag{2.26}$$

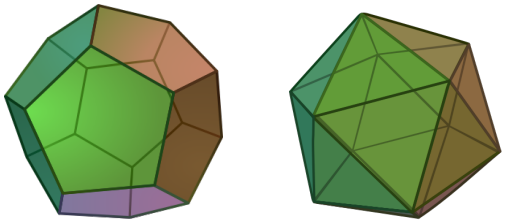
Since we assume that $W(\omega_r)$ is real and always positive, the second and third versions in Equation 2.26 have j and $-j$ as their multipliers, and the angular function of the fourth one is in absolute value to have real coefficients in the spatio-temporal domain. In Statement 1, we express the number of the oriented filters and their center frequencies in the orientation space. However, before we make the statement, we need to introduce the platonic solids. The platonic solids are the only convex volumetric shapes that have full symmetry. There are five known platonic solids. The readers can find more about the platonic solids in [73]. All of the known platonic solids are shown in Figure 2.12.



(a) Tetrahedron (four faces)

(b) Cube or Hexahedron (six faces)

(c) Octahedron (eight faces)



(d) Dodecahedron (twelve faces)

(e) Icosahedron (twenty faces)

Figure 2.12: Platonic solids figures which are captured from the website `http://en.wikipedia.org/wiki/Platonic_solid`. (Permission to use these figures are granted under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation. Copyright from the user in wikipedia with the nickname DTR.)

Now we proceed with the following statement;

Statement 1. *Let us assume that the platonic solids we mentioned above except the cube are circumscribed by a sphere. If the angular part of the filters is equal to* $\cos_{\theta_i,\phi_i}(\alpha)$*, the perfect reconstruction can be provided if the center frequencies of the filters in the orientation space,* (θ_i, ϕ_i)*, are the spherical coordinates of the vertices of the tetrahedron (four oriented filters) or the octahedron (three oriented filters). If the angular part is equal to* $\left[\cos_{\theta_i,\phi_i}(\alpha)\right]^2$*, The angles* (θ_i, ϕ_i)*, can be the spherical coordinates of the vertices of the dodecahedron (ten oriented filters) or the icosahedron (six oriented filters).*

An explicit proof of the above statement is not given in [72]. We include a proof for $\cos_{\theta_i,\phi_i}(\alpha)$ case in Appendix A.

Statement 1 somewhat different than the statement given in [72]. In that article, it is stated that, in the orientation space, if the center frequencies of the filters, which have angular function as $\cos_{\theta_i,\phi_i}(\alpha)$, are chosen as the vertices of the cube and the octahedron, the filters provide the perfect reconstruction. On the other hand, in Statement 1, instead of the cube, we mention that vertices of the *tetrahedron* provide perfect reconstruction. Actually, if only the vertices of the tetrahedron are chosen as the center frequencies (not vertices of any composition of rotated tetrahedrons and octahedrons), in the magnitude spectrum it is exactly the same as chosing the vertices of the cube. The reason for that is, as it is explained at the end of section 2.3.4, if the angular function is a cosine, orienting a filter towards a certain frequency point also means orienting the same filter towards the symmetric frequency points (symmetry is with respect to origin). And, the set of the coordinate points of the vertices of a tetrahedron with their symmetric ones is the same as the set of the coordinate points of the vertices of the cube.

However, the difference here is that, if the orientation selectivity function is chosen as $\cos_{\theta_i,\phi_i}(\alpha)$, the perfect reconstruction should also be satisfied when the coordinates of the vertices of compounds of differently rotated tetrahedrons, cubes or octahedrons are chosen. Here the word *compound* means that, platonic solids are combined such a way that their circumscribed spheres have the same center point in the three-dimensional space, but the platonic solids differ in their rotations. And a certain volumetric shape, whose vertices satisfy perfect reconstruction requirement, can be a compound of only rotated tetrahedrons. In other words, it may not include any cubes or octahedrons. Therefore, we find it beneficial to state the vertices of the tetrahedron, instead of the cube. For example, if the angular function of the filters are chosen as $\cos_{\theta_i,\phi_i}(\alpha)$, the coordinates of the vertices of the dodecahedron (compound of five tetrahedrons) or the icosahedron (compound of three tetra hedrons) provide the perfect reconstruction.

In this paragraph, we explain the characteristics of the orientation selectivity function, $\left[\cos_{\theta_i,\phi_i}(\alpha)\right]^n$, when the center frequencies in the orientation space are selected as the spherical coordinates of the vertices of four platonic solid. For $n = 1$, the characteristics of the cube and the octahedron, for $n = 2$, the characteristics of the dodecahedron and the icosahedron are given. The 3 dB bandwidth of the orientation selectivity function, $\alpha_{BW}(n)$, is stated in Equation 2.20. In addition, we express the angle, $\varphi(a, r)$, between the vectors whose elements are the cartesian coordinates of two closest points of the vertices of platonic solids. It is calculated by the formula; $\varphi = 2\arcsin(\frac{a}{2r})$, where a is the edge length of the platonic solid and r is the radius of the circumscribed sphere of platonic solid. The reason of giving the characteristics of the cube instead of the characteristics of the tetrahedron lies here. Since the angle, $\varphi(a, r)$, is smaller for a cube, giving the characteristics of the tetrahedron leads to wrong results. The magnitude of the angular functions of the filters in the half of $\varphi(a, r)$ is also given as $M(\varphi) = 20nlog_{10}(\cos(\varphi/2))$. By computing $M(\varphi)$, we find the most suitable platonic solid to use. Here, the suitability criteria is that, if $M(\varphi)$ becomes too low, then, some frequency content in the video may fade away in all the subbands. On the other hand, if $M(\varphi)$ becomes too high, some frequency content may be represented in more than one subband. Therefore, we desire $M(\varphi)$ to be close to $-3\ dB$, which is the known magnitude for the cut-off. However, depending on the application, any combination of the vertices of platonic solids can be used. Finally, we give the sum of the squares of the orientation selectivity functions for the vertices of the platonic solids. The reason of taking square is that in Equation 2.23 and Equation 2.24, the squares of the oriented filters are added up. That is;

$$F(n, V) = \sum_{i=1}^{V} \left[\cos_{\theta_i,\phi_i}(\alpha)\right]^{2n}, \tag{2.27}$$

where, V is half of the number of the vertices of the corresponding platonic solid. We give all these properties with their associated platonic solid in Table 2.2. In Appendix C, we give the cartesian coordinates, $C(x, y, t)$, and the spherical

coordinates, $S(\theta, \phi)$, of the selected vertices of the platonic solids based on the coordinate axes as in Fig. 2.5. However, these points can be rotated altogether by multiplying the coordinates with the rotation matrix stated in Equation 2.10.

	$\alpha_{BW}(n)$	$\varphi(a,r)$	M(φ) (dB)	F(n,V)
Cube or Regular Hexahedron	0.25π	0.39π	-1.76	1.33
Octahedron	0.25π	0.5π	-3	1
Dodecahedron	0.18π	0.23π	-1.18	2
Icosahedron	0.18π	0.35π	-2.81	1.2

Table 2.2: Some properties of the orientation selectivity function with their associated platonic solid are shown. Angles are expressed in radians.

By looking at M values, it can be said that, if n is equal to 1, the coordinates of vertices of the octahedron, if n is equal to 2, the coordinates of vertices of icosahedron are more appropriate to use. In this project, in order not to get results which are depending on the choice of the steerable pyramid filters, we designed two type of filter sets. One set uses the coordinates of the vertices of the compounds of three differently rotated octahedron and one cube and the other uses the coordinates of the vertices of the icosahedron. In the next section, we mention detailed design steps of the filters.

2.4.3 Explicit Design Steps of Three-Dimensional Steerable Pyramid Filters

Method I: In this method, we designed the oriented filters by using 1^{st} derivative of Gaussian filters. The low-pass and high-pass filters also have Gaussian characteristics. This filter set is useful when the video to be filtered has only a few frames. In addition, computational requirements can be reduced by the separable convolutions with this filter set. If a video has a small number of frames, the undesired border effects (extremely low or high responses) may dominate the

output of the filter. In our case, since we extract statistical information on the subbands, it is important that a filtered movie should not have significant border artifacts. The main drawback of this filter set is that the reconstruction is not perfect. In below paragraphs we give comments on the reconstruction error.

With the Gaussian and 1^{st} derivative of Gaussian filters, it is not possible to satisfy the flat system response and recursion requirements as given by Equation 2.23 and Equation 2.24 at the same time. In order to minimize the reconstruction error and see how much distortion occurs on the magnitude spectrum, we combined the flat system response and recursion restrictions into one equation. In this way, we found the overall frequency response of the steerable pyramid filterbank. In order to do this, we need to predetermine the number of the scales and we choose it as 3. This number is chosen not to have too few frames after the downsamplings.

It can be seen from Equation 2.15 that 1^{st} derivative of Gaussian filters orientation selectivity is determined by the funtion $\cos(\alpha)$. Therefore, according to the discussion given in the previous section and after trials of different volumetric shapes, we decided to use the spherical coordinates of the vertices of the compound of three octahedrons and a cube. Therefore, the number of the oriented filters, k, becomes 13. The spherical coordinates that we used are given in Table 2.3.

	1	2	3	4	5	6	
θ_c	0	$\pi/2$	$3\pi/2$	$\pi/2$	0	π	
ϕ_c	$\pi/2$	$\pi/4$	$\pi/4$	$\pi/2$	$\pi/4$	$\pi/4$	
	7	8	9	10	11	12	13
θ_c	0	$\pi/4$	$7\pi/4$	$\pi/4$	$3\pi/4$	$5\pi/4$	$7\pi/4$
ϕ_c	0	$\pi/2$	$\pi/2$	0.96π	0.96π	0.96π	0.96π

Table 2.3: Spherical coordinates which are used in the design of the steerable pyramid filters in Method I. Angles are expressed in radians

After achieving perfect reconstruction in the orientation space, the overall frequency response in the radial frequency, $S(\omega_r)$, of the pyramid becomes,

$$\begin{aligned} S(\omega_r) =& |H_0(\omega_r)|^2 + \\ & |L_0(\omega_r)|^2 \Big\{ \sum_{i=1}^{3} |O(\omega_r, \theta_i, \phi_i)|^2 + |L_1(\omega_r)|^2 \Big[\sum_{i=1}^{3} |O(\omega_r/2, \theta_i, \phi_i)|^2 + \\ & |L_1(\omega_r/2)|^2 \Big(\sum_{i=1}^{3} |O(\omega_r/4, \theta_i, \phi_i)|^2 + |L_1(\omega_r/4)|^2 \Big) \Big] \Big\} \; + \; a.t., \end{aligned} \tag{2.28}$$

where, $a.t.$ refers to aliasing terms and

$$\begin{aligned} L_0(\omega_r) = e^{-\frac{\sigma_{l0}^2}{2} \cdot \omega_r^2}, \; H_0(\omega_r) = \sqrt{1 - e^{-\sigma_{l0}^2 \cdot \omega_r^2}}, \; L_1(\omega_r) = e^{-\frac{\sigma_{l1}^2}{2} \cdot \omega_r^2}, \\ O(\omega_r, \theta_i, \phi_i) = N_i \sigma_d e^{\frac{1}{2}} j e^{-\frac{\sigma_d^2}{2} {\omega_r}^2} \omega_r cos_{\theta_i, \phi_i}(\alpha) \end{aligned} \tag{2.29}$$

We choose σ_{l1} as 1.6, so that $L_1(\pi/2) = 0.04$ which is close to zero. Therefore, aliasing terms can be ignored. Since we choose σ_{l1} as 1.6, we choose σ_{l0} as 0.8 so as to have $L_0(\omega_r) = L_1(\omega_r/2)$. If we choose σ_{l0} larger than 0.8, the reconstruction error increases rapidly. If it becomes smaller than 0.8, the band pass regions in different scales are not represented by the equal magnitudes. We also select σ_d as 1. This selection is suitable in terms of small reconstruction error, the center frequency and the bandwidth of the bandpass filters. Radially center frequency of the band pass filters is 1 rad. This value is not too close either to zero or π rad. The 3 dB bandwidth of these filters is 1.15 rad. Moreover, in order to represent the subbands efficiently, the passbands of $L_0(\omega)$ and $O(\omega_r, \theta_i, \phi_i)$ should overlap and we considered this fact in the choice of σ parameters. The 3 dB cutoff frequency of $L_0(\omega)$ is 0.33π rad and the center frequency of the band-pass filters is 0.4π rad, which are close enough. The normalization values N_1, N_2 and N_3 are selected such that to provide minimum reconstruction error and normalize the sum of the oriented filters. If $N_1 = 1/\sqrt{4.33}$, $N_2 = 0.95/\sqrt{4.33}$ and $N_3 = 0.87/\sqrt{4.33}$, the frequency response $S(\omega_r)$ becomes as in Figure 2.13.
From this figure, it can be seen that, the frequency response takes its minimum

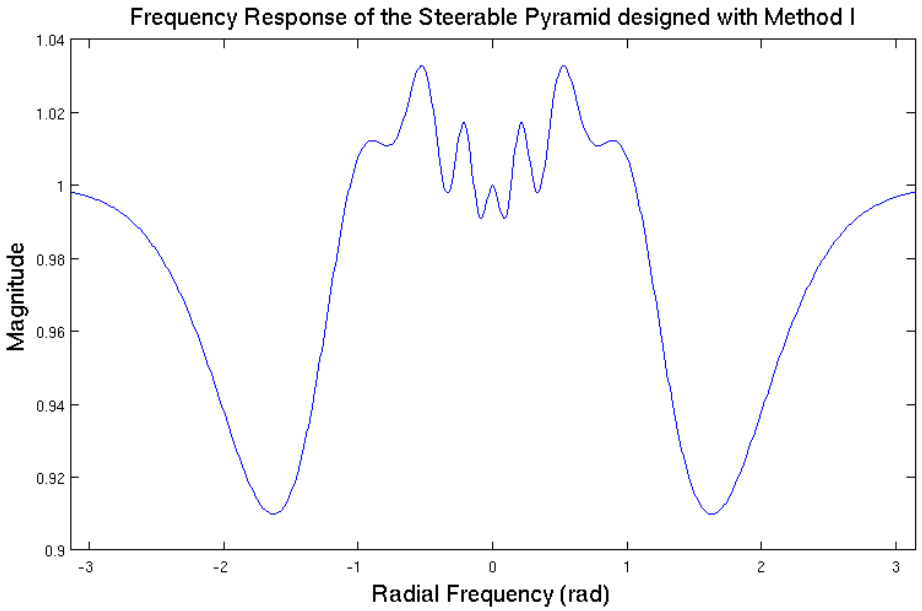


Figure 2.13: Frequency response of the steerable pyramid designed with Method I

value as 0.91 at around radial frequency of 0.5π rad and it takes its maximum value as 1.035 at around radial frequency of 0.22π rad.

All the filters except the high-pass one are separable. We found spatio-temporal domain coefficients of the high pass filter by taking inverse three-dimensional DFT of the function given in Equation 2.29. We designed all other filters as explained in the previous section. 1^{st} derivative of Gaussian filter in any orientation can be written in the steerable form as

$$\begin{aligned} G^{(1)}_{\theta_i,\phi_i}(\omega_x,\omega_y,\omega_t) &= \cos(\theta_i)\sin(\phi_i)(j\omega_x e^{-\frac{\sigma_d^2}{2}\omega_x^2})e^{-\frac{\sigma_d^2}{2}\omega_y^2}e^{-\frac{\sigma_d^2}{2}\omega_t^2} \\ &+ \sin(\theta_i)\sin(\phi_i)(j\omega_y e^{-\frac{\sigma_d^2}{2}\omega_y^2})e^{-\frac{\sigma_d^2}{2}\omega_x^2}e^{-\frac{\sigma_d^2}{2}\omega_t^2} \\ &+ \cos(\phi_i)(j\omega_t e^{-\frac{\sigma_d^2}{2}\omega_t^2})e^{-\frac{\sigma_d^2}{2}\omega_x^2}e^{-\frac{\sigma_d^2}{2}\omega_y^2} \end{aligned} \tag{2.30}$$

As it can be seen from these filters, the filtering operations can be done by one-dimensional convolutions. As it is stated before, we used spherical coordinates of the vertices of the compound of three rotated octahedrons and a cube.

We give the spatio-temporal filter coefficients mentioned in this section in Appendix B.

The filters that we mentioned until now are for the decomposition (analysis) part. Since the low-pass and the high-pass filters are zero phase filters, they can be directly used in the synthesis part. However, the band-pass filters are not zero phase filters, thus, in the reconstruction part we use their symmetric versions with respect to origin. Moreover, the downsampling and upsampling reduce the energy of the video to 1/8 of its value. We compensate this reduction by multiplying the video by $2\sqrt{2}$ while downsampling and upsampling. Finally, in the convolutions, we pad the movies in a symmetric manner. Let one dimensional signal to be convolved be $f(n)$, where n is an integer such that $0 \leqslant n \leqslant L-1$. Also let the filter length be T. In the convolutions, we assume that $f(-k) = f(k)$ and $f(L-1+k) = f(L-1-k)$, where k is an integer and $0 \leqslant k \leqslant (T-1)/2$. For example, if the signal to be convolved is $[a\ b\ c\ d\ e]$ ($L = 5$), and the filter length, T, is 3, we assume the signal to be convolved as $[b\ a\ b\ c\ d\ e\ d]$. The reason of this is to prevent the spurious responses after upsampling and filtering with $L_1(\omega)$ and the spurious responses occur if the signal has two consecutive zeros or two consecutive nonzeros at the borders after upsampling. In order to have the same dimensions at the input of the pyramid and the output of the pyramid, in each scale, the signal size after upsampling should be the same as the signal size before downsampling. Let the signal to be upsampled has length L. Depending on the signal length before downsampling, the signal length after upsampling can be either $2L$ or $2L+1$. The padding method that we described prevents being two consecutive zeros or two consecutive nonzeros at the borders in either case. For example, let us assume that the signal $[a\ b\ c]$ is to be upsampled. Depending on the signal length before downsampling, the upsampling operation results in either $[a\ 0\ b\ 0\ c]$ or $[a\ 0\ b\ 0\ c\ 0]$. Then, for example, if the length of $L_1(\omega)$ is three, then the signal to be filtered is either $[0\ a\ 0\ b\ 0\ c\ 0]$ or $[0\ a\ 0\ b\ 0\ c\ 0\ c]$. Any

other padding method (e.g. circularly, symmetric or zero padding) might result in border artifacts by placing two consecutive zeros or nonzeros. However, the padding method that we use guarantees that the border artifacts do not occur whatever the signal after upsampling is. If all the videos would have dimensions of integer exponents of 2, the circular convolution could also prevent the border artifacts.

Method II: In this method, the filters are nonseperable and have larger sizes. The advantage of this set of filters is that they have narrower bandwidth both in radial and angular frequencies. In addition, they give better results in terms of reconstruction. However, filter size is higher compared to the filters designed with Method I and the filters are nonseparable.

In this method we first designed the radially symmetric low pass filter, $L_0(\omega_r)$, by using the algorithm for the multidimensional radially symmetric filters explained in [74]. This algorithm provides zero phase filters. Therefore, specifying the magnitude spectrum is enough to design the filters. In that algorithm, there are some parameters about the filter to be specified. They are, the cut off radial frequency of the pass band region, ω_{pass}, the cut off radial frequency of the stop band region, ω_{stop} and the tolerance values in the stopband region, δ_{stop} and in the passband region, δ_{pass}. They are shown in Figure 2.14.

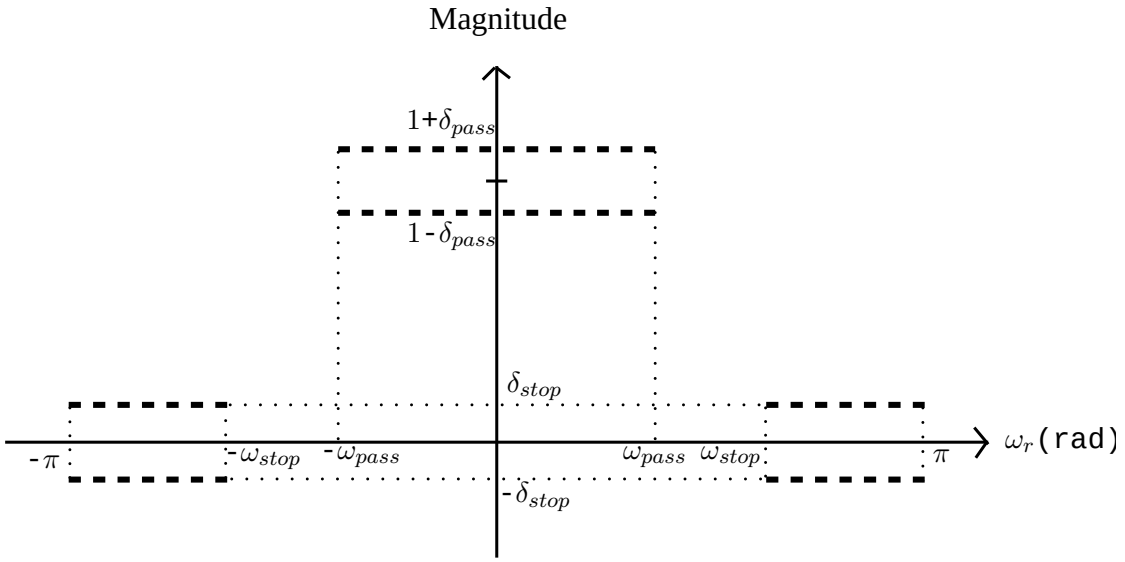


Figure 2.14: The filter specifications are shown. The magnitude of the filter is specified for the intervals shown within dashed lines. The regions that remains outside of these intervals are the transition regions.

We chose the parameters as $\omega_{pass} = \pi/2$ rad, $\omega_{stop} = 3\pi/4$ rad, $\delta_{stop} = \delta_{pass} = 0.05$ rad. This is an iterative algorithm to get a radially symmetric filter whose DTFT is as specified by the parameters. We decided the number of iterations as 30, and the filter size as $9 \times 9 \times 9$. After generating a low pass filter by using the algorithm, we make a small modification on the filter. That is, we brought its DFT coefficients between 0 and 1. This modification is required in order to satisfy flat system response and recursion constraints. To do the modification, we computed the DFT, $F(k_{x_{old}}, k_{y_{old}}, k_{t_{old}})$, of the coefficients of the designed filter and applied a simple linear operation on the DFT coefficients. That is, $F(k_{x_{new}}, k_{y_{new}}, k_{t_{new}}) = \big[F(k_{x_{old}}, k_{y_{old}}, k_{t_{old}}) - min\big]/(max - min)$, where min and max are the minimum and maximum of the coefficients $F(k_{x_{old}}, k_{y_{old}}, k_{t_{old}})$. Then, by computing the IDFT from new coefficients, we got the final low pass filter, $L_0(\omega_r)$.

After we designed $L_0(\omega_r)$ in the spatio-temporal domain, we proceeded to design the remaining filters according to flat system response and recursion requirements. In order to do this, we computed the DFT coefficients of $H_0(\omega_r)$ from the DFT coefficients of $L_0(\omega_r)$, according to $\sqrt{1-|L_0(\omega_r)|^2}$. Then, by computing IDFT of $H_0(\omega_r)$, we find the spatio temporal domain coefficients of the high-pass filter.

We designed the filter, $L_1(\omega_r)$, by using the DFT of $L_0(\omega_r)$. After computing the DFT of $L_0(\omega_r)$, we padded zeros to this DFT and increased the DFT size to $17 \times 17 \times 17$ (As a reminder, the size of the filter, $L_0(\omega_r)$, is $9 \times 9 \times 9$). This provides the equality $L_0(\omega_r) = L_1(\omega_r/2)$ and $L_1(\omega_r) = 0$ when $\omega_r > \pi/2$. After computing IDFT of this $17 \times 17 \times 17$ sized filter, we find the spatio-temporal coefficients of the filter $L_1(\omega_r)$.

Then we moved on to the design of the oriented filters. In order to design, we assumed the filters as separable functions as explained in section 2.1.4. Since the radial part of the oriented filters are the same, we first designed that part. The radial part of the oriented filters have been assumed to be high-pass filters, contrary to the first method of filter design. The design procedure is exactly the same as the design of $H_0(\omega_r)$. Here, instead of using $L_0(\omega_r)$, we used the DFT coefficients of$L_1(\omega_r)$ and computed the DFT coefficients of $O_i(\omega_r)$ from the equation $\sqrt{1-|L_1(\omega_r)|^2}$. In order to make the spherically symmetric filter to have orientation selectivity, we multiplied the filter coefficients with the function;

$$\cos_{\theta_i,\phi_i}(\alpha)^2 = \big[\sin(\phi_i)\sin(\phi)\cos(\theta_i)\cos(\theta) + \sin(\phi_i)\sin(\phi)\sin(\theta_i)\sin(\theta) + \cos(\phi_i)\cos(\phi)\big]^2 \tag{2.31}$$

This equation is written from Equation 2.16. The angle pairs (θ_i, ϕ_i) are the spherical coordinates of the vertices of the *icosahedron* and the (θ,ϕ) are the spherical coordinates of the grid points of the DFT of the oriented filters.

As an example, let the index of a coefficient in the cartesian coordinates be $(5, -3, 4)$ and the filter index goes from -8 to 8 for each axis. (Since the filter size is $17 \times 17 \times 17$). The spherical coordinates, (θ_i, ϕ_i) of $(5, -3, 4)$ are $\arccos\left[5/\sqrt{5^2 + (-3)^2}\right]$ and $\arccos\left[4/\sqrt{5^2 + (-3)^2 + 4^2}\right]$, respectively. Moreover, we normalized the functions by multiplying them by $1/\sqrt{1.2}$, in order to make the sum of the oriented filters equal to 1.

As we explained at the end of Method I, we used here symmetric padding to prevent undesired border effects and multiplied the filter outputs by $2\sqrt{2}$ both in downsampling and upsampling.

As we mentioned before, some movies showing individual responses of the steerable pyramid filters as well as reconstruction of those movies by using each steerable pyramid in `http://www.umram.bilkent.edu.tr/~kulce/`.

To sum up, in this chapter we give the technical details of the image processing tools that we utilize on the surface reflectance recognition. We first mention the orientation and scale concept and a design technique of orientation and scale selective filters. Then, we introduce the steerable filter concept and we move on to the derivative of Gaussian filters. In that section, we give orientation and scale selectivity characteristics and the design methods of derivative of Gaussian filters for arbitrary derivative orders. Then, since we designed second derivative of Gaussian filters, we explain the details about second derivative of Gaussian filters. Finally, we mention the steerable pyramid and its properties. Moreover, we give our adopted filter design techniques for steerable pyramid.

Chapter 3

EXPERIMENTAL RESULTS

This chapter is devoted to the experimental results related to recognition of surface reflectance from statistics of band-pass filter outputs.

First of all we give information about our dataset and mathematical expression of the statistical parameters that we used. Then, we explain the details of our analysis methods. Our analysis steps can be grouped under four sections. First, since the data set includes movies of one matte and one shiny version for each object, we show pairwise statistical differences between matte and shiny versions. Secondly, we find the average subband statistics of matte and shiny object motions and express their differences. Thirdly, we designed a classifier from statistics of matte and shiny objects and we give the classification results. Finally, in order to check whether the subband statistics are the sufficient cues to surface reflectance recognition, we attempted to synthesize new movies which have the same subband statistics with the motions of matte and shiny objects by using the texture synthesis algorithm proposed in [55]. In that section, since we use two steerable pyramids, we first mention the general success rate of the steerable pyramids on three dimensional texture synthesis and then we give the results related to synthesis of textured movies.

3.1 Data Set and Statistical Parameters

3.1.1 Data Set

The dataset includes 20 matte object and 20 shiny object motions. These motions are in pairs. In other words, there are 20 pairs and in each pair there are one matte and one shiny version of an object. The motion of the objects are generated with a simple rotation. The movies are computer-generated motions, so, they are not real scenes. The motions of shiny objects are generated such that they reflect the environment that they are placed into. Their matte versions, on the other hand, are painted as though they reflect their environments when they are stationary. However, when they start to rotate, it is understood that they just carry the painting, instead of reflecting the environment. The main difference between appearances of the matte and shiny objects is that, in flat regions on the object surface, specular features move faster than the object itself and in convexly curved regions, specular features move slower than the object. On the other hand, if the object is matte, the paintings on the object surface move with the same speed along with the object surface. In our dataset, the general shape of the objects has the appearance of a potato, but they have different curved and flat regions. The data set is outlined in Figure 3.1. In addition, in Figure 3.2 and Figure 3.3, we give sample video frames of a matte object and its shiny counterpart.

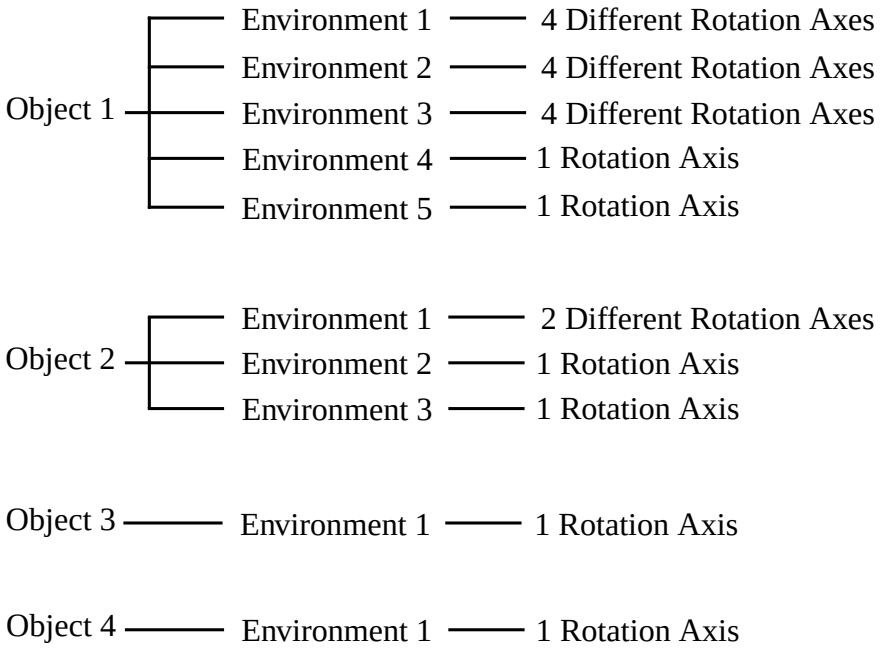
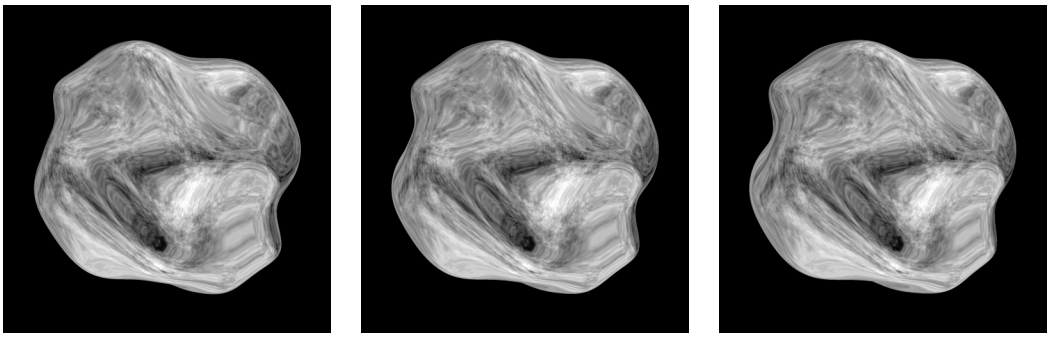
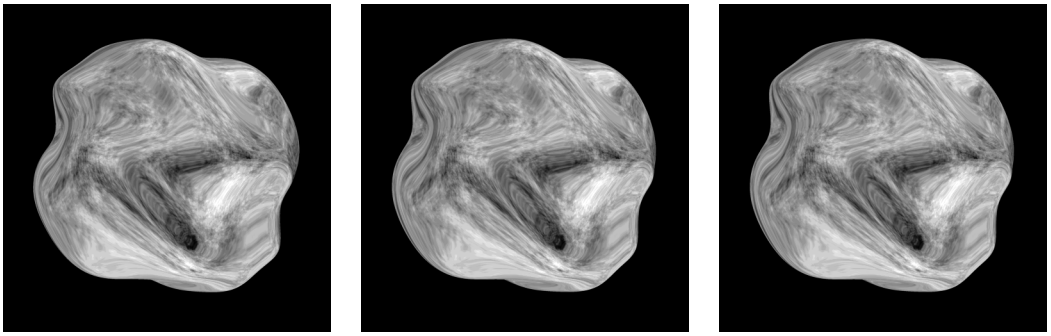


Figure 3.1: The dataset scheme is shown. At the end of each branch, there is a pair of movie which includes a shiny object and a matte object.

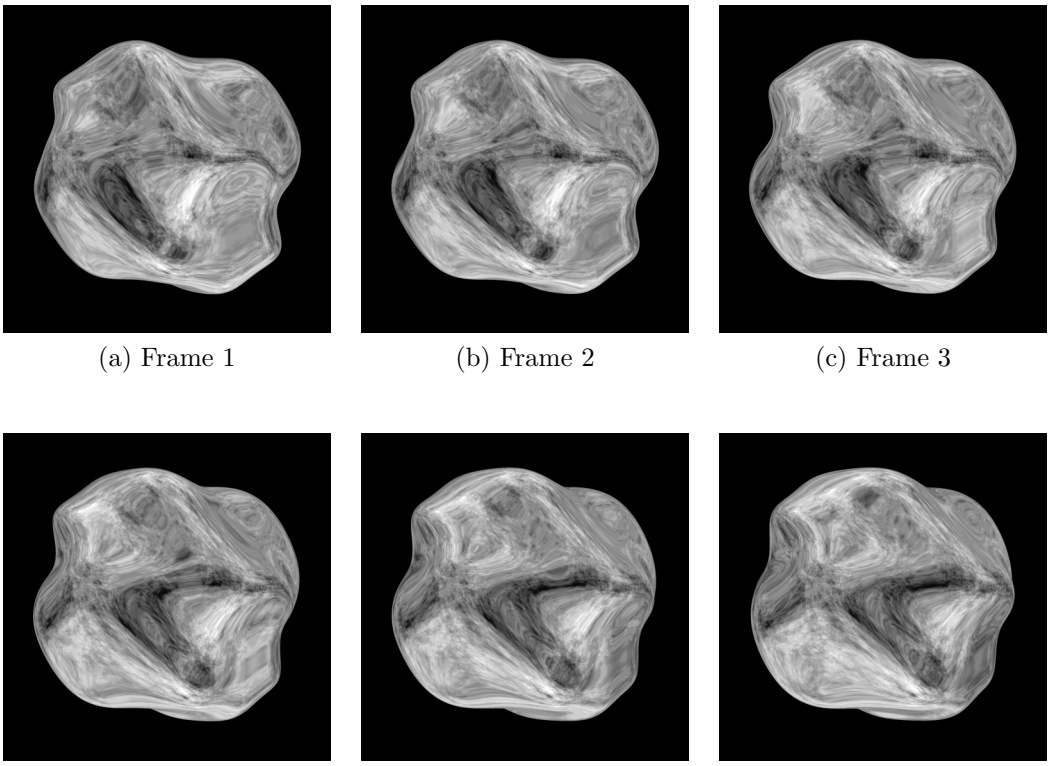


(a) Frame 1 (b) Frame 2 (c) Frame 3



(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.2: Six frames from a moving matte object

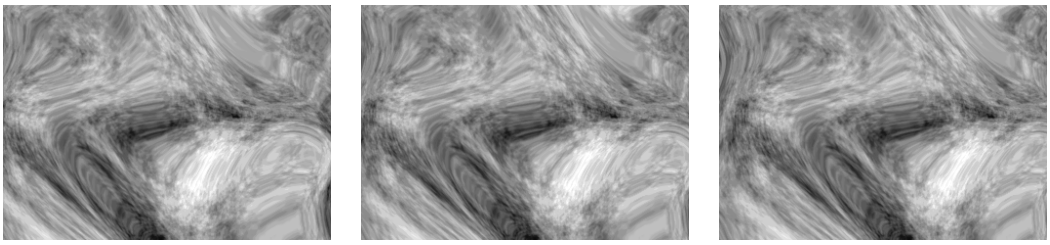


(a) Frame 1 (b) Frame 2 (c) Frame 3

(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.3: Six frames from a moving shiny object

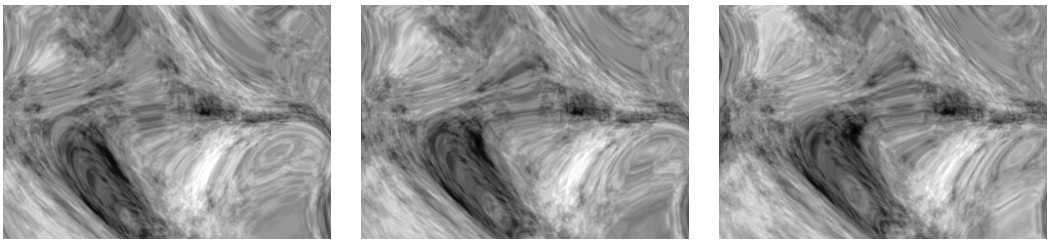
Since the black background behind the objects has dominant effect on the statistics, we use only the center part without the background. After removing the background, the frames given in Figure 3.2 and Figure 3.3 look like as in Figure 3.4 and Figure 3.5



(a) Frame 1 (b) Frame 2 (c) Frame 3

(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.4: Six frames showing the center parts of a moving matte object



(a) Frame 1 (b) Frame 2 (c) Frame 3

(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.5: Six frames showing the center parts of a moving shiny object

3.1.2 Statistical Parameters

Since we use subband statistics for the analysis, we first compute the outputs of the bandpass filters, which are second derivative of Gaussian filters that we presented in Chapter 2. We compute 24 filter outputs (eight orientations for three different scales) for each movie. Although we use symmetric padding for convolutions, in order not to have border artifacts, we exclude borders in all for three dimensions after filtering.

The statistics that we measured are all extracted from pixel histograms of the subbands. The pixel histogram is an array whose elements are the counts of the pixels which have the same intensity levels in an image or a video. The first index of the histogram corresponds to minimum intensity level (black color) and the last index corresponds to maximum intensity level (white color). After computing the outputs of the subbands, we linearly normalized pixel intensities to have minimum pixel intensity as 0 and maximum pixel intensity as 1. Making this normalization for each subband separately is logical, since each orientation selective cell gives response according to the local intensity differences, instead of the overall brightness level [11]. Since the graphic card in our computer supports 8-bit resolution, we assume that there are $2^8 = 256$ intensity levels in the interval [0,1] and we used those quantized values for the intensity values in the subbands. Moreover, we divided each element in the histogram by the total number of pixels in the filter output. Therefore, the histogram of each subband can be assumed as the representation of the probability density function (pdf) and the corresponding subband output can be assumed as a random process. In Figure 3.6, a sample histogram is shown.

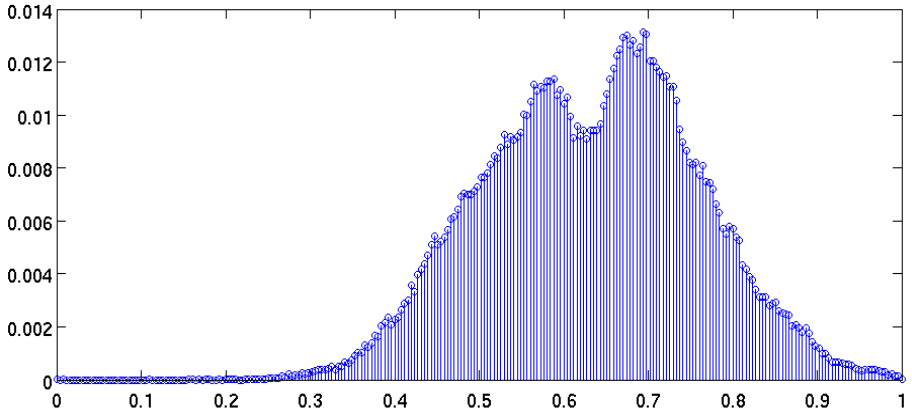


Figure 3.6: A sample histogram

Let a histogram represent the pdf, $p_X(i)$, where X is a random variable which takes one of the 256 intensity levels in the interval [0 1]. From this pdf we extracted four statistical parameters. These parameters are the mean, standard deviation, skewness and kurtosis.

Mean: The mean, μ, is,

$$\mu = E[X] = \sum_{i=0}^{1} i \cdot p_X(i), \text{ where } i = 0, \ \frac{1}{255}, \ \frac{2}{255}, \ \ldots, \ 1. \tag{3.1}$$

Standard Deviation: The standard deviation, σ, is,

$$\sigma = \sqrt{E[(X-\mu)^2]} = \sum_{i=0}^{1} (i-\mu)^2 \cdot p_X(i), \text{ where } i = 0, \ \frac{1}{255}, \ \frac{2}{255}, \ \ldots, \ 1. \tag{3.2}$$

The standard deviation is a measure of average contrast.

Skewness: The skewness, μ_3, is,

$$\mu_3 = E[(X-\mu)^3] = \sum_{i=0}^{1} (i-\mu)^3 \cdot p_X(i), \text{ where } i = 0, \ \frac{1}{255}, \ \frac{2}{255}, \ \ldots, \ 1. \tag{3.3}$$

The skewness is a measure of the symmetry with respect to mean of the histogram. This value takes a positive number if the histogram plot is skewed towards right and takes a negative number if the histogram plot is skewed towards left.

Kurtosis: The kurtosis, κ, is,

$$\kappa = \frac{E[(X-\mu)^4]}{\sigma^4} = \frac{\sum_{i=0}^{1}(i-\mu)^4 \cdot p_X(i)}{\sigma^4}, \text{ where } i = 0, \frac{1}{255}, \frac{2}{255}, \ldots, 1. \quad (3.4)$$

The kurtosis is a measure of the peakedness of the histogram.

3.2 Numerical Results

3.2.1 Comparison of Matte-Shiny Pairs

As we stated previously, the data set consists of videos of moving matte and shiny versions of each object. We have a total of 20 pairs of movies and we investigated statistical differences of each pair. Since we have 24 subband histograms for each movie and four statistical parameters from each histogram, we compared each pair for $24 \times 4 = 96$ parameters.

The comparison was made in a way that, for a certain parameter and a certain subband, we find the number of pairs in which the specified parameter is greater for subband histogram of matte object motion than its shiny pair. For example, let the parameter that is subject to comparison be the standard deviation, σ_i, which is extracted from histograms of a particular scale and orientation selective filter outputs. The number, c, represents that in c of twenty movies, the specified parameter is greater for matte object than that of the shiny object. We give those numbers in Table 3.1.

		O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8
μ	S_1	7	7	11	8	9	8	12	8
	S_2	11	9	12	13	13	12	14	11
	S_3	10	11	10	10	11	10	13	9
σ	S_1	8	14	12	13	16	17	15	15
	S_2	12	14	15	13	13	17	16	14
	S_3	14	14	17	18	18	17	14	11
μ_3	S_1	10	13	8	7	8	9	9	8
	S_2	7	7	7	6	5	6	7	5
	S_3	7	8	7	7	6	10	8	6
κ	S_1	14	14	11	8	6	7	7	9
	S_2	11	11	9	9	9	5	5	7
	S_3	6	10	4	3	5	6	8	10

Table 3.1: Each column is a specific orientation and each row is a specific scale. The symbol O_i represents the i^{th} orientation and S_i represents the i^{th} scale. The scale index 1 indicates the highest radial frequency and 3 indicates the lowest radial frequency. The parameters which are subject to comparison are written on the left. μ, σ, μ_3 and κ are mean, standard deviation, skewness and kurtosis respectively. The explanation on how to read this table is given in the text.

Comparison of Means

By looking at the results given in Table 3.1, it can be stated that, there is not much difference between matte and shiny objects in terms of subband histogram means. There are two reasons of it. Firstly, if the subbands are investigated separately, the number of matte object motions whose mean is higher is approximately half of the twenty pairs in all of the subbands. Secondly, when the subbands are investigated as a whole, in eight of twenty four subbands, the mean of the shiny object motion is greater and in twelve of twenty four subbands, the mean of the matte object motion is greater. In addition, in four subbands, the mean of the matte object is higher in ten pairs and the mean of the shiny object is higher in remaining ten pairs. Therefore, we can state that an accurate recognition in a matte-shiny motion pair cannot be made by just looking at their subband histogram means.

Comparison of Standard Deviations

In terms of standard deviation, the separation between matte and shiny object motions is more obvious. In only one subband, the number of motion pairs whose standard deviation of histogram of shiny object is higher than its matte counterpart. In all the remaining subbands, the number of the matte object motions whose standard deviation is higher than their shiny counterparts is larger. Furthermore, in 13 of the 24 subbands, the standard deviation of matte object motion is greater in 15 or more of 20 motion pairs. Therefore, in our dataset, by looking at the standard deviations of the subbands of a given pair, a more accurate decision about which movie is the motion of matte one can be made.

Comparison of Skewnesses

In many of the subbands (21 of 24), the skewness is larger for motions of shiny objects. As in the standard deviation case, for a given pair, an accurate decision can be made whether a motion belongs to matte object or not by looking at the subband histogram skewnesses. However, although in many subbands the skewness is higher for shiny object motions, the difference is not as large as the standard deviation case: only in two of 24 subbands, the skewness of shiny object motion is greater in 15 or more of 20 motion pairs.

Comparison of Kurtosises

In terms of kurtosis, the difference between shiny and matte object motions is not large compared to the standard deviation and the skewness cases, but, the

separation is better compared to the mean case. In 17 of 24 subbands, the histograms of shiny object motions are more kurtotic than their matte counterparts and in five of them the kurtosis of shiny object motion is greater in 15 or more of 20 motion pairs

3.2.2 Average Statistics of Matte and Shiny Motions

In order to find the difference of average statistics, first we summed statistical parameters which belong to matte motions or shiny motions and divided each parameter by 20 which is the number of the matte object and shiny object movies. Then, we subtracted shiny motion parameters from matte motion parameters. That is,

$$\mu_{dif}^{ij} = \frac{1}{20}\sum_{k=1}^{20}(\mu_{mk}^{ij} - \mu_{sk}^{ij}), \qquad \sigma_{dif}^{ij} = \frac{1}{20}\sum_{k=1}^{20}(\sigma_{mk}^{ij} - \sigma_{sk}^{ij}),$$

$$\mu_{3_{dif}}^{ij} = \frac{1}{20}\sum_{k=1}^{20}(\mu_{3_{mk}}^{ij} - \mu_{3_{sk}}^{ij}), \qquad \kappa_{dif}^{ij} = \frac{1}{20}\sum_{k=1}^{20}(\kappa_{mk}^{ij} - \kappa_{sk}^{ij}),$$

where ij represents the i^{th} scale and j^{th} orientation and mk and sk represent the k^{th} movie of the *matte* motion dataset and k^{th} movie of the *shiny* motion dataset, respectively. In Table 3.2, we give these values.

From Table 3.2, it can be stated that the average of the means do not give much information about shininess of an object. Since in approximately half of the subbands, the mean which is extracted from the average subband histogram is higher for matte object motions. On the other hand, in all of the subbands, the variance of the average histogram of the matte object motion is higher. Moreover, in a few subbands (2 of 24), the skewness of the shiny object motion is smaller. As being in the mean parameter, in half of the subbands, the shiny motion is more kurtotic than the matte motion. These results agree with the ones expressed in

		O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8
μ_{dif}	S_1	-0.0216	-0.0234	0.0040	0.0018	-0.0067	-0.0120	0.0022	-0.0159
	S_2	0.0029	-0.0059	0.0104	0.0060	0.0094	0.0161	0.0222	0.0150
	S_3	-0.0042	0.0034	0.0060	-0.0051	-0.0121	-0.0111	0.0116	0.0058
σ_{dif}	S_1	0.0001	0.0039	0.0039	0.0083	0.0091	0.0096	0.0047	0.0050
	S_2	0.0084	0.0068	0.0134	0.0110	0.0087	0.0128	0.0111	0.0114
	S_3	0.0108	0.0099	0.0208	0.0166	0.0150	0.0175	0.0134	0.0105
$\mu_{3_{dif}}$	S_1	0.0001	0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
	S_2	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
	S_3	-0.0002	-0.0001	-0.0001	-0.0001	-0.0002	-0.0001	-0.0001	-0.0001
κ_{dif}	S_1	1.7869	0.6211	0.5747	-0.1531	-0.1253	0.7740	0.3900	1.0191
	S_2	1.1207	0.6473	-0.3088	-0.4635	-0.5107	-0.4001	-0.0653	0.8296
	S_3	0.4280	-0.1003	-0.4693	-0.7003	-0.6134	-0.6276	0.0551	0.4814

Table 3.2: Each column is a particular orientation and each row is a particular scale. The symbol O_i represents the i^{th} orientation and S_i represents the i^{th} scale. The scale index 1 indicates the highest radial frequency and 3 indicates the lowest radial frequency. The parameters which are subject to comparison are written on the left. μ_{dif}, σ_{dif}, $\mu_{3_{dif}}$ and κ_{dif} are mean, standard deviation, skewness and kurtosis respectively.

Section 3.2.1. As in separate investigation of the matte and shiny objects, the variance and the skewness of the subband histograms give more information than the mean and the kurtosis about shininess.

3.2.3 Classification Results

In this thesis, we tested five machine learning algorithms for recognition of a motion which belongs to either a matte or a shiny object. We compared the accuracy of the results of these algorithms in order to verify classification results are independent of the specific classifier. These algorithms are the least-squares, the support vector machine (SVM), the Bayesian classifier, the Bayesian classifier after the principal component analysis (PCA) and the Bayesian classifier after the linear discriminant analysis (LDA). Since we get most classification success rates from the least squares and the SVM algorithms and the success rates of the least-squares and the SVM algorithms are very much the same, we give just the results of the SVM algorithm. In these algorithms, we used the statistical parameters which are extracted from motion subband histograms as the movie features. As

we stated before, these parameters are the mean, standard deviation, skewness and kurtosis. Since there are 24 subbands, in this classification algorithm we used $24 \times 4 = 96$ features for each movie. Therefore, a movie is represented by a feature vector which contains 24 mean, 24 standard deviation, 24 skewness and 24 kurtosis values. Moreover, we also classified the movies just based on only their mean, standard deviation, skewness and kurtosis parameters. We used the classification algorithms provided by PRTools for Matlab [75].

Preprocessing Stage: In order to get an independent result from selection of the training and test samples, we ran the classification algorithm 1000 times. In each run, first of all, we randomly selected 70% of the matte movies as matte motion training samples and 70% of the shiny movies as shiny motion training samples. Therefore, in the learning step, we have 28 of 40 movies whose half is matte object and half is shiny object motions.

Secondly, in order to classify the movies correctly, the classification algorithm requires that each feature which belongs to a certain movie should be in the same order. For example, the mean values which are extracted from Equation 3.1 given in Section 3.1.2 are in the order of 10^{-1} (for example, 0.5, 0.7, 0.3 etc.), whereas the skewness values extracted from Equation 3.3 are in the order of 10^{-5} (for example 0.00005, 0.00007, 0.00003 etc.). In order to fix this problem, we made a normalization on the training data. In the PRToolbox, the function in *scalem.m* file does this normalization. The main idea of this normalization is that each feature is provided to be an outcome of a pdf whose mean is zero and variance is one. Specifically, let p_{ij} be a parameter (μ, σ, μ_3 or κ) array whose entries are the parameter values which are extracted from histograms of i^{th} orientation and j^{th} scale filter output of the. That is, $p_{ij} = [p_{ij}^{m1}\ p_{ij}^{m2}\ \ldots\ p_{ij}^{m14}\ p_{ij}^{s1}\ p_{ij}^{s2}\ \ldots\ p_{ij}^{s14}]$. The letters m and s represent the words *matte* and *shiny* and their indexes go up to 14, since we use 14 matte object and 14 shiny object movies in the training

set. Let the mean, $\mu_{p_{ij}}$ calculated from these values be;

$$\mu_{p_{ij}} = \sum_{k=1}^{14} \frac{\left(p_{ij}^{mk} + p_{ij}^{sk}\right)}{28},$$

and the standard deviation, σ_{ij}, be;

$$\sigma_{p_{ij}} = \sqrt{\sum_{k=1}^{14} \Big[\frac{\left(p_{ij}^{mk} - \mu_{p_{ij}}\right)^2 + \left(p_{ij}^{sk} - \mu_{p_{ij}}\right)^2}{28} \Big]}.$$

Then, each entry in p_{ij} is updated according to the equation $(c - \mu_{p_{ij}})/\sigma_{p_{ij}}$, where c represents an element in p_{ij}. Finally, each parameter in the test set is also normalized by using the parameters $\mu_{p_{ij}}$ and $\sigma_{p_{ij}}$.

Classification Stage: In this part, we give the classification results based on the SVM classifier. This classifier is based on finding a linear decision boundary which separates each class from others. The mathematical background of this algorithm is out of the scope of this thesis. Details of the procedure, as well as other types of classification algorithms can be found in [76].

The classification results based on **all four statistical** parameters for 1000 run of the SVM algorithm are as follows;

- The average success rate is 74%.
- The standard deviation of the success rates is 11.
- The minimum success rate is 33%.
- The maximum success rate is 100%.
- Confusion matrix for average success rate is;

		Estimated Label	
		Matte	Shiny
True Label	Matte	71	29
	Shiny	23	77

- The standard deviation of the success rates of matte object motion classification is 18.
- The minimum success rate for matte object motion classification is 17%.
- The maximum success rate for matte object motion classification is 100%.
- The standard deviation of the success rates of shiny object motion classification is 18.
- The minimum success rate for shiny object motion classification is 0%.
- The maximum success rate for shiny object motion classification is 100%.

The results of the classification algorithms show that if the mean, standard deviation, skewness and kurtosis parameters are used altogether in the feature set, approximately 75% of the movies can be classified correctly. The motions of the shiny objects can be classified with a bit more success. Although the minimum and maximum success rates indicate the classification success may depend on the choice of training and test sets, the low standard deviation values signal the robustness of the algorithm against the choice of these sets.

The classification results based on just the **mean** values for 1000 run of the SVM algorithm are as follows;

- The average success rate is 57%.
- The standard deviation of the success rates is 13.
- The minimum success rate is 17%.
- The maximum success rate is 100%.
- Confusion matrix for average success rate is;

		Estimated Label	
		Matte	Shiny
True Label	Matte	58	42
	Shiny	45	55

- The standard deviation of the success rates of matte object motion classification is 21.
- The minimum success rate for matte object motion classification is 0%.
- The maximum success rate for matte object motion classification is 100%.
- The standard deviation of the success rates of shiny object motion classification is 21.
- The minimum success rate for shiny object motion classification is 0%.
- The maximum success rate for shiny object motion classification is 100%.

The classification results based on just the **standard deviation** values for 1000 run of the SVM algorithm are as follows;

- The average success rate is 54%.
- The standard deviation of the success rates is 13.
- The minimum success rate is 0%.
- The maximum success rate is 92%.
- Confusion matrix for average success rate is;

		Estimated Label	
		Matte	Shiny
True Label	Matte	48	52
	Shiny	39	61

- The standard deviation of the success rates of matte object motion classification is 20.
- The minimum success rate for matte object motion classification is 0%.
- The maximum success rate for matte object motion classification is 100%.
- The standard deviation of the success rates of shiny object motion classification is 23.
- The minimum success rate for shiny object motion classification is 0%.
- The maximum success rate for shiny object motion classification is 100%.

The classification results based on just the **skewness** values for 1000 run of the SVM algorithm are as follows;

- The average success rate is 53%.
- The standard deviation of the success rates is 12.
- The minimum success rate is 8%.
- The maximum success rate is 83%.
- Confusion matrix for average success rate is;

		Estimated Label	
		Matte	Shiny
True Label	Matte	43	57
	Shiny	37	63

- The standard deviation of the success rates of matte object motion classification is 20.
- The minimum success rate for matte object motion classification is 0%.

- The maximum success rate for matte object motion classification is 100%.
- The standard deviation of the success rates of shiny object motion classification is 23.
- The minimum success rate for shiny object motion classification is 0%.
- The maximum success rate for shiny object motion classification is 100%.

The classification results based on just the **kurtosis** values for 1000 run of the SVM algorithm are as follows;

- The average success rate is 61%.
- The standard deviation of the success rates is 11.
- The minimum success rate is 25%.
- The maximum success rate is 92%.
- Confusion matrix for average success rate is;

		Estimated Label	
		Matte	Shiny
True Label	Matte	65	35
	Shiny	42	57

- The standard deviation of the success rates of matte object motion classification is 19.
- The minimum success rate for matte object motion classification is 0%.
- The maximum success rate for matte object motion classification is 100%.
- The standard deviation of the success rates of shiny object motion classification is 18.

- The minimum success rate for shiny object motion classification is 0%.
- The maximum success rate for shiny object motion classification is 100%.

From the results of the success rates of the classifications based on the individual parameters, it can be stated that if just the mean, standard deviation or skewness values are used in the feature set, the success rate rapidly decreases relative to the success rate of the classification based on the feature set which includes all the parameters. These success rates become approximately 50%, which is the expected success rate of a random classification. On the other hand, the success rate of the classification based on the kurtosis values is around 60%, which is higher than the cases of the other parameters. That means, in our dataset, the separation between matte and shiny objects in terms of their kurtosis values of the subband histograms is more obvious than the mean, standard deviation and skewness values. Another important observation about these classifications can be made as, in the skewness case, most of the matte objects are classified as a shiny object. The reason of this misclassification may be that the skewness of the moving shiny objects may be more stationary than the skewness of the matte objects. In other words, the skewness values which belong to matte objects show more variation than the skewness of the shiny objects. And, if the most of the matte objects in the test set resemble to shiny objects in terms of their subband histogram skewnesses, they are classified as shiny.

As a result of these classification success rates, the highest classification success is observed when all the statistical parameters are used in the feature set.

3.3 Motion Synthesis

In the previous section, we give some statistical differences between motions of matte and shiny objects. In order to understand to what extent the motion subband statistics are important on the surface reflectance recognition, we synthesized new movies. These movies are aimed to have the same subband statistics with a source movie, but it looks completely different from that movie. The source movie is either the motion of a matte object or a shiny object. We give the method of motion synthesis and results in the later stages of this section.

In order to generate a synthethic motion, we used texture synthesis algorithm proposed in [55]. Although that algorithm was proposed for two-dimensional textures, we adapted it for three dimensional motion synthesis. The aim of the algorithm is to modify a random noise movie such that its subband histograms are the same as subband histograms of a chosen source movie. The steps of the algorithm can be summarized as follows;

- **Preprocessing Step:** There are two accomplishments in this step. Firstly, a noise movie is generated by assigning independent random values to the video pixels. Secondly, the source movie is decomposed by the steerable pyramid which is mentioned in Section 2.4. Then, the subband histograms are extracted from the source movie subbands and saved for the later usage.

- **Step 1:** The histogram of the noise is equalized to the histogram of the source movie. This process is called as *histogram equalization* and it is based on equating the cumulative distribution functions (CDF) of two histograms. Detailed information about histogram equalization can be found in [77].

- **Step 2:** The noise movie is decomposed into their subbands by using the steerable pyramid.

- **Step 3:** In each subband, the histograms of the noise are equalized to the histograms of the source movie.
- **Step 4:** The histogram equalized subbands of the noise movie are used in the reconstruction of the modified noise.
- **Step 5:** It is went back to Step 1 and continued until a desired number of iterations reached.

In our case, we decided to have 5 iterations, since, if the iteration number increases, the artifacts in the resulting image also increase due to reconstruction errors.

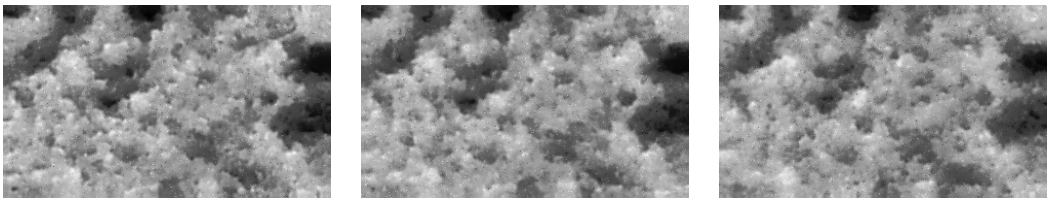
3.3.1 Test of the Success of the Algorithm on Moving Textures

In [55], it is shown that this algorithm is useful if source image does not have a dominant orientation and it has a homogeneous structure. If this conditions are met, although there are some failure examples, synthesized texture looks like similar to source image. In order to test the algorithm for three dimensional images, we first synthesized a texture from the motion of clouds which has isotropic features both in space and time. Since we designed two different steerable pyramids, we give the results for each. In Figure 3.7, six frames of the real cloud motion are shown in gray scales. In Figure 3.8 and Figure 3.9 six frames of the synthesized texture are also shown for each steerable pyramid.

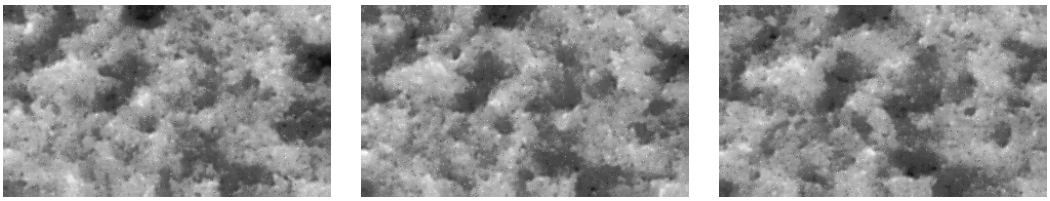
(a) Frame 1 (b) Frame 2 (c) Frame 3

(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.7: Six frames from a motion of clouds. This is a test movie for the texture synthesis algorithm.

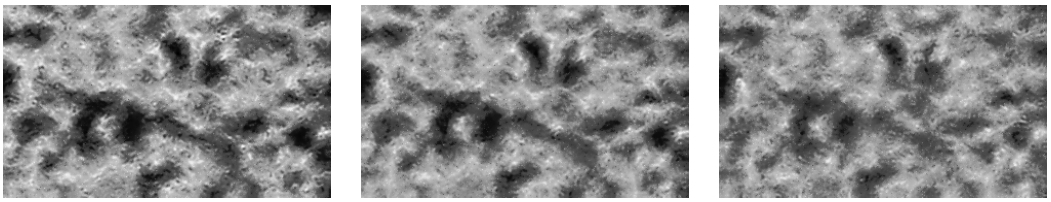


(a) Frame 1 (b) Frame 2 (c) Frame 3

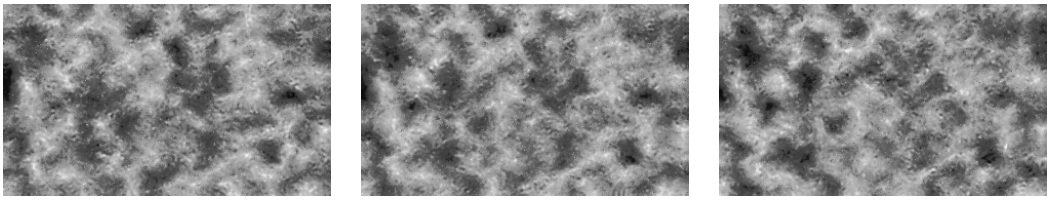


(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.8: Six frames from the synthesized texture by using the steerable pyramid mentioned in Method I in Section 2.4.3.



(a) Frame 1 (b) Frame 2 (c) Frame 3



(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.9: Six frames from the synthesized texture by using the steerable pyramid mentioned in Method II in Section 2.4.3.

As it can be seen from the figures, although there are some artifacts, the synthesized textures are similar to the source movie for the cloud motion, especially the ones synthesized using the steerable pyramid whose bandpass filters are the first derivative of Gaussian filters. One of the most different parts is the sense of depth between the source movie and the synthesized movies.

The movies that we used for surface reflectance recognition are not homogenous both in time and space. Therefore, we do not expect that the synthesized textures resemble the source movies in this particular feature. However, although the spatial structure of the objects are not captured, the synthesized dynamic textures have the same motion direction with the motion direction of their corresponding source movies. In order to test the success of the algorithm, we developed a classification algorithm which resembles to the one mentioned in Section 3.2.3. In order to apply the classification algorithm, we first decomposed each synthesized texture into its subbands by using the second derivative of Gaussian filters, then extracted the mean, standard deviation, skewness and kurtosis from the subband histograms and used all of them in the feature space. In this version of the classification algorithm, we again randomly selected 14

motions of matte objects and 14 motions of shiny objects. Then, we used all the synthesized movies in the test set. Since the aim of this algorithm is to whether the synthesized textures resemble to the source movies in terms of their subband statistics, we check whether the success rate is to 100%. The results of the classification can be summarized as follows:

The classification results for 1000 run of the SVM algorithm on synthesized textures by the steerable pyramid mentioned in Method I in 2.4 are as follows;

- The average success rate is 54%.
- The standard deviation of the success rates is 4.
- The minimum success rate is 40%.
- The maximum success rate is 63%.
- Confusion matrix for average success rate is;

		Estimated Label	
		Matte	Shiny
True Label	Matte	25	75
	Shiny	17	83

- The standard deviation of the success rates of matte object motion classification is 8.
- The minimum success rate for matte object motion classification is 5%.
- The maximum success rate for matte object motion classification is 55%.
- The standard deviation of the success rates of shiny object motion classification is 14.
- The minimum success rate for shiny object motion classification is 35%.
- The maximum success rate for shiny object motion classification is 100%.

The classification results for 1000 run of the SVM algorithm on synthesized textures by the steerable pyramid mentioned in Method II in 2.4 are as follows;

- The average success rate is 54%.
- The standard deviation of the success rates is 3.
- The minimum success rate is 45%.
- The maximum success rate is 68%.
- Confusion matrix for average success rate is;

		Estimated Label	
		Matte	Shiny
True Label	Matte	17	83
	Shiny	9	91

- The standard deviation of the success rates of matte object motion classification is 10.
- The minimum success rate for matte object motion classification is 0%.
- The maximum success rate for matte object motion classification is 55%.
- The standard deviation of the success rates of shiny object motion classification is 7.
- The minimum success rate for shiny object motion classification is 55%.
- The maximum success rate for shiny object motion classification is 100%.

From the results given above, the success rates for both steerable pyramids are around 50%. However, the main point here is that, the shiny objects are classified with a high success rate, which is around 90%. On the other hand, the matte objects are classified with a low success rate, which is around 20%. These

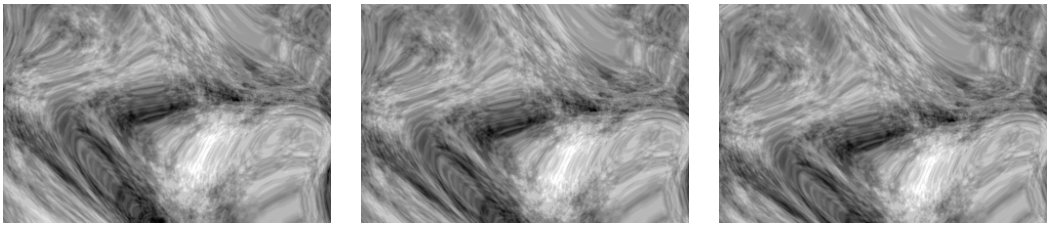
results indicate that, in terms of subband statistics, the synthesized textures, regardless of their source movies, resemble more to the shiny object motions than the matte object motions. Therefore, the synthesized textures can be used in a subjective test which checks whether they resemble to matte or shiny one. We give the results of this experiment below.

3.3.2 Results of Motion Texture Synthesis

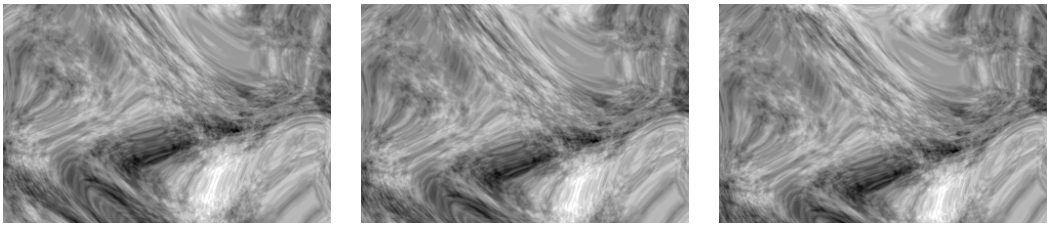
Since we synthesized two dynamic textures for each motion. One movie is synthesized by using the steerable pyramid mentioned in Method I and the other is synthesized by using the steerable pyramid mentioned in Method II in 2.4. Therefore, there are 80 different synthetic dynamic textures. As we stated before, most of the synthesized textures resemble more the shiny object movies in terms of the subband statistics. Therefore, in this section, we qualitatively checked whether a human observer judged the synthesized textures to be more similar to shiny than matte. The main result that can be inferred from synthesized textures is that, although synthesized textures capture some of the properties of their corresponding source movies, they do not appear to provide sufficient information to make a decision whether the synthesized textures resemble a moving shiny or a moving matte object.

The detailed information about the synthesized textures can be summarized as follows. First of all, synthesized textures move in the same direction with their corresponding source movies. Secondly, the speed of the textures also looks similar to speed of the source movie textures. In addition, since the bandwidth of the bandpass regions of the steerable pyramid designed with Method II is narrower in terms of orientation, the textures synthesized with this steerable pyramid looks more articulated than the ones which are synthesized by using the steerable pyramid designed with Method I. In Figures 3.10, 3.11, 3.12, 3.13, 3.14

and 3.15, an example motion pair and their corresponding synthesized textures are shown.

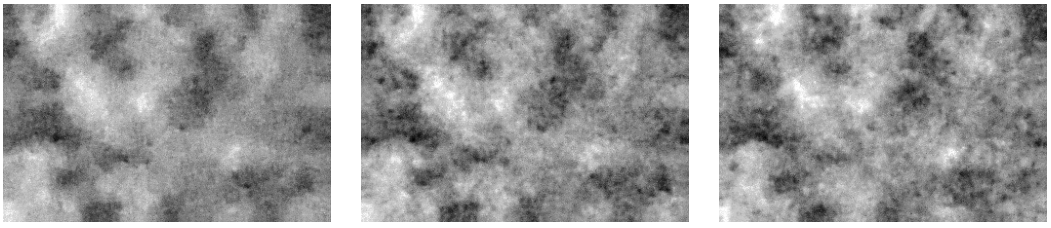


(a) Frame 1 (b) Frame 2 (c) Frame 3

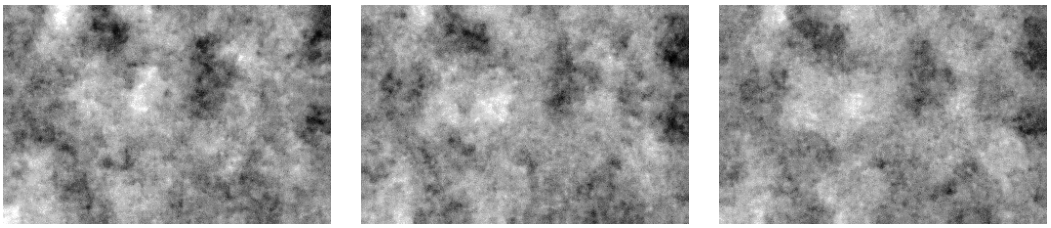


(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.10: Six frames from the source moving matte object.

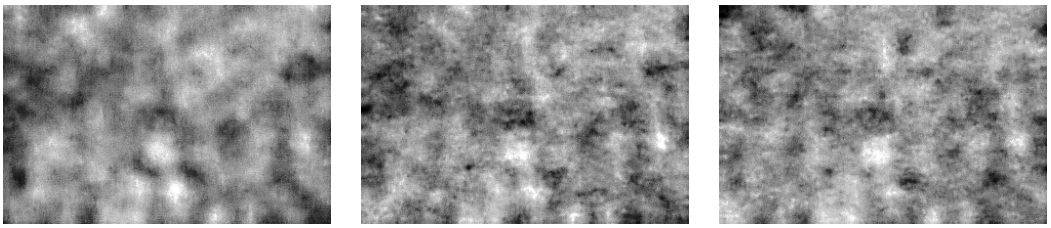


(a) Frame 1 (b) Frame 2 (c) Frame 3

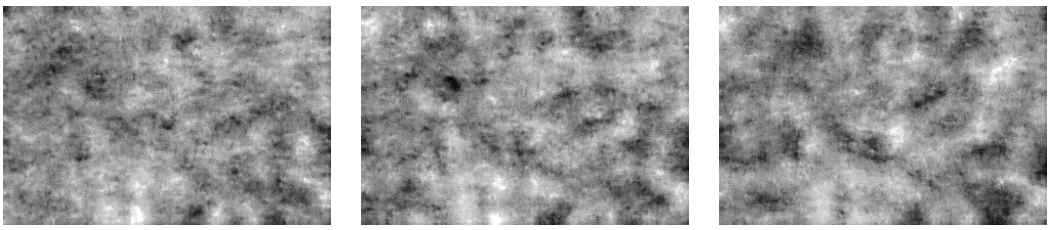


(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.11: Six frames from synthesized texture whose source motion is shown in Figure 3.10. This texture is synthesized by using the steerable pyramid mentioned in Method I in Section 2.4.3.

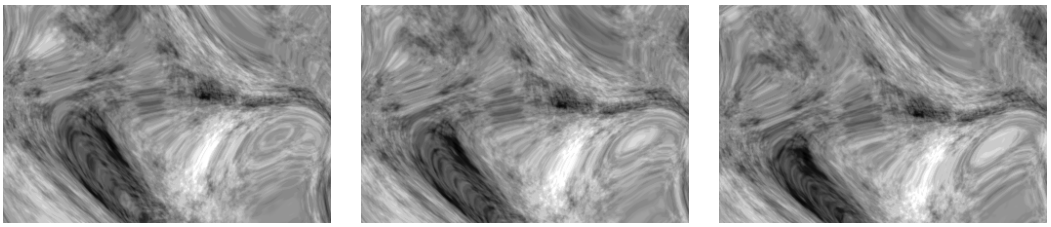


(a) Frame 1 (b) Frame 2 (c) Frame 3

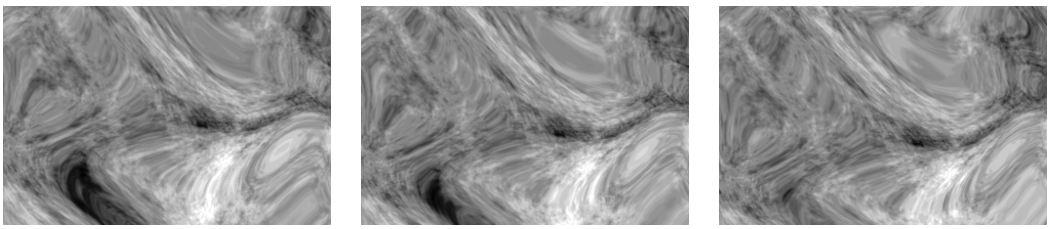


(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.12: Six frames from the synthesized texture whose source motion is shown in Figure 3.10. This texture is synthesized by using the steerable pyramid mentioned in Method II in Section 2.4.3.

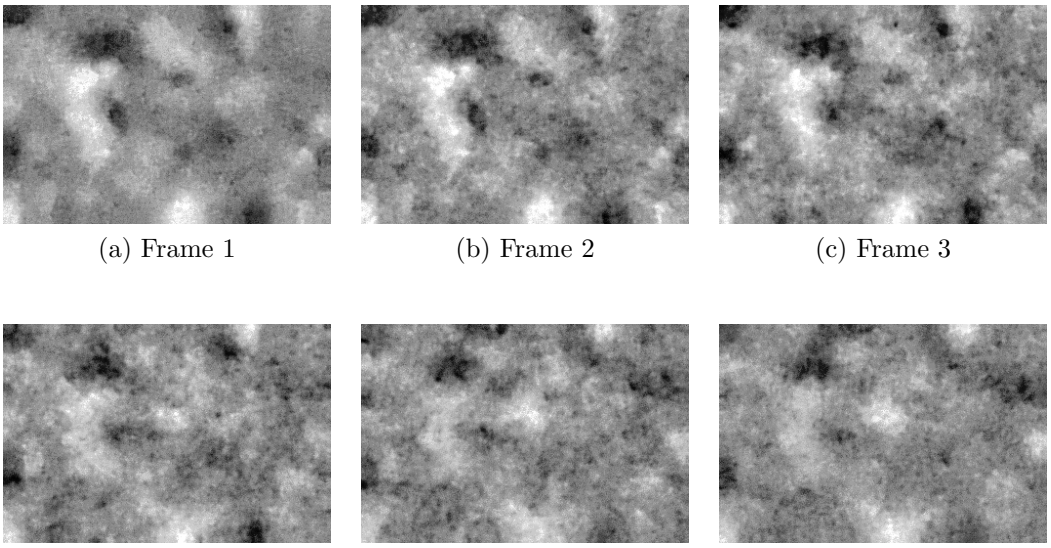


(a) Frame 1 (b) Frame 2 (c) Frame 3



(d) Frame 4 (e) Frame 5 (f) Frame 6

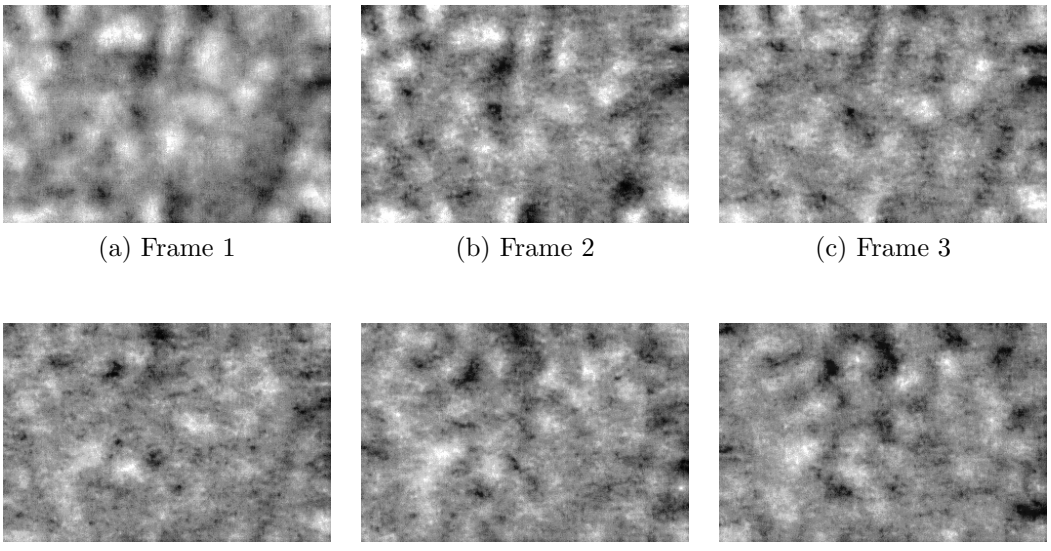
Figure 3.13: Six frames from a source moving shiny object which has the same shape as of the object in Figure 3.10.



(a) Frame 1 (b) Frame 2 (c) Frame 3

(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.14: Six frames from synthesized texture whose source motion is shown in Figure 3.13. This texture is synthesized by using the steerable pyramid mentioned in Method I in Section 2.4.3.



(a) Frame 1 (b) Frame 2 (c) Frame 3

(d) Frame 4 (e) Frame 5 (f) Frame 6

Figure 3.15: Six frames from the synthesized texture whose source motion is shown in Figure 3.13. This texture is synthesized by using the steerable pyramid mentioned in Method II in Section 2.4.3.

In addition to these movies, some other movie and synthesized texture examples can be found in `http://www.umram.bilkent.edu.tr/~kulce/`.

One of the possible reasons that the synthesized textures do not give the sense of the reflectance is that the algorithm does not capture any information about the object shape. We already saw that shape information is crucial for judging surface glossines of static objects [38]. Since specular features slide on the surface of a shiny object, perceiving no object in a movie may somehow interfere with the recognition of specularity. In addition, our subband statistics were not an exact match due to overlap in the frequency domain in our filterset. Using a better algorithm that synthesizes textures whose subband statistics are exactly the same as the source movies may lead to a better classification of matte and shiny synthesized textures.

In this chapter, we provided indices to the importance of subband statistics on the perception of the surface specularity. In the begining of the chapter, we gave details on the dataset and analytical definitions of the four statistical parameters, mean, standard deviation, skewness and kurtosis. Then, we explained the numerical results, starting from the comparison of matte and shiny object motion pairs. In comparison of the pairs, we highlighted some properties that are in common in the motion pairs. In the second phase of the numerical results section, we compute the average statistics of the matte and shiny objects. Then, we developed a machine learning algorithm from two different classifiers. In both of the methods, we got approximately 75% success rate on the classification of matte and shiny objects. Finally, we synthesized dynamic texture, to see whether the subband statistics are the sufficient characteristics on the perception of the reflectance, and we see that there should be some other characteristics that should be taken into account to succefully predict surface reflectance perception.

Chapter 4

SUMMARY AND CONCLUSIONS

In this thesis, we investigated the role of subband motion statistics on surface reflectance appearance. Our results indicate that, although the subband statistics can provide cues to surface reflectance, they do not capture all the required information necessary for the identification of surface material properties.

Summary: The work we have done for this thesis can be grouped under two main titles:

- Technical details of three dimensional filters used for subband decomposition/reconstruction.
- Examination of the first-order subband motion statistics of moving matte and shiny objects.

After introducing the fundamental concepts that provide a basis for the rest of the thesis in Chapter 1, we describe the technical details in Chapter 2. We

first explain *orientation* and *scale* concepts. We, then, give their interpretations in the frequency domain. Next, we introduce steerable filter concept, which is a method of implementing an oriented filter as a linear sum of differently oriented filters. After that, since we used second derivative of Gaussian filters as the orientation and scale selective filters, we first provide a few properties of the derivative of Gaussian filters, such as steerability, orientation and scale selectivity characteristics and design considerations. Then, we describe the explicit design steps of the second derivative of Gaussian filters in three dimensions. These filters are designed in three scales and eight orientations that extract total of 24 subbands. The filter coefficients of these filters can be found in Appendix B. In the last part of the chapter, we introduce a video decomposition/reconstruction technique, which is called the *steerable pyramid*. After that, we describe our adopted filter design techniques for three-dimensional steerable pyramid. In order to provide perfect reconstruction in the orientation space, we used the technique in [72] and in order to provide perfect reconstruction in scale space we used two different methods; **1)** First derivative of Gaussian filters, **2)** Designing the low pass filter, which has broader bandwidth than bandwidth of the other low pass filter in the filterbank, by applying the technique given in [74] and subsequent design of the remaining filters in the steerable pyramid filterbank.

In Chapter 3, we introduce the dataset and mathematical definitions of the statistical parameters. These parameters are the mean, standard deviation, skewness and kurtosis. All these parameters are extracted from the subband histograms of the movies. Then, we compare statistics of each motion pair which contains one matte and one shiny version of an object. According to the results, means and kurtosises of the motion subbands do not differ much within the pairs. On the other hand, in many subband histograms, the standart deviation of most of the matte object motions is higher than their shiny counterparts. Whereas, the skewness of the subband histograms of matte object motions is generally

higher than in its shiny counterpart in most of the subbands. Moreover, we also compared the averages of the statistical parameters which are extracted from the subband histograms of 20 moving matte objects and the statistical parameters which are extracted from the subband histograms of 20 moving shiny objects. We observed that the standard deviation and skewness values in many subbands of the moving matte objects generally differ from the moving shiny objects. The average standard deviations of the subband histograms of the matte objects are larger than their shiny counterparts and the skewness shows the opposite behaviour, that is the skewness values are higher for the subband histograms of shiny object motions in many subbands. Then, we developed a classifier algorithm based on those statistical parameters. We used the SVM as the classification algorithm. The results of the classification showed that, nearly 75% of all movies are classified correctly. The correct classification of shiny movies is around 77% and of matte movies is around 71%. In addition, we classified the movies just based on one statistical parameter. In the mean, standard deviation and skewness cases, the success rate is nearly 50%. On the other hand, the success rate in the classification based on the kurtosis values increases up to 60%. Finally, we synthesized artificial reflectance videos by using the algorithm proposed in [55]. From a classification algorithm applied on these new videos, we found that most of the videos were classified as shiny. Visual inspection, however, indicates that it is not possible to make an accurate estimation about the resemblance of the synthesized movies as either matte or shiny textures.

Comments: According to the results which are obtained from our dataset, the hypothesis that states the statistics of a visual stimuli play role on perception and identification is partially validated on the surface reflectance recognition from movies. One of the strongest findings that have us reach such a conclusion is that the differences between the standard deviation and skewness values in the video subbands. These results indicate that in many subbands the standard

deviation of matte object motions is higher than their shiny counterparts. That means, the outputs of the filters have more intensity contrasts in the matte case. Moreover, since the skewness values are generally higher in the shiny case, it can be said that, the histograms of the filter outputs have longer tails towards brighther intensity than darker intensity values. In addition, although it is not obvious as in the standard deviation and skewness cases, the kurtosis of the subband histograms of the moving shiny objects is generally higher than the matte case. Since the kurtosis is a measure of the peakedness of a histogram, it can be deduced that the intensity values in the shiny object motion subbands are accumulated around a certain value.

Another finding in this thesis is the success rate of the classification of the movies based on four statistical parameters extracted from the subband histograms. Although we have a small data set, we got a 75% success on the classification of all movies. Therefore, this can be another cue that the statistics are important on the perception of the shininess. These were the results that supports the hyphothesis that we mentioned at the beginning of this paragraph. However, classification results on the synthesized textures suggests otherwise. In terms of first order subband statistics, a large number of the synthesized textures resemble shiny moving objects, regardless whether they were originally derived from shiny or matte. However, most human observers would not judge the synthesized textures to look shiny, in fact, they find it hard to make any comment on the resemblance of the synthesized textures to either matte or shiny objects. This may be beause the synthesized motions do not give the feeling of objectness. In other words, a solid shape cannot be perceivied and as noted before, the motion of the specularities on a shiny object surface is related to the object shape. They move faster than the object surface on the flat regions, slower on the convexly curved regions and tend to move towards convexities [30]. In the synthesized textures, this information is absent. The crucial implication is that

for the HVS the estimation of 3D shape and reflectance may be coopearative, i.e. not independent.

Future Work:

- The work can be repeated over a larger dataset. The reason is that, the number of the sample movies is relatively small when it is compared to the dimension of the feature sets which are used in the classifications. The dataset can include both computer generated and real world examples. Moreover, the dataset does not have to include just the solid shapes. For example, in order to investigate the shiny features on a liquid, motion of water waves can be included.

- In our dataset, the objects are rotating in a fixed rectangle within the movie frames. Thus, we could cut out the background from the movies easily. However, if our classification algorithm is going to be used in the classification of a motion captured from a real world environment, just cutting out the background may not be possible. The reason is that, in the real scenes, the objects generally move unevenly within the scenes. Therefore, applying a background extraction algorithm to the movies before classification can be useful.

- We worked on the first order statistics which are extracted from the subband histograms. However, second or higher order statistics or Markov models may also have the influence on the perception of the surface reflectance. As an intuition, these statistics may explain the relationship between the motion of specular features and the motion of object surface.

- We saw that the synthesis algorithm that we used is not good enough for our aim, which is to synthesize a movie whose subbands have exactly the same histograms with a particular source movie. Therefore, another

algorithm that provides synthesizing a movie whose only similarity to a source movie is the first order statistics can be developed or the existing algorithm can be modified.

- In order to see the response of the early stages of the HVS to a visual stimuli which includes motion of a matte or shiny object, functional magnetic resonance imaging (fMRI) experiments can be performed.

APPENDIX A

PROOF OF STATEMENT I

Let the cartesian coordinates of the frequency components of an arbitrary cosine signal be $\boldsymbol{\omega_x} = [\omega_x\ \omega_y\ \omega_t]^T$ and the cartesian coordinates of the directed frequency points of a filter be $\boldsymbol{\omega_{x_i}} = [\omega_{x_i}\ \omega_{y_i}\ \omega_{t_i}]^T$, where i is from 1 to P. We assume that the frequency points are on the unit sphere. That is $\omega_x^2+\omega_y^2+\omega_t^2 = \omega_{x_i}^2+\omega_{y_i}^2+\omega_{t_i}^2 = 1$. However, even if the frequency points are not on the unit sphere, the proof is still valid (The only thing that should be added to the proof is a normalization constant, which does not have an effect on the proof). Let $\boldsymbol{R}$ be the rotation matrix defined in Equation 2.10. The coordinate axis points $\boldsymbol{\omega_{x_i}}$ can be rotated to new coordinate axes, $\boldsymbol{\omega_{x_i'}} = [\omega_{x_i'}\ \omega_{y_i'}\ \omega_{t_i'}]^T$ as;

$$\boldsymbol{\omega_{x_i'}} \quad = \quad \boldsymbol{R\omega_{x_i}} \tag{A.1}$$

The cosine of the angle, α, between the vectors whose elements are the cartesian coordinates of a point can be calculated through inner product of two vectors. In the same way, $\cos^2(\alpha_i)$, can be written as multiplication of two inner products;

$$\cos(\alpha_i)^2 = \left(\boldsymbol{\omega_x}^T \boldsymbol{\omega_{x'_i}}\right)\left(\boldsymbol{\omega_{x'_i}}^T \boldsymbol{\omega_x}\right) \tag{A.2}$$

$$= \boldsymbol{\omega_x}^T \left(\boldsymbol{\omega_{x'_i}} \boldsymbol{\omega_{x'_i}}^T\right) \boldsymbol{\omega_x} \tag{A.3}$$

The sum of the square of the cosine functions can be written as;

$$\sum_{i=1}^{P} \cos(\alpha_i)^2 = \sum_{i=1}^{P} \left[\boldsymbol{\omega_x}^T \left(\boldsymbol{\omega_{x'_i}} \boldsymbol{\omega_{x'_i}}^T\right) \boldsymbol{\omega_x}^T\right] \tag{A.4}$$

$$= \boldsymbol{\omega_x}^T \left(\sum_{i=1}^{P} \boldsymbol{\omega_{x'_i}} \boldsymbol{\omega_{x'_i}}^T\right) \boldsymbol{\omega_x} \tag{A.5}$$

$$= \begin{bmatrix}\omega_x & \omega_y & \omega_t\end{bmatrix} \begin{bmatrix} \sum_{i=1}^{P} \omega_{x'_i}^2 & \sum_{i=1}^{P} \omega_{x'_i}\omega_{y'_i} & \sum_{i=1}^{P} \omega_{x'_i}\omega_{t'_i} \\ \sum_{i=1}^{P} \omega_{x'_i}\omega_{y'_i} & \sum_{i=1}^{P} \omega_{y'_i}^2 & \sum_{i=1}^{P} \omega_{y'_i}\omega_{t'_i} \\ \sum_{i=1}^{P} \omega_{x'_i}\omega_{t'_i} & \sum_{i=1}^{P} \omega_{y'_i}\omega_{t'_i} & \sum_{i=1}^{P} \omega_{t'_i}^2 \end{bmatrix} \cdot \begin{bmatrix}\omega_x \\ \omega_y \\ \omega_t\end{bmatrix} \tag{A.6}$$

In order to have a constant frequency response for all signals, we want to have sum of the square of the cosines independent from ω_x, ω_y and ω_t. If the cartesian coordinates of frequency points, $\omega_{x'_i}$, $\omega_{y'_i}$, $\omega_{t'_i}$, are selected such that, the matrix in the middle in Equation A.6 is the identity matrix times a scalar constant, that is cI, $\sum_{i=1}^{P} \cos(\alpha_i)^2$ becomes equal to $c \cdot (\omega_x^2 + \omega_y^2 + \omega_t^2) = c$. Any point set which satisfies this equality can be used. It can also be seen that, if $\boldsymbol{\omega_{x'_i}}$ equals the cartesian coordinates of the vertices of the cube or the octahedron, which are given in Appendix C, the matrix in the middle in Equation A.6 becomes an identity matrix times a constant. Finally, since we rotated the coordinate system at the beginning, any set of points that can be rotated to cartesian coordinates of the vertices of the cube or the octahedron can be used.

APPENDIX B

FILTER COEFFICIENTS

The filter coefficients are given in the below table. Each column gives 1D spatial filter coefficients of the functions $F_0(\omega) = e^{-\frac{\sigma^2}{2}\omega^2}$, $F_1(\omega) = j\omega e^{-\frac{\sigma^2}{2}\omega^2}$ and $F_2(\omega) = \omega^2 e^{-\frac{\sigma^2}{2}\omega^2}$. The numbers in the first column indicate filter coefficient index. Since filter coefficients of $F_0(\omega)$ and $F_2(\omega)$ are even symmetric and coefficients of $F_1(\omega)$ are odd symmetric with respect to tap number 0, negative index coefficients are not shown. The coefficients are calculated according to the IDFT equation;

$$f_i(t) = \frac{1}{M} \sum_{k=-(\lfloor \frac{M-1}{2} \rfloor)}^{\lceil \frac{M-1}{2} \rceil} F_i(k\frac{2\pi}{M}) e^{j\frac{2\pi kt}{M}}, \tag{B.1}$$

where, $i = 0, 1$ or 2, $t \in \mathbb{Z}$ and $-(\lfloor \frac{M-1}{2} \rfloor) \leqslant t \leqslant (\lceil \frac{M-1}{2} \rceil)$ and M is the filter length. A generic table that shows placement of coefficients is as follows;

	σ		
	$F_0(\omega)$	$F_1(\omega)$	$F_2(\omega)$
0	$f_0(0)$	$f_1(0)$	$f_2(0)$
1	$f_0(1)$	$f_1(1)$	$f_2(1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\frac{M-1}{2}$	$f_0(\frac{M-1}{2})$	$f_1(\frac{M-1}{2})$	$f_2(\frac{M-1}{2})$

	$\sigma = 1$			$\sigma = 1.8$			$\sigma = 3.2$		
	$F_0(\omega)$	$F_1(\omega)$	$F_2(\omega)$	$F_0(\omega)$	$F_1(\omega)$	$F_2(\omega)$	$F_0(\omega)$	$F_1(\omega)$	$F_2(\omega)$
0	0.3984	0	0.3922	0.2216	0	0.0684	0.1247	0	0.0122
1	0.2425	-0.2427	0.006	0.1899	-0.0586	0.0405	0.1187	-0.0116	0.0105
2	0.0536	-0.1067	-0.1663	0.1196	-0.0738	-0.0087	0.1026	-0.02	0.0061
3	0.0046	-0.0148	-0.033	0.0553	-0.0512	-0.0303	0.0803	-0.0235	0.001
4	0.0001	0.0011	-0.0028	0.0188	-0.0232	-0.0228	0.0571	-0.0223	-0.0031
5				0.0047	-0.0072	-0.0097	0.0368	-0.018	-0.0052
6				0.0008	-0.0016	-0.0027	0.0215	-0.0126	-0.0053
7				0.0001	-0.0002	-0.0006	0.0114	-0.0078	-0.0042
8							0.0055	-0.0043	-0.0028
9							0.0024	-0.0021	-0.0016
10							0.0009	-0.0009	-0.0008
11							0.0003	-0.0004	-0.0004
12							0.0001	-0.0001	-0.0002

	$\sigma = 0.8$	$\sigma = 1.6$
	$F_0(\omega)$	$F_0(\omega)$
0	0.4936	0.2493
1	0.2326	0.2051
2	0.0194	0.1142
3	0.0012	0.0432
4		0.0128

The nonseparable spatio-temporal filter kernel, $h(x, y, t)$ of the high pass filter $\sqrt{1 - e^{-0.8^2 \cdot (\omega_x^2 + \omega_y^2 + \omega_t^2)}}$ is shown below. Since the filter is three dimensional, in order to write it here, we separate the filter coefficients according to their time index. If a filter kernel is shown under a label $h(0:3, 0:3, t)$, the filter kernel is for from 0^{th} index to 3^{rd} index for both spatial axes of the t^{th} time index. Since

the coefficients are symmetric with respect to origin, we give the kernels only for positive time indexes. Moreover, we only give the first quadrant of the filter kernels, however, since the filter coefficients are symmetric with respect to origin, $x = 0$ and $y = 0$ axes, they can be extrapolated in a straightforward manner to have a size of $9 \times 9 \times 9$.

$H(0:3, 0:3, 0) =$

	y_0	y_1	y_2	y_3
x_0	0.9745	-0.0180	-0.0068	-0.0023
x_1	-0.0180	-0.0128	-0.0051	-0.0020
x_2	-0.0068	-0.0051	-0.0025	-0.0014
x_3	-0.0023	-0.0020	-0.0014	-0.0011

$H(0:3, 0:3, 1) =$

	y_0	y_1	y_2	y_3
x_0	-0.0180	-0.0128	-0.0051	-0.0020
x_1	-0.0128	-0.0093	-0.0039	-0.0017
x_2	-0.0051	-0.0039	-0.0021	-0.0013
x_3	-0.0020	-0.0017	-0.0013	-0.0010

$H(0:3, 0:3, 2) =$

	y_0	y_1	y_2	y_3
x_0	-0.0068	-0.0051	-0.0025	-0.0014
x_1	-0.0051	-0.0039	-0.0021	-0.0013
x_2	-0.0025	-0.0021	-0.0014	-0.0011
x_3	-0.0014	-0.0013	-0.0011	-0.0009

$H(0:3, 0:3, 3) =$

	y_0	y_1	y_2	y_3
x_0	-0.0023	-0.0020	-0.0014	-0.0011
x_1	-0.0020	-0.0017	-0.0013	-0.0010
x_2	-0.0014	-0.0013	-0.0011	-0.0009
x_3	-0.0011	-0.0010	-0.0009	-0.0009

APPENDIX C

CARTESIAN AND SPHERICAL COORDINATES OF VERTICES OF PLATONIC SOLIDS

The coordinate points presented in this part are taken from [73]. The cartesian coordinates, $C(x, y, t)$, of the vertices of the cube;

	C_1	C_2	C_3	C_4
x	1	1	-1	-1
y	1	-1	-1	1
t	1	1	1	1

The spherical coordinates, $S(\theta, \phi)$, of the vertices of the cube in radians;

	S_1	S_2	S_3	S_4
θ	$\frac{\pi}{4}$	$\frac{7\pi}{4}$	$\frac{5\pi}{4}$	$\frac{3\pi}{4}$
ϕ	0.3041π	0.3041π	0.3041π	0.3041π

The cartesian coordinates, $C(x, y, t)$, of the vertices of the octahedron;

	C_1	C_2	C_3
x	1	0	0
y	0	1	0
t	0	0	1

The spherical coordinates, $S(\theta, \phi)$, of the vertices of the octahedron in radians;

	S_1	S_2	S_3
θ	0	$\frac{\pi}{2}$	Not Important
ϕ	$\frac{\pi}{2}$	$\frac{\pi}{2}$	0

The cartesian coordinates, $C(x, y, t)$, of the vertices of the dodecahedron; ($g = (1 + \sqrt{5})/2$);

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
x	1	-1	1	-1	1/g	-1/g	g	$-$g	0	0
y	1	1	-1	-1	0	0	1/g	1/g	g	$-$g
t	1	1	1	1	g	g	0	0	1/g	1/g

The spherical coordinates, $S(\theta, \phi)$, of the vertices of the dodecahedron in radians;

	S_1	S_2	S_3	S_4	S_5
θ	$\frac{\pi}{4}$	$\frac{3\pi}{4}$	$\frac{7\pi}{4}$	$\frac{5\pi}{4}$	0
ϕ	0.3041π	0.3041π	0.3041π	0.3041π	0.116π

	S_6	S_7	S_8	S_9	S_{10}
θ	π	0.116π	0.8839π	$\frac{\pi}{2}$	$\frac{3\pi}{2}$
ϕ	0.116π	$\frac{\pi}{2}$	$\frac{\pi}{2}$	0.3839π	0.3839π

The cartesian coordinates, $C(x, y, t)$, of the vertices of the icosahedron;

	C_1	C_2	C_3	C_4	C_5	C_6
x	0	0	1	1	g	$-$g
y	1	-1	g	$-$g	0	0
t	g	g	0	0	1	1

The spherical coordinates, $S(\theta, \phi)$, of the vertices of the icosahedron in radians;

	S_1	S_2	S_3	S_4	S_5	S_6
θ	$\frac{\pi}{2}$	$\frac{3\pi}{2}$	0.3238π	1.1762π	0	π
ϕ	0.1762π	0.1762π	$\frac{\pi}{2}$	$\frac{\pi}{2}$	0.3238π	0.3238π

Bibliography

[1] D. H. HUBEL and T. N. WIESEL, "Receptive fields of single neurones in the cat's striate cortex.," *The Journal of physiology*, vol. 148, pp. 574–591, Oct. 1959.

[2] E. T. Rolls, Martin, and J. Tovee, "Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex," *Journal of Neurophysiology*, vol. 73, pp. 713–726, 1995.

[3] E. Philip and A. Laplante, *Electrical Engineering Dictionary*. CRC Press LLC, 2000.

[4] R. Dror, *Surface reflectance recognition and real-world illumination statistics*. PhD thesis, Massachusetts Institute of Technology. Dept. of Electrical Engineering and Computer Science, 2003.

[5] B. K. P. Horn, "Shape from shading," ch. Obtaining shape from shading information, pp. 123–171, Cambridge, MA, USA: MIT Press, 1989.

[6] J. J. Koenderink and A. J. van Doorn, "Shape from shading," ch. Photometric invariants related to solid shape, pp. 301–321, Cambridge, MA, USA: MIT Press, 1989.

[7] A. Pentland, "Shape information from shading: A theory about human perception," in *Computer Vision., Second International Conference on*, pp. 404 –413, dec 1988.

[8] I. Ihrke, K. N. Kutulakos, H. P. A. Lensch, M. Magnor, and W. Heidrich, "State of the art in transparent and specular object reconstruction," 2008.

[9] F. Attneave, "Some informational aspects of visual perception," *Psychological Review*, vol. 61, pp. 183–193, May 1954.

[10] H. Barlow, "Possible principles underlying the transformation of sensory messages," *Sensory Communication*, pp. 217–234, 1961.

[11] W. S. Geisler, "Visual Perception and the Statistical Properties of Natural Scenes," *Annual Review of Psychology*, vol. 59, no. 1, pp. 167–192, 2008.

[12] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, pp. 1193–1216, 2001.

[13] B. A. Olshausen, "Principles of image representation in visual cortex," *The Visual Neurosciences*, p. 1603–1615, 2003.

[14] B. A. Olshausen and D. J. Field, "Vision and the coding of natural images," *American Scientist*, vol. 88, no. 3, pp. 238–245, 2000.

[15] H. Barlow, "Redundancy reduction revisited.," *Network*, vol. 12, pp. 241–253, Aug. 2001.

[16] S. Zeki, "Cerebral akinetopsia (visual motion blindness). a review.," *Brain*, vol. 114 (Pt 2), pp. 811–24, 1991.

[17] D. Kersten, H. H. Bülthoff, B. L. Schwartz, and K. J. Kurtz, "Interaction between transparency and structure from motion," *Neural Computation*, vol. 4, no. 4, pp. 573–589, 1992.

[18] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *Int. J. Comput. Vision*, vol. 9, pp. 137–154, Nov. 1992.

[19] C. J. Poelman and T. Kanade, "A paraperspective factorization method for shape and motion recovery," in *Proceedings of the Third European Conference-Volume II on Computer Vision - Volume II*, ECCV '94, (London, UK, UK), pp. 97–108, Springer-Verlag, 1994.

[20] D. D. Morris and T. Kanade, "A unified factorization algorithm for points, line segments and planes with uncertainty models," in *Proceedings of the Sixth International Conference on Computer Vision*, ICCV '98, (Washington, DC, USA), pp. 696–, IEEE Computer Society, 1998.

[21] D. Forsyth, S. Ioffe, and J. Haddon, "Bayesian structure from motion," 1999.

[22] D. P. Huttenlocher and S. Ullman, "Recognizing solid objects by alignment with an image," *Int. J. Comput. Vision*, vol. 5, pp. 195–212, Nov. 1990.

[23] D. G. Lowe, "Fitting parameterized three-dimensional models to images," *IEEE Transactions on Pattern Aanalysis and Machine Intelligence*, vol. 13, pp. 441–450, 1991.

[24] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun, "Structure from motion without correspondence," in *In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR*, pp. 557–564, 2000.

[25] R. Blake and M. Shiffrar, "Perception of human motion," *Annual Review of Psychology*, vol. 58, no. 1, pp. 47–73, 2007.

[26] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception & Psychophysics*, vol. 14, pp. 201–211, 1973.

[27] E. D. Grossman and R. Blake, "Brain areas active during visual perception of biological motion," *Neuron*, vol. 35, no. 6, pp. 1167 – 1175, 2002.

[28] B. Hartung and D. Kersten, "Distinguishing shiny from matte," *Journal of Vision*, vol. 2, no. 7, p. 551, 2002.

[29] K. Doerschner, R. W. Fleming, O. Yilmaz, P. R. Schrater, B. Hartung, and D. Kersten, "Visual motion and the perception of surface material," *Curr Biol*, vol. 21, pp. 2010–2016, Dec. 2011.

[30] K. Doerschner, D. Kersten, and P. R. Schrater, "Rapid classification of specular and diffuse reflection from image velocities," *Pattern Recogn.*, vol. 44, pp. 1874–1884, Sept. 2011.

[31] D. Zang, K. Doerschner, and P. R. Schrater, "Rapid inference of object rigidity and reflectance using optic flow," in *Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns*, CAIP '09, (Berlin, Heidelberg), pp. 881–888, Springer-Verlag, 2009.

[32] H. Schiffman, *Sensation and perception. An integrated approach*. New York: John Wiley & Sons, Inc., 2001.

[33] G. H. Granlund and H. Knutsson, *Signal Processing for Computer Vision*. Norwell, MA, USA: Kluwer Academic Publishers, 1995.

[34] L. Sharan, Y. Li, I. Motoyoshi, S. Nishida, and E. H. Adelson, "Image statistics for surface reflectance perception," *J. Opt. Soc. Am. A*, vol. 25, pp. 846–865, Apr 2008.

[35] R. W. Fleming, R. O. Dror, and E. H. Adelson, "Real-world illumination and the perception of surface reflectance properties," *Journal of Vision*, vol. 3, pp. 347–368, 2003.

[36] R. O. Dror, E. H. Adelson, and A. S. Willsky, "Estimating surface reflectance properties from images under unknown illumination," *Proceedings of the SPIE 4299: Human Vision and Electronic Imaging IV*, 2001.

[37] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson, "Image statistics and the perception of surface qualities," *Nature*, vol. 447, pp. 206–209, Apr. 2007.

[38] B. L. Anderson and J. Kim, "Image statistics do not explain the perception of gloss and lightness," *Journal of Vision*, vol. 9, no. 11, 2009.

[39] E. H. Adelson, "Image statistics and surface perception," *Human Vision and Electronic Imaging XIII, Proceedings of the SPIE*, vol. 6806, pp. 680602–680602–9, 2008.

[40] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 891–906, 1991.

[41] K. G. Derpanis and J. M. Gryn, "Three-dimensional nth derivative of gaussian separable steerable filters," in *ICIP (3)*, pp. 553–456, 2005.

[42] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *ICIP '95: Proceedings of the 1995 International Conference on Image Processing (Vol. 3)-Volume 3*, (Washington, DC, USA), IEEE Computer Society, 1995.

[43] W. K. Pratt, *Digital Image Processing: PIKS Inside*. New York, NY, USA: John Wiley & Sons, Inc., 3rd ed., 2001.

[44] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London. Series B, Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.

[45] E. P. Simoncelli, *Distributed representation and analysis of visual motion*. PhD thesis, Cambridge, MA, USA, 1993. Not available from Univ. Microfilms Int.

[46] E. H. Adelson and J. R. Bergen, "Spatio-temporal energy models for the perception of motion," *J. OPT. SOC. AM. A*, vol. 2, no. 2, p. 1861, 1985.

[47] R. P. Wildes and J. R. Bergen, "Qualitative spatiotemporal analysis using an oriented energy representation," in *Proceedings of the 6th European*

Conference on Computer Vision-Part II, ECCV '00, (London, UK, UK), pp. 768–784, Springer-Verlag, 2000.

[48] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Norwell, MA, USA: Kluwer Academic Publishers, 1994.

[49] S. Winder, *Analog and Digital Filter Design*. Edn Series for Design Engineers, Newnes, 2002.

[50] K. R. Castleman, *Digital Image Processing*. Prentice Hall Professional Technical Reference, 1st ed., 1979.

[51] T. Herter and K. Lott, "Algorithms for decomposing 3-d orthogonal matrices into primitive rotations," *Computers & Graphics*, vol. 17, no. 5, pp. 517–527, 1993.

[52] E. P. Simoncelli, "Design of multi-dimensional derivative filters," in *Proc 1st IEEE Int'l Conf on Image Proc*, vol. I, (Austin, Texas), pp. 790–794, IEEE Sig Proc Society, Nov 13-16 1994.

[53] M. van Ginkel, P. Verbeek, and L. van Vliet, "Improved orientation selectivity for orientation estimation," in *Proc. 10th Scandinavian Conference on Image Analysis*, SCIA '97, (Lappeenranta, Finland), pp. 533–537, M. Frydrych, J. Parkkinen, A. Visa (eds.), 1997.

[54] J. S. Lim, *Two-dimensional signal and image processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1990.

[55] D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '95, (New York, NY, USA), pp. 229–238, ACM, 1995.

[56] E. J. Candès and D. L. Donoho, "Ridgelets: a key to higher-dimensional intermittency?," *Philosophical Transactions of The Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 357, pp. 2495–2509, 1999.

[57] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling & Simulation*, vol. 5, no. 3, pp. 861–899, 2006.

[58] L. Demanet and L. Ying, "Wave atoms and sparsity of oscillatory patterns," *Applied and Computational Harmonic Analysis*, vol. 23, pp. 368–387, Nov. 2007.

[59] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, pp. 2091–2106, Dec. 2005.

[60] D. L. Donoho, "Wedgelets: nearly-minimax estimation of edges," *Ann. Statist*, pp. 859–897, 1999.

[61] D. L. Donoho, X. Huo, I. Jermyn, P. Jones, G. Lerman, O. Levi, and F. Natterer, "Beamlets and multiscale image analysis," in *in Multiscale and Multiresolution Methods*, pp. 149–196, Springer, 2001.

[62] K. Guo and D. Labate, "Optimally sparse multidimensional representation using shearlets," *SIAM journal on mathematical analysis*, vol. 39, no. 1, p. 298, 2008.

[63] N. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. Royal Society London A*, vol. 357, pp. 2543–2560, 1997.

[64] Labate, "Sparse multidimensional representation using shearlets," *SPIE Proc. 5914, SPIE, Bellingham*, pp. 254–262, 2005.

[65] T. S. Lee, "Image representation using 2d gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, pp. 959–971, Oct. 1996.

[66] Y. M. Lu and M. N. Do, "Multidimensional Directional Filter Banks and Surfacelets," *Image Processing, IEEE Transactions on*, vol. 16, pp. 918–931, Mar. 2007.

[67] S. Mallat and G. Peyré, "A review of bandlet methods for geometrical image representation.," *Numerical Algorithms*, vol. 44, no. 3, pp. 205–234, 2007.

[68] J. Neumann and G. Steidl, "Dual-tree complex wavelet transform in the frequency domain and an application to signal classification," *IJWMIP*, vol. 3, no. 1, pp. 43–65, 2005.

[69] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *Image Processing, IEEE Transactions on*, vol. 14, pp. 423 –438, april 2005.

[70] A. Karasaridis and E. Simoncelli, "A filter design technique for steerable pyramid image transforms," in *Proceedings of the Acoustics, Speech, and Signal Processing, 1996. on Conference Proceedings., 1996 IEEE International Conference - Volume 04*, ICASSP '96, (Washington, DC, USA), pp. 2387–2390, IEEE Computer Society, 1996.

[71] K. Castleman, M. Schulze, and Q. Wu, "Simplified design of steerable pyramid filters," in *Circuits and Systems, 1998. ISCAS '98. Proceedings of the 1998 IEEE International Symposium on*, vol. 5, pp. 329 –332 vol.5, may-3 jun 1998.

[72] C. Delle Luche, F. Denis, and A. Baskurt, "Three-dimensional steerable pyramid based on conic filters," 2003. Wavelet Applications in Industrial Processing , part of symposium Photonics East 2003, 27-31 October 2003, Providence, Rhode Island, USA.

[73] H. S. M. Coxeter, *Regular Polytopes*. Dover Publications, 1973.

[74] A. Cetin, O. Gerek, and Y. Yardimci, "Equiripple fir filter design by the fft algorithm," *Signal Processing Magazine, IEEE*, vol. 14, pp. 60 –64, mar 1997.

[75] F. van der Heijden, R. Duin, D. de Ridder, and D. M. J. Tax, *Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB*. Wiley, 1 ed., Nov. 2004.

[76] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Third Edition*. Academic Press, 3 ed., Mar. 2006.

[77] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2003.