Hierarchical Fusion Based Deep Learning Framework for Lung Nodule Classification

Kazim Sekeroglu
Louisiana State University and Agricultural and Mechanical College, ksekeroglu@gmail.com

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Artificial Intelligence and Robotics Commons, Bioimaging and Biomedical Optics Commons, and the Biomedical Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/4115

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
HIERARCHICAL FUSION BASED DEEP LEARNING FRAMEWORK FOR LUNG NODULE CLASSIFICATION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Electrical Engineering

by

Kazim Sekeroglu
B.S., Gazi University, 2007
M.S., Southern University and A&M College, 2011
December 2017
to my family...
ACKNOWLEDGMENTS

I would like to first express all my deepest respect and gratitude to my academic advisor, Dr. Ömer M. Soysal, for his continuous support to earn my degree. I thank him for guiding me with his patience and immense knowledge and being more than an advisor from beginning to the end. Nothing is enough to express my sincere respect to him.

Beside my advisor, I thank my committee members especially Dr. Xin (Shane) Li for being a co-chair in my dissertation committee and Dr. Jerry Trahan, Dr. Jian Zhang, and Dr. Kenneth L. (Kip) Matthews II, for assisting me during my research and giving their valuable advises.

My sincere appreciation goes to my parents for their guidance and moral support throughout my whole life. And, my deepest gratitude goes to my wife from my heart for all of her sacrifice, support and patience during my doctorate. I could have never completed this Ph.D. without her and my beloved son and daughter.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ............................................................................................................. iii

ABSTRACT ............................................................................................................................... v

CHAPTER 1. INTRODUCTION ................................................................................................. 1
  1.1 Motivation ....................................................................................................................... 1
  1.2 Problem Definition and Proposed Framework .............................................................. 2

CHAPTER 2. LITERATURE REVIEW ....................................................................................... 5
  2.1 Computer-Aided Detection of Lung Nodules Based on Hand Crafted Features .......... 5
  2.2 Computer-Aided Detection of Lung Nodules Based on Deep Learning ................... 9

CHAPTER 3. DEEP CONVOLUTIONAL NEURAL NETWORKS IN A HIERARCHICAL
  LEARNING MODEL ............................................................................................................. 12
  3.1 Multi-Perspective Fusion Based Deep Learning (MPF) ................................................. 12
  3.2 Fusion of Hierarchical Deep Learning Models and Modular Training ..................... 13
  3.3 Single Feature & Multi-Perspective Fusion (SFMPF) .................................................... 17
  3.4 Multi-Feature & Multi-Perspective Fusion (MFMPF) .................................................... 21

CHAPTER 4. COMPARISON OF SIFT, BI-SIFT, AND TRI-SIFT AND THEIR FREQUENCY
  SPECTRUM ANALYSIS ...................................................................................................... 23
  4.1 Related Work ................................................................................................................ 26
  4.2 Method .......................................................................................................................... 27
  4.3 Experiments and Results ............................................................................................. 48
  4.4 Conclusion .................................................................................................................... 67

CHAPTER 5. EXPERIMENTS AND RESULTS ........................................................................ 69
  5.1 Data Preparation .......................................................................................................... 69
  5.2 Performance Assessment Measures ............................................................................. 71
  5.3 Experimental Results of MPF Model ........................................................................... 73
  5.4 Experimental Results of SFMPF Models .................................................................... 80
  5.5 Experimental Results of MFMPF Model .................................................................... 95

CHAPTER 6. CONCLUSION AND FUTURE WORK .............................................................. 97
  6.1 Contributions and Findings ......................................................................................... 97
  6.2 Future Work ................................................................................................................ 101

REFERENCES ....................................................................................................................... 103

APPENDIX: PERMISSIONS .................................................................................................. 109

VITA ..................................................................................................................................... 113
ABSTRACT

Lung cancer is the leading cancer type that causes the mortality in both men and women. Computer aided detection (CAD) and diagnosis systems can play a very important role for helping the physicians in cancer treatments. This dissertation proposes a CAD framework that utilizes a hierarchical fusion based deep learning model for detection of nodules from the stacks of 2D images. In the proposed hierarchical approach, a decision is made at each level individually employing the decisions from the previous level. Further, individual decisions are computed for several perspectives of a volume of interest (VOI). This study explores three different approaches to obtain decisions in a hierarchical fashion. The first model utilizes the raw images. The second model uses a single type feature images having salient content. The last model employs multi-type feature images. All models learn the parameters by means of supervised learning. In addition, this dissertation proposes a new Trilateral Filter to extract salient content of 2D images. This new filter includes a second anisotropic Laplacian kernel in addition to the Bilateral filter’s range kernel. The proposed CAD frameworks are tested using lung CT scans from the LIDC/IDRI database. The experimental results showed that the proposed multi-perspective hierarchical fusion approach significantly improves the performance of the classification.
CHAPTER 1. INTRODUCTION

1.1 Motivation

Although the lung cancer is the second most commonly diagnosed cancer in both men and women, it is the leading cancer type that causes the mortality in both men and women [1]. Lung nodule detection is a very challenging task. The research team in [2] explored the effect of the low-dose CT scans in cancer mortality. Utilizing either low-dose CT or chest radiography, they screened around 53K high lung cancer risk patients three times a year between August 2002 and April 2004. The results of their study show that there is a 20% reduction in mortality of the patients who were screened by low-dose CT scan. Even though the CT scans helps to reduce the mortality rate, the radiologists’ decision may differ significantly in identification of the lung nodules from the CT scans. As an example, [3] shared the results of two radiologists examination over 25 CT scans; the results show that one of the radiologist detected 20 nodules, whereas the other radiologist detected 63 nodules from the same CT scans.

A CAD system increases the performance of the nodule detection substantially. The study conducted by [4] showed that the CAD system reduced significantly the number of false positives. The research in [5] that studied the effect of a CAD system in detection of small nodules shared the results of 6 radiologists examination over 52 CT scans with/without a CAD system. The results show that the CAD system improves a radiologist’s performance considerably. In [6], the performance of the commercial CAD software Lung-CAD VB10A and Siemens AG Healthcare were compared with the performance of two independent readers for detecting the pulmonary nodules in NELSON dataset. The study showed that sensitivity of CAD was 96.7% with a 3.7 FPs/scan and sensitivity of double reader was 78.3% with 0.5 FPs/scan. Therefore, CAD system
with a higher nodule detection rate can be a good help for radiologist to decrease the number of missed nodules, particularly, the small nodules in their early stages.

1.2 Problem Definition and Proposed Framework

The main goal of this dissertation is to explore several models for detection of lung nodules from CT scans by means of hierarchical fusion based deep learning. Specific aims of this dissertation are

1- Exploring fusion of decisions obtained from multi-slice and from multi-perspective in a hierarchical manner:

Using multiple slices from each view and fusing them in hierarchical manner has not been explored before. Among few studies which explores the deep learning for lung nodule classification, the only method that uses multiple views of a volume is proposed in [7]. However, the method utilizes only a single slice from each view and has only one level of fusion.

2- Supervised-learning based fusion of decisions obtained by multi-deep learning models:

According to [7] and [8] that studied fusion of multi-deep learning models, the fusion-based models increase performance of the classification. To the best of our knowledge, the fusion strategies used in deep learning are averaging, multiplication, or using a voting scheme [9], [10], [11], and [8]; all these methods are based on simple combination of decisions from multiple deep learning models. There is no learning at fusion level explored in the previous studies.

3- Exploration of single feature image based classification within the hierarchical multi-deep learning scheme (HMDLS):
In the previous studies, raw CT scans are used for the lung nodule detection by means of deep learning methods. However, the effect of using feature images instead of raw CT scans has not been explored yet.

4- Exploration of multi-feature image based classification within the hierarchical multi-deep learning scheme

5- Exploring the effect of newly proposed Trilateral filter-based feature image in HMDLS:

Bilateral and Trilateral filters are used to obtain the feature images and the use of these feature images are explored for the feature image based deep learning model for lung nodule classification.

The proposed work is inspired from the work done by Dr. Soysal et al [12] where he proposed a modular learning approach for lung nodule classification using hand crafted features. The proposed framework is seen in Figure 1. Once the volume of interest (VOI) is extracted, the slices from different perspectives are fed into the hierarchical deep fusion network, and the class score of the input data is computed in the output. Different type of fusion schemes are proposed in hierarchical deep fusion network. Proposed basic scheme is multi-perspective hierarchical fusion of raw images (MPF) where the slices from different perspectives are classified hierarchically and the class scores are fused at the decision level by the proposed supervised learning based fusion method. Another proposed fusion schema is based on a single feature image and it is called single feature multi-perspective fusion (SFMPF) in which the feature images are used as an input to the hierarchical deep fusion network. The single feature image approach is used to extend the basic MPF scheme to a multi-feature and multi-perspective fusion (MFMPF) by using different type of feature images from different perspectives and fusing them with the proposed hierarchical fusion
approach. MFMPF scheme allows to fuse the decisions made by looking at different features and different perspectives of the object.

Rest of the dissertation is organized as follows: In the second chapter, literature review for the lung nodule classification from CT scans using hand crafted features and using deep learning models are provided. In the third chapter, proposed hierarchical fusion based deep learning models are explained in detail and three different models are proposed. Newly proposed Trilateral filter and its use of creating anisotropic scale-space for feature extraction is explained in detail in the fourth chapter. In chapter five, data preparation, performance assessment measures, experimental setup and analysis of results are given. Finally, in the sixth chapter, contributions and findings with the feature directions of the proposed research are provided.
CHAPTER 2. LITERATURE REVIEW

2.1 Computer-Aided Detection of Lung Nodules Based on Hand Crafted Features

Computer-aided detection and diagnosis (CAD) systems have been studied for decades to get more accurate detection and decrease the work load on the radiologists. A complete computer-aided detection and diagnosis algorithms are usually composed of three main steps: 1. detection of the nodule candidates, 2. extraction of the features from the nodule candidates, 3. false positive reduction and classification. There are different approaches used for the detection of the nodule candidates based on 2D or 3D segmentation. Since the intensity value of the nodule and the other structures in the lung region are differ from each other, most of the segmentation methods are based on gray level thresholding. After segmenting out the nodule candidates, extracting the robust features to be used in classification is the next step. The most common features extracted from the nodules are shape and texture based features. Once the features are extracted from the nodule candidates, to reduce the false positives, one of the classification methods such as k-nearest neighbor, support vector machine, linear discriminant, or random forest classifier is used. In [13] after detecting the nodule candidates, local image features; number of voxels, compactness, ratio, sphericity are used with 2 stage k-NN classifier for the false positive reduction. 813 CT scans from NELSON Trial data generated in Europe were used and the proposed method achieved a sensitivity of 80% with an average 4.2 false positives per scan (FPs/scan).

In [14] a fully automated CAD system for lung nodule detection algorithm is proposed. Authors are claiming that detecting and segmenting the nodules at the same time is one of the advantage of their candidate detection algorithm. Once the nodule candidates are detected, a total of 245 features based on geometric, intensity and gradient are extracted from each nodule candidate. A sequential forward selection process is used to select the best descriptive features
from out of 245 features and these features are used in Fisher Linear Discriminant (FLD) classifier and a quadratic classifier. The comparison result of the two classifiers shows that the FLD classifier performs better than the quadratic classifier. According to the 7 fold cross validation, sensitivity of the proposed CAD system with FLD classifier is 82.66% with an average of 3 FPs/Scan using LIDC dataset. There are 84 scans and 143 nodules in LIDC data set.

In [15] authors aim to develop a CAD system that can automatically detect a pulmonary nodule greater than or equal to 3mm. Once they segmented the nodule candidates by using 3D mass spring models, seven features 1. surface area, 2. volume, 3. sphericity, 4. mean of the nodule intensity, 5. standard deviation of the nodule intensity, 6. skewness of the nodule intensity, and 7. kurtosis of the nodule intensity are extracted from each nodule candidate. They have 2 stages for false positive reduction and classification. In the first stage, they are eliminating the noodles candidates smaller than 3mm and greater than 50mm. In the second stage, they are using neural network with one input, one hidden and one output layers for the classification of nodule candidates. They have 84 CT scans from LIDC dataset and 148 nodules. The proposed algorithm reaches the sensitivity of 88% with 2.5 FPs/scan.

In [16] as in most of the CAD system, the proposed method also has two main stages, nodule candidate detection and false positive reduction. A hierarchical 3D block analysis method is used for nodule detection and SVM classifier is used for false positive reduction. After nodule candidates are detected, 2D and 3D geometric features such as area, diameter, circularity, volume, compactness, elongation and 2D texture features such as mean, variance, skewness, kurtosis, and eigenvalues are extracted to be used as an input to the SVM classifier. LIDC dataset is used for the experiments. There are 84 scans in LIDC dataset but only 58 of them contains nodule and only
those 58 scans were used in the experiments. The proposed method achieved 95.28% sensitivity with 2.27 FPs/scan.

[17] Concentrates on computer aided detection of subsolid pulmonary nodules. Authors used a threshold based methods for the nodule candidate detection. Once the nodule candidates are detected, set of 128 features based on intensity, texture, shape and context are extracted from each of the nodule candidates. Then these features are used in different type of classifiers such as GentleBoost, SVM, k-nearest neighbor, linear discriminant, nearest mean, and random forest classifiers. According to the results from the FROC curves, GentleBoost classifier perform best and it reaches to the sensitivity of 80% with 1 FPs/scan.

In most of the proposed CAD algorithms, the data set used in training and testing and the way of the performance is assessed differ from method to method. Therefore, there is a bottleneck in the comparison between the performances of the proposed CAD algorithms [18]. There are few studies which compares the performance of the existing CAD systems by using the same dataset and the same evaluation method.

In [19] existing CAD methods were compared by testing and evaluating them with the same data and same method and also authors proposed a method for combining the tested CAD system for a better performance. In this study ANODE09 database which includes 55 scans from a lung cancer screening program is introduced. Performance of the six different CAD algorithms were compared and each CAD method is evaluated based on their average sensitivity of seven different FP rates: 1/8, 1/4, 1/2, 1, 2, 4, and 8. According to the results, there is a significant performance difference between the algorithms and combining the results of each CAD system leads to a better performance.
Another study to improve the performance of the existing CAD system by combination is proposed in [20]. They propose a set of 4 different methods to combine the existing CAD systems for 4 different scenarios for a better performance. The first method is proposed where there is only the location of the nodule is available as an output of the CAD system. In this case the method suggests combining the detected locations of the CAD systems. In the second and the third scenario, in addition to the location of the nodule, the level of suspicion for each detected nodule is available. Lastly in the fourth scenario, most of the internal details such as training data, feature vectors, classifiers etc. of the CAD systems are available. However, authors are not discussing a combination method for this case. Since it is not likely to have access to the internal details of most of the CAD system in practice.

In [21] the performance of the state of the art CAD systems VISIA, Herakles, and ISICAD for detection of the pulmonary nodules are compared by using LIDC/IDRI dataset. After comparison of the CAD systems, the false positives of the best performing one was examined by four radiologists to see if the CAD system can detect any nodule that was missed by the radiologist during the annotation. Out of these 3 CAD systems, Herakles performed best with the sensitivity of 82% with 3.1 FPs/scan for nodules annotated by all four readers. While Herakles achieving a more robust performance, other two CAD systems VISIA and ISICAD showed substantial performance difference on LIDC/IDRI dataset. The reason for the performance drop on ISICAD is that it is trained exclusively on NELSON dataset which “consists of homogeneous thin-slice data reconstructed with a soft/standard reconstruction kernel”. Thus, it is important to use heterogeneous dataset like LIDC/IDRI to train and test the CAD system. Lastly, there were 45 nodules which were accepted as nodule\textgreater=3mm by all four radiologists detected by the CAD system but overlooked by the radiologist during the annotation procedure.
According to the review of CAD systems for lung cancer in CT scans, the CAD systems are still not used widely by the community of the radiologists. Therefore, further research and development is needed in CAD system particularly for decreasing the "number of false positives (FP), having high processing speed, presenting high level of automation, low cost (of implementation, training, support and maintenance), the ability to detect different types and shapes of nodules, and software security assurance" [22].

2.2 Computer-Aided Detection of Lung Nodules Based on Deep Learning

So far, the CAD systems which use the hand-crafted features for the classification is reviewed. However, there are existing CAD algorithms for pulmonary nodule detection which are based on the deep learning methods such as convolutional neural networks, deep belief networks, and autoencoders. One of the earliest study that uses the deep learning system for lung nodule classification is [23]. In [23] classification of the pulmonary nodules as malignant or benign by using deep learning methods were explored. Specifically, the deep belief network (DBN) and convolutional neural network (CNN) models were tested. This is one of the first study that explores the application of the deep learning techniques for the classification of pulmonary nodules. LIDC-IDRI dataset that includes 1010 scans and 2545 nodules which are greater than 3mm is used for testing the proposed methods. For comparison of the deep learning methods and the feature based methods, two of the well performing features SIFT and local binary pattern (LBP) features with k-NN classifier is used. DBN was able to classify pulmonary nodules with 82.2% sensitivity and the SIFT+LBP feature based classifier reached the sensitivity of 66.8%. Another earlier study for classifying the pulmonary nodules as malignant or benign is [24]. The classification is done by using the deep features extracted from 2D images by the autoencoder and classified by the binary decision tree. Publicly available LIDC/IDRI dataset is used to train and test the algorithm.
Although there are 1010 CT scans available in LIDC/IDRI dataset, only 157 scans have the proper annotation for the nodules for being benign or malignant. The proposed method achieved the sensitivity of 83.35% with 0.39 FPs/scan over a 10-fold cross validation.

In [25] 3D convolutional neural network based lung nodule classification algorithm is proposed. Authors are claiming that the proposed method can work with weakly labeled 3D data as in the case of only the label of the central voxel and the size of the largest expected nodule are provided. Once they estimate the labels of the 3D training data by using basic thresholding and 2D SLIC super-pixels of the 2D slices, they use these data to train 3D CNN for the nodule classification. The negative samples are extracted from the lung area by randomly sampling the locations based on the threshold. SPIE-AAPM-LUNGx dataset is used to train and test the proposed method. The dataset contains 70 CT scans. 15K positive and 20K negatives samples are labeled by the proposed method. The proposed method achieved 80% sensitivity with 10 FPs/scan.

One of the latest study for lung nodule detection using deep learning methods is done in [7]. Authors are proposing a multi-view CNN for lung nodule detection. In the proposed method, they are extracting the volume of interest as a cube. Then 2-D patches from nine symmetrical perspectives of the extracted volume are fed into a separate CNNs. The outputs of the CNNs are fused based on different architecture. First fusion structure is committee fusion where the fusion is done at decision level. Once the class scores from each CNN is computed, they are fused using a product rule on the output probabilities [26]. Another fusion method is late fusion where the fusion is done at feature level by concatenating the outputs of the first fully connected layers. Lastly, they are using mixed fusion which is the combination of the committee and the later fusion. Although this proposed method is fusing the slices from multi-view, they are using a single slice
from each view and the way the fusion is done similar to the previously proposed fusion approaches.
CHAPTER 3. DEEP CONVOLUTIONAL NEURAL NETWORKS IN A HIERARCHICAL LEARNING MODEL

3.1 Multi-Perspective Fusion Based Deep Learning (MPF)

The proposed method is based on the hierarchical fusion of the multiple perspectives of the 3D seen. Three different perspectives, transverse, coronal and sagittal are used as shown in Figure 2. Multiple slices are used from the same object for each perspective. In the hierarchical learning model shown in Figure 3, the first module is called a Slice Module (MS) which is a Deep Convolutional Neural Network. Slice module outputs a class score for each single 2D slice. Then the class scores from the multiple slices from the same object is concatenated and transposed to create an input to the next module namely Perspective Module (MP). The perspective module computes the classification score for each object at the perspective level. Once the decision made for each perspective, the class scores from different perspectives are concatenated to create an input to the last module namely Volume Module (MV). At the final layer, the class score is obtained for each 3D object at the volume level. In the following sections, more detailed information is provided for the hierarchical fusion of deep learning models as well as each module used in the proposed method.

Figure 2. Raw input images. Slices from three perspectives are used.
3.2 Fusion of Hierarchical Deep Learning Models and Modular Training

We proposed a supervised learning based fusion of hierarchical deep learning models. In the proposed method, there are three levels of hierarchical predictions: 1. Slice level, 2. Perspective level, and 3. Volume level. Each module at each level are trained separately in a hierarchical modular fashion such as; first the slice modules are trained, then the class scores obtained at each slice module is concatenated for each object to create a new feature vector for the next level of prediction which is the perspective level. For instance, if there are 10 slices in each input object, every 10 class scores produced from each slice module is concatenated and transposed to create an input feature vector for the perspective module. Then each of the perspective module is trained by using the features created by the slice modules. After training each perspective module, class scores are computed for each object at perspective level. Since there are three perspectives used in the proposed method, the class score from each of the three perspectives are concatenated to create a input feature for the next level which is the volume level. Finally, the volume module is trained by using the features computed by the perspective module and the final class score at volume level for each 3D object is computed at volume module. This process is illustrated in Figure 4.
Figure 4. Re-arrangement of the class scores from slice level classification to create an input feature for the Perspective Module

### 3.2.1 Slice Module (Deep Convolutional Neural Network)

Deep learning models are learning the representation of the data by processing it with multiple layers and consecutively obtaining the abstract representation [27]. Deep convolutional neural networks (DCNN) are one of the best deep learning models for the object recognition task [28].
The inspiration of the DCNN is LeNet5 which is the 7-layer convolutional neural network designed for the handwritten character recognition in 1998 by Yann LeCun et al. [29]. The block diagram of the LeNet5 is shown in Figure 5. It is composed of 2 convolution layers followed by pooling layers and 2 fully connection layers followed by the output layer. In the convolution layer, the input image is convolved with the filters (kernels) which are in fact the shared weights adjusted during the training by using the backpropagation algorithm. To extract different type of structures, more than one filter can be used in convolution layer. After convolving the image with different filters, in the next stage feature selection is done by pooling. In pooling operation, the selected region (which is usually determined as 2x2 or 3x3) is represented by the max or the average intensity of the region. This abstraction by convolution and pooling can be done multiple times to get a better representation of the data. After obtaining the final representation, the features are fed into the fully connected layer and then the class core is obtained at the output layer.

![Figure 5. LeNet5 developed by Yann LeCun et al [29]](image)

In the hierarchical deep learning model, there is a distinct slice module for each perspective, so there are three different slice modules for three different perspectives. Since the structure of all three slice modules are the same, only a single structure of the slice module is covered in this section. Slice module is a deep convolutional neural network (DCNN) and it is depicted in Figure 6. It consists of four convolutional and pooling layers following with the regular one hidden layer.
neural network. Input of the slice module is 2D slices from the selected perspective of the 3D object. The input size of each 2D slice is 56x56 pixels. At the first convolution layer, there are 8 filters in the size of 3x3. The filter sizes are the same for all convolution layers. Also, the number of filters at the second, third, and the fourth convolutional layers are the double of the number of filters at their previous convolutional layer. Hence, at the last convolutional layer, there are 64 filters. After the last pooling layer, there is a fully connected layer comprised of 32 number of units. At the convolutional and fully connected layers rectified linear function defined by (1) and at the output layer softmax function defined by (2) are used. The filters at each convolution layer are adjusted by back-propagating the error obtained at the output based on the cross-entropy loss function defined by (3).

\[ f(x) = \max(0, x) \quad (1) \]

\[ \sigma(x)_j = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}} \quad (2) \]

where \( K \) is the total number of neurons in the layer, and \( j \) is the index of the neuron at the output layer.

\[ L = -\sum_j t_j \log(p_j) \quad (3) \]

where \( t \) is the target and \( p \) is the predicted values at the output layer, and \( j \) is the index of the neuron at the output layer.
3.2.2 Perspective and Volume Modules

In the perspective module and the volume module any supervised classifier such as Support Vector Machine, ANN, Bayesian Network, or a multi-dimensional regression model can be used. In this dissertation, a regular feedforward ANN is used for perspective and volume level predictions. As shown in Figure 7, perspective modules and the volume module are ANNs composed of an input layer following with a single hidden layer and the output layer. In the hidden layer, rectified linear function and at the output layer softmax function are used as an activation function for the non-linear transformation.

3.3 Single Feature & Multi-Perspective Fusion (SFMPF)

Extracting salient content from the input data can lead to a better representation and better classification accuracy. Therefore, in the proposed method, the feature images are used instead of the raw images to learn the representation of the data. Features images can be obtained by applying filters such as Bilateral, Trilateral, LOG, or Gabor filters to the raw images as shown in Figure 8 where $V_1$, $V_2$ and $V_3$ are the raw images from three different perspectives and $FI_1$, $FI_2$, and $FI_3$ are...
the feature images. Once the feature images are created, they are fed into the same proposed hierarchical fusion network architecture as shown in Figure 9.

Figure 7. Perspective and volume modules are regular feedforward ANN with one hidden layer

3.3.1 Creating Feature Images

In the proposed feature based hierarchical deep fusion, 4 different methods Bilateral Filtering, Trilateral Filtering, Laplacian of Gaussian (LoG) Filtering, and Gabor Filtering are used to acquire the feature images. In this section, background for LoG and Gabor filters are provided. Bilateral and Trilateral filters are explained in detail in chapter four.

Laplacian operator ($\Delta$) can be used to measure the rapid changes in the image. Laplacian of an input image $I(x, y)$ at a pixel point $(x, y)$ is given by (4)

$$\Delta I(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$ (4)

However, before applying Laplacian operator, smoothing the input image to reduce the effect of noise is a very common approach. Therefore, the input image is convolved with a Gaussian filter, defined by (5), with the shape parameter $\sigma$ before applying Laplacian operator as in (6).
\[ G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (5) \]

\[ \Delta[G(x, y, \sigma) * I(x, y)] \quad (6) \]

To reduce the cost of computation, one can use (7) instead of (6)

\[ \Delta[G(x, y, \sigma) * I(x, y)] = \Delta G(x, y, \sigma) * I(x, y) \quad (7) \]

where (7) is derived using the convolution property defined by (8).

\[ \frac{d}{dt}[h(t) * f(t)] = \frac{d}{dt} \int f(\tau)h(t-\tau)d\tau \quad (8) \]

\[ = \int f(\tau) \frac{d}{dt}h(t-\tau)d\tau = f(t) * \frac{d}{dt}h(t) \]

Hence the LoG \( \Delta G(x, y, \sigma) \) is given by

\[ \Delta G(x, y, \sigma) = -\frac{1}{\pi\sigma^4} \left[ 1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-\frac{x^2 + y^2}{2\sigma^2}} \quad (9) \]

Figure 8. Feature images are created by filtering the slices from each perspective.
Figure 9. Block diagram of the proposed SFMPF model

Figure 10. (a) Raw nodule, (b) through (h) are LoG filtered nodule with $\sigma = 1$ to $\sigma = 7$ by 1 incremental

In addition to LoG filter, Gabor filter is also used to create the feature image. Gabor filters are typically used to extract the textures in the images. Gabor filter is composed by multiplying a Gaussian kernel with a complex sinusoid.

$$G(x, y) = g(x, y) \cdot s(x, y)$$ \hspace{1cm} (10)

Where $g(x, y)$ is a 2D Gaussian kernel with the standard deviation of $\sigma_x$ and $\sigma_y$,

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right)}$$ \hspace{1cm} (11)
and $s(x,y)$ is the complex sinusoid with the center frequency of $\omega_{x_0}$ and $\omega_{y_0}$

$$s(x,y) = \cos(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y) + i\sin(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y)$$

Using the Euler’s formula (13)

$$e^{i\theta} = \cos \theta + i\sin \theta$$

(12) can be written as

$$s(x,y) = e^{i(2\pi(\omega_{x_0}x + \omega_{y_0}y))}$$

Therefore, the complex Gabor filter is

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} e^{i(2\pi(\omega_{x_0}x + \omega_{y_0}y))}$$

Real part of the Gabor filter is

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} \cos(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y)$$

Imaginary part of the Gabor filter is

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} \sin(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y)$$

### 3.4 Multi-Feature & Multi-Perspective Fusion (MFMPF)

In Figure 11, the block diagram of the hierarchical fusion of multiple SFMPF models is illustrated. This proposed fusion scheme is the composition of previously proposed SFMPF scheme with different type of feature images. The idea is first making multiple decisions by different SFMPF models which make the predictions based on different type of feature images. Then the decisions obtained from the multiple SFMPF models are fused to make the final decision.
Figure 11. Block diagram of MFMPF model
CHAPTER 4. COMPARISON OF SIFT, BI-SIFT, AND TRI-SIFT AND THEIR FREQUENCY SPECTRUM ANALYSIS

The Scale Invariant Feature Transform (SIFT) proposed by David Lowe is one of the pioneering algorithms for describing image content [30]. The SIFT can be utilized for image matching, object or scene recognition, stereo correspondence, and motion tracking. This algorithm provides a mechanism to achieve invariance to scaling and rotation and to changes in illumination and camera viewpoint. The original SIFT utilizes the Gaussian kernel in construction of an isotropic scale-space. The Bi-SIFT and Tri-SIFT that are proposed in this paper utilize the Bilateral kernel for anisotropic scale-space processing of an image in detection of stable keypoints. The rationale behind the use of these anisotropic kernels is briefly explained in the sequel. This paper aims to analyze the frequency spectrum of the kernels used in these three SIFT versions and compare their performance in terms of detected number of matching keypoints, warping intensity error, and scatteredness.

There are two main stages in the SIFT algorithm: 1) Detection of keypoints including scale-space extrema detection, keypoint localization, and orientation assignment and 2) forming keypoint descriptor.

The SIFT achieves scale invariance by iterative search of discrete scales through the Gaussian scale-space. It is reported that the Gaussian kernel has the best form for the scale-space search [31] [32]. The Gaussian response \( L(x, y, \sigma) \) of an image \( I(x, y) \) for the Gaussian \( G(x, y, \sigma) \) kernel that has the scale parameter \( \sigma \) is given by the convolution

\[
L(x, y, \sigma) = (I * G)(x, y, \sigma)
\]

where

\[ \text{This chapter previously appeared as [Şekeroğlu, Kazim, and Ömer Muhammet Soysal. "Comparison of SIFT, Bi-SIFT, and Tri-SIFT and their frequency spectrum analysis", Machine Vision and Applications, 2017]. It is reprinted by permission of Springer as attached in the appendix.} \]
\[ G(x, y, \sigma) = \frac{1}{2\pi \sigma^2} e^{-(x^2 + y^2)/2\sigma^2} \]  

The SIFT utilizes difference of Gaussian (DoG) in detection of the stable pixels throughout the scale-space. The DoG image, as illustrated in Figure 12, is obtained by computing the difference \( D_G(x, y, \sigma) \) between two consecutive Gaussian smoothed images, one with scale \( k \) times the other. To produce the scale-space for each octave, the initial image is repeatedly convolved with Gaussians as shown on the left, and to produce the difference of Gaussian images, starting from the second scale, each consecutive scale is subtracted from the previous scale as shown on the right. \( D_G(x, y, \sigma) \) is given by

\[ D_G(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \]

\[ = [G(x, y, k\sigma) - G(x, y, \sigma)] \ast I(x, y) \]  

Figure 12. Forming scale-space [30]

The second step of the detection stage searches for candidate pixels in terms of local extrema of DoG responses within a 3D 26-neighbor at each sample point as seen in Figure 13. That is, a sample point is considered as a candidate if it has larger (or smaller as a choice) DoG response than all of its 26-neighbors.
The difference of Gaussian scale-space used by SIFT is produced by an isotropic Gaussian filter. Consequently, the high frequency structures such as edges are not preserved throughout the scale-space constructed by using such an isotropic filter. An anisotropic filter that yields a nonlinear scale-space can overcome this drawback of a linear Gaussian scale-space. The Bilateral filter \cite{33} that is anisotropic inherently preserves these fine details while smoothing an image. Substituting the Gaussian filter $G(*)$ in equation (19) with the Bilateral filter $B(*)$, the difference of Bilateral image, $D_B$ is given by

$$D_B(x, y, \sigma_S, \sigma_R) = [B(x, y, k \sigma_S, \sigma_R) - B(x, y, \sigma_S, \sigma_R)] \ast I(x, y)$$

(20)

where $\sigma_S$ and $\sigma_R$ are the parameters for the spatial and range (or intensity) kernels, respectively.

The proposed Trilateral filter employs a Laplacian kernel as a third kernel in addition to scale and range (pixel intensity) kernels. This new anisotropic filter is capable of preserving more high frequency content while blurring the image.

The rationale behind seeking the stable pixels at the extrema of the DoG response relies on the fact that it is an approximation to the scale-normalized Laplacian of Gaussian function $LoG = \sigma^2 \nabla^2 G$. The extrema of LoG provide the most stable image features compared to gradient, Hessian, or Harris corner function as experimentally reported in \cite{34}. The relation between the DoG and
LoG within the scale-space scheme can be understood with the help of the diffusion equation as described in [30]:

\[
(k - 1) \sigma_s^2 \nabla^2 G \approx G(x, y, k \sigma_s) - G(x, y, \sigma_s) \tag{21}
\]

As a remark, equation (21) shows that the normalized LoG is a special case of the diffusion equation and that the DoG is an approximation to the normalized LoG.

The rest of the paper is as follows: A brief on isotropic and anisotropic is given in section 2. Section 3 introduces the proposed Trilateral filter, discusses the frequency spectrum of isotropic and anisotropic filters under consideration, and presents measures to compare SIFT, Bi-SIFT, and Tri-SIFT. In section 4, the comparison results are presented.

4.1 Related Work

The SIFT is one of the most popular feature detection and description method in computer vision research. Since the SIFT has been invented, many variations of SIFT methods as listed in [35], such as PCA-SIFT [36], GSIFT [37], CSIFT [38], and ASIFT [39] have been developed. All of these variants of SIFT algorithms utilize a smoothing kernel that is isotropic over the spatial domain. This type of smoothing applies to all variants equally without distinguishing the local character of an image signal. Consequently, possible ‘good candidates’ that are observed at higher frequencies are eliminated from the search list. In addressing this problem, the authors in [40] have proposed an adaptive anisotropic Gaussian-smoothing kernel, based on the gradient magnitude, gradient direction, and the curvature at each pixel location in creation of the scale-space. This method requires an initial edge detection based on an experimentally determined threshold for each image. They have tested their method by using only two different classes. In their initial dataset, there were two images, and they extended their dataset by scaling each image by 0.1, 0.3, 0.5, 0.7, and 0.9 and rotating each scaled image from 0° to 360° in increments of 2°. They have
implemented the smoothing operations in CUDA by assigning a separate thread for each pixel. This parallel implementation is reduced the computation from 47 hours to 45 minutes for each class. In [41], the authors have created the nonlinear scale-space by using the adapted anisotropic Gaussian filter based on the local statistical characteristics represented by a second-moment matrix. The authors pointed out that the influence of noise should be removed by applying nonlinear diffusion formulation; such a scheme would particularly have more effective noises located around the edges.

In [42], the authors have proposed a nonlinear scale-space based on efficient Additive Operator Splitting techniques and variable conductance diffusion. In [43], the authors proposed an anisotropic scale-space based on a Bilateral filter to overcome the inadequacy of SIFT-based methods for the SAR image registration, which are suffering from unavoidable speckles as well as losing the details in the coarser Gaussian scale-space. They have smoothed the input image with a Gaussian filter and then down-sampled by a factor of two to reduce the speckle. In [44], the authors have proposed a Laplacian of Bilateral scale-space. Their approach constructs the Bilateral scale-space by filtering the images with a Bilateral filter; followingly, a Laplacian filter is applied to these anisotropically smoothed images. The authors claim that they have speeded up the process by using the Laplacian of Bilateral instead of difference of Bilateral.

4.2 Method

In this section, we analyze the behavior of the SIFT, Bi-SIFT, and Tri-SIFT in the frequency domain. A Bilateral filtered image at a pixel $c$ within a $N$ by $N$ neighbor is defined as

$$I_B(c) = \frac{1}{W} \sum_{r=c-N/2}^{c+N/2} B(r; c, \sigma_S, \sigma_R) I(r)$$  \hspace{1cm} (22)
where $\sigma_S$ and $\sigma_R$ are the shape parameters of the spatial Gaussian filter $S$ and the range (intensity) Gaussian filter $R$, respectively, and the Bilateral filter that operates at the central pixel $c$ and the neighboring pixels $r$ is defined by

$$B(r; c, \sigma_S, \sigma_R) = S(r; c, \sigma_S) \ R(I(r); I(c), \sigma_R)$$  \hspace{1cm} (23)

where

$$S(r; c, \sigma_S) = e^{-\frac{|r-c|^2}{2\sigma_S^2}}$$

$$R(I(r); I(c), \sigma_R) = e^{-\frac{|I(r)-I(c)|^2}{2\sigma_R^2}}$$  \hspace{1cm} (24)

$$W = \sum_{r=c-N/2}^{c+N/2} S(r; c, \sigma_S) \ R(I(r); I(c), \sigma_R)$$  \hspace{1cm} (25)

The normalization factor $W$ ensures the sum of the weights is one. The range filter $R(I(r); I(c), \sigma_R)$ determines the effect of the neighbor pixels on the central pixel value $I(c)$ during the smoothing operation. In other words, the contribution of the neighbor pixels is adaptively determined. This adaptive characteristic of the Bilateral filter preserves high frequency structures such as edges during the smoothing operation. Note that the spatial filter $S(r; c, \sigma_S)$ penalizes a neighbor pixel more when it moves away from the center while the range filter penalizes a pixel more when its intensity differs more from the central pixel intensity.
4.2.1 Trilateral Filter

We propose a new filter called Trilateral Filter. This filter is formed by adding a Laplacian kernel as a second range kernel in addition to spatial and range kernels in the Bilateral filter. This additional anisotropic filter measures the variation of the gradient and adjusts the weight of neighbor pixels accordingly. The Trilateral filter is more effective at higher frequency structures. As seen in Figure 14, in a high frequency region, the Laplacian kernel in the Trilateral filter penalizes the neighboring pixels comparing their speed of variation with the central pixel. Hence, the high frequency content is preserved. On the other hand, the range kernel of the Bilateral filter only compares the intensity difference; hence, it does not penalize as much as the Trilateral filter. Over a lower frequency region, the Bilateral and Trilateral filters behave similar. As an alternative to a Laplacian kernel, a gradient kernel would be used; however, the Laplacian kernel has an advantage to measure the speed of the gradient, which is important in high frequency regions. Similar to the Bilateral filter, the Trilateral filtered image is defined by

\[
l_T(c) = \frac{1}{W} \sum_{r = c - N/2}^{c + N/2} T(r; c, \sigma_S, \sigma_R, \sigma_L) I(r)
\]

where \(\sigma_L\) is the shape parameter of the Laplacian kernel \(L(*).\) The difference of Trilateral images \(D_T\) to construct the scale-space for the extrema detection is given by the following equation

\[
D_T(x, y, \sigma_S, \sigma_R, \sigma_L)
\]

\[
= [T(x, y, k \sigma_S, \sigma_R, \sigma_L) - T(x, y, \sigma_S, \sigma_R, \sigma_L)] * I(x, y)
\]
The behavior of the Gaussian, Bilateral and Trilateral filters in a 2D high frequency region can be seen in Figure 15 through Figure 21. The Gaussian filter smoothes the sharp edges as seen in Figure 16. However, the Bilateral filter preserves the sharpness of the edges while removing the noise as shown in Figure 17. On the other hand, the Trilateral filter keeps more details than the Bilateral filter while smoothing as seen in Figure 18.
Figure 15. Input image

Figure 16. Input image filtered by Gaussian filter
Figure 17. Input image filtered by Bilateral filter

Figure 18. Input image filtered by Trilateral filter
Figure 19. Input image filtered by difference of Gaussian

Figure 20. Input image filtered by difference of Bilateral
4.2.2 Frequency Spectrum Analysis

In the image domain, the edges correspond to the high frequency regions and the smooth areas correspond to the low frequency regions. In this section, the behavior of the difference of Bilateral and the difference of Trilateral filters in low and high frequency regions are analyzed by using their frequency spectrum obtained with the Fourier Transform.

For the sake of analysis, assume that we will have an input signal to the Bilateral and the Trilateral filters such that it makes the range kernel function and the Laplacian kernel function a regular Gaussian-shaped function. If the input function is a linear function such that

\[ I(x) = -mx + A_I \]  \hspace{1cm} (29)

\[ \Delta I(x) = A_I - I(x) = mx \]  \hspace{1cm} (30)
Then equation (24) becomes,

\[
R(\Delta I(x); \bar{\sigma}_R) = e^{-\frac{\Delta I(x)^2}{2\sigma_R^2}} = e^{-\frac{m^2x^2}{2\sigma_R^2}} = e^{\frac{-x^2}{2\sigma_R^2}}
\]

(31)

where \(\sigma_R = \bar{\sigma}_R/m\). Based on the assumption that we made, intuitively, the parameter \(\sigma_B\) of the Bilateral kernel will be smaller than or equal to the spatial kernel parameter \(\sigma_S\) as shown below.

\[
B(r; c, \sigma_B) = S(\Delta r(r, c); \sigma_S) R(\Delta I(r, c); \sigma_R)
\]

(32)

\[
\sigma_B = \frac{\sigma_R \sigma_S}{\sqrt{\sigma_S^2 + \sigma_R^2}}
\]

(33)

\[
\frac{\sigma_B}{\sigma_S} = \frac{\sigma_R}{\sqrt{\sigma_S^2 + \sigma_R^2}} \leq 1
\]

(34)

Thus,

\[
\sigma_B \leq \sigma_S, \text{ where } \sigma_R, \sigma_S > 0
\]

(35)

The same logic can apply to the Trilateral filter, giving

\[
\sigma_T \leq \sigma_B \leq \sigma_S
\]

(36)

where, \(\sigma_T\) is the shape parameter of Trilateral filter.

The difference of Gaussian can be expressed in the frequency domain as follows

\[
F\{g(r; k \sigma) - g(r; \sigma)\} = G_{k\sigma}(f_S) - G_{\sigma}(f_S)
\]

(37)

\[
G_{k\sigma}(f_S) - G_{\sigma}(f_S) = e^{-\frac{(2nk\sigma f_S)^2}{2}} - e^{-\frac{(2n\sigma f_S)^2}{2}}
\]

(38)

The maximum occurs at
\[ f_S = \frac{1}{2\pi^2 \sigma_S} \sqrt{\frac{2 \ln k}{k^2 - 1}} \]  

(39)

Therefore,

\[ f_S \propto \frac{1}{\sigma_S} \]  

(40)

Applying the same derivation for difference of Bilateral and the difference of Trilateral filters, similarly

\[ f_B \propto \frac{1}{\sigma_B} \quad \text{and} \quad f_T \propto \frac{1}{\sigma_T} \]  

(41)

where, \( f_B \) and \( f_T \) are the frequency at which maximum occur in the frequency spectrum.

Based on equation (36), one can conclude that

\[ f_S \leq f_B \leq f_T \]  

(42)

As we show in the sequel, the DoG, DoB, and DoT act as a band-pass filter. Equation (42) reads that the frequency band is shifted towards higher frequencies in the case of the difference of Bilateral and Trilateral filters.

To illustrate this shifting, we used a Gaussian kernel as an input function. The Gaussian-shaped input signal for the Bilateral and Trilateral filters also produce an approximate Gaussian-shaped range and Laplacian kernels. Note that the response of a Laplacian kernel vanishes for a linear input signal; consequently, the behavior of the Bilateral and the Trilateral filters will be the same.

We start our analysis with a low frequency input signal in Figure 22. Since the intensity values are very close to each other in a low frequency signal, the range and the Laplacian kernels become very wide; consequently, high weights for the neighbor pixels are produced. The resulting Bilateral and Trilateral kernels will be the same as the Spatial (Gaussian) kernel as seen in Figure 23. This
similarity for the difference of Gaussian, difference of Bilateral, and the difference of Trilateral kernels are preserved when we increase the shape parameter of the spatial kernel from $\sigma$ to $k\sigma$ as seen in Figure 24. The frequency spectrum of DoG, DoB, and DoT kernels are shown in Figure 25 and they are all overlapping. Thus, one can conclude that the behavior of the Gaussian, Bilateral, and the Trilateral filters are the same over a smooth region.

![Figure 22. Low frequency input signal](image-url)
Figure 23. Bilateral and Trilateral kernels for a low frequency signal

Figure 24. DOG, DOB and DOT kernels for a low frequency signal
When the frequency of the input signal is increased as in Figure 26, the difference between the central pixel intensity and the surrounding pixel intensities are also increased. This makes the range and the Laplacian kernels narrower as seen in Figure 27 compared to the one in Figure 23. As it was before, an increase in shape parameter of the spatial kernel from $\sigma$ to $k\sigma$, the resultant Bilateral and the Trilateral kernels become narrower than the spatial (Gaussian) kernel as seen in Figure 28.
Figure 26. High frequency input signal

Figure 27. Bilateral and Trilateral kernels for a high frequency signal-1
If we keep increasing the frequency of the input signal as in Figure 30 and Figure 33, the range and the Laplacian kernels get narrower and narrower, causing to the Bilateral and the Trilateral kernels becoming narrower. The higher input signal frequency leads to narrower Bilateral and Trilateral kernels in a spatial domain. However, this is vice versa in the frequency domain, and the higher input signal frequency leads to wider Bilateral and Trilateral kernels in the frequency domain as shown in Figure 29. When we obtain the difference of Gaussian, difference of Bilateral, and the difference of Trilateral in the frequency domain, there will be a shifting toward higher frequencies as shown in Figure 29, Figure 33, and Figure 37.
Figure 29. Frequency spectrum of DOG, DOB, and DOT kernels for a high frequency signal

Figure 30. High frequency input signal
Figure 31. Bilateral and Trilateral kernels for a high frequency signal

Figure 32. DOG, DOB, and DOT kernels for a high frequency signal
Figure 33. Frequency spectrum of DOG, DOB, and DOT kernels for a high frequency signal

Figure 34. High frequency input signal
Figure 35. Bilateral and Trilateral kernels for a high frequency signal

Figure 36. DOG, DOB, and DOT kernels for a high frequency signal
4.2.3 Comparison of SIFT, Bi-SIFT, and Tri-SIFT

In this section, we explore how SIFT, Bi-SIFT, and Tri-SIFT behave when there is a variation in rotation, scale, and pose view. In comparison, we measure the number of matching keypoints, warping intensity error, and scatteredness between the reference image and the test (transformed) image. The shape parameters $\sigma_R$ and $\sigma_L$ for the range and the Laplacian kernels are determined experimentally, and the ones that give the best result with respect to number of matching keypoints, warping intensity error, and scatteredness are used.

The procedure to find the matching keypoints is as follows [30]: The keypoints and their feature vector are extracted for a reference image and its test image. A keypoint $P_T$ in the test image is matched initially to the keypoint $P_R$ whose feature vector has the lowest Euclidean distance $d_1$ to that of $P_T$ in the reference image. In addition, a second closest keypoint $P_{R2}$ in the reference image is found based on the Euclidean distance $d_2$ between the feature vectors of $P_T$ and
If the ratio $d_1/d_2 > 80\%$, then it is considered as not-matching. This matching scheme eliminates $\%90$ of the false matches and less than $\%5$ of the true matches. Finally, the RANSAC method is used to remove outliers from the matching set. The time complexity of the matching process is $O(NM)$, where $N$ and $M$ denote the number of keypoints for test and reference images, respectively.

After the detection of keypoints, the kernels are compared by measuring the similarity between the reference image $I_R$ and the warped test image $I_T$ by means of the Average Intensity Error (AIE) as defined by equation (43). Once the matching keypoints are found, a homography matrix is obtained by using the epipolar geometry. The test image is transformed by using this homography matrix and warped on to the reference image to calculate AIE.

$$AIE = \frac{1}{NM} \sum_{x}^{N} \sum_{y}^{M} |I_R(x, y) - I_T(x, y)|$$

where $N$ and $M$ are the number of pixels in the overlapping area of $I_R$ and $I_T$, respectively.

The scatteredness $S$ measured by the Eigen values of the second moment matrix of the keypoints as illustrated in Figure 38, is defined by

$$S = \sqrt{\lambda_{max} + \lambda_{min}}$$

where $\lambda_{max}$ and $\lambda_{min}$ are the maximum and the minimum Eigen values. As a remark, when the keypoints are scattered more, the ellipse fitted to the keypoints have a larger diagonal. Note that the local features extracted from a wider region would be more descriptive compared to those extracted from a part of the image.
4.3 Experiments and Results

4.3.1 Data

We employed two different datasets. The face datasets [45] and [46] have 90 different classes and each class has 4 different images: Profile image, scaled, rotated, and 35 degree pose variation of the profile image. Totally, there are 360 images in the face dataset. The object-dataset [47] consists of 2 classes with change in blur, 2 classes with change in viewpoint, 2 classes with change in zoom plus rotation, one class with change in lighting condition and one class with jpeg compression. In this dataset, each class has 6 images and there are 48 images in total. Figure 39 shows the sample images from the object-dataset.
Figure 39. Sample images from the object dataset. (a) graffiti - change in viewpoint angle in structured scene, (b) wall - change in viewpoint angle in textured scene, (c) boat – change in scale in structured scene, (d) bark - change in scale in textured scene, (e) bikes - change in blur in structured scene, (f) trees - change in blur in textured scene, (g) ubc - jpeg compression, (h) leuven - change in lighting [48]
4.3.2 Results and Discussion

4.3.2.1 Number of matching keypoints

The comparison results from the face dataset is given by the box-whisker plot that shows the overall variation in the dataset. The number of matching keypoints from Tri-SIFT in scale, rotation, and viewpoint change are always greater than the number of matching keypoints from Bi-SIFT and SIFT for the face dataset as seen in Figure 40, Figure 41, and Figure 42. As shown in Figure 43 through Figure 50, in the object dataset, again the number of matches from Tri-SIFT is greater than the number of matches from Bi-SIFT and SIFT for all cases except the change in blur. This is because when the blur is increased, high frequency contents such as edges and object boundaries are removed from the scene and Tri-SIFT is not able to catch more keypoints.

Under viewpoint change, the number of matching keypoints decreases dramatically for SIFT, Bi-SIFT, and Ti-SIFT. Hence, the importance of catching more matching keypoints can easily be seen under viewpoint change.

![Box-whisker plot showing the variation of the number of matching keypoints under scale transformation for SIFT, Bi-SIFT, and Tri-SIFT.](image)

Figure 40. Variation of the number of matching keypoints under scale transformation for SIFT, Bi-SIFT, and Tri-SIFT
Figure 41. Variation of the number of matching keypoints under rotation transformation for SIFT, Bi-SIFT, and Tri-SIFT.

Figure 42. Variation of the number of matching keypoints under 35 degree view-point change for SIFT, Bi-SIFT, and Tri-SIFT.
Figure 43. Number of matching keypoints under scale changes for the textured scene (bark images)

Figure 44. Number of matching keypoints under scale changes for the structured scene (boat images)
Figure 45. Number of matching keypoints under different viewpoint angles for the structured scene (graffiti images)

Figure 46. Number of matching keypoints under different viewpoint angles for the textured scene (wall images)
Figure 47. Number of matching keypoints under different level of blur for the structured scene (bike images)

Figure 48. Number of matching keypoints under different level of blur for the textured scene (tree images)
Figure 49. Number of matching keypoints under different level of lighting condition

Figure 50. Number of matching keypoints under different level of Jpeg compression
4.3.2.2 Warping intensity error

The warping intensity error obtained from SIFT, Bi-SIFT, and Tri-SIFT is almost the same for all of the cases except the change in viewpoint. The reason is all of the three methods are able to catch a sufficient enough number of correct matching keypoints under change in scale, rotation, light, blur, and jpeg compression. However, under change in viewpoint, the number of matching keypoints decreases intensely compared to other cases, and the computed homography matrix is not very accurate to transform the test image and warp it. Therefore, the warping intensity error is higher for the change in viewpoint angle and differs from SIFT to Bi-SIFT and to Tri-SIFT.

As seen in Figure 52, the intensity error from all three methods is almost the same up to a 50 degree viewpoint angle. At a 50 degree viewpoint angle, Tri-SIFT has the lowest warping intensity error than SIFT and Bi-SIFT. Because Tri-SIFT is able to catch enough number of good matching keypoints as seen in Figure 53, by using these matching keypoints, the reference image can be warped to the test image correctly as shown in Figure 54. On the other hand, SIFT and Bi-SIFT could not find any good matching keypoints between the reference and the test image as seen in Figure 55 and Figure 57, and they were not able to compute the accurate homography matrix and transform the test image. Therefore, the warping obtained by using the matches from SIFT and Bi-SIFT are not accurate as seen in Figure 56 and Figure 58.
Figure 51. Variation of the warping intensity error under 35 degree pose transformation for SIFT, Bi-SIFT, and Tri-SIFT

Figure 52. Intensity error under different viewpoint angles for the structured scene (graffiti images)
Figure 53. Tri-SIFT matching keypoints between the reference image and its 50 degree viewpoint transformed version. There are 49 matching keypoints to be used to compute homography matrix.

Figure 54. Mosaic image obtained by warping the test image (which has the 50 degree of viewpoint angle) to reference image using the matching keypoints found by Tri-SIFT.

Figure 55. SIFT matching keypoints between the reference image and its 50 degree viewpoint transformed version. There is no correct matching keypoints to be used to compute homography matrix.
Figure 56. Mosaic image obtained by warping the test image (which has the 50 degree of viewpoint angle) to reference image using the matching keypoints found by SIFT

Figure 57. Bi-SIFT matching keypoints between the reference image and its 50 degree viewpoint transformed version. There is no correct matching keypoints to be used to compute homography matrix.

Figure 58. Mosaic image obtained by warping the test image (which has the 50 degree of viewpoint angle) to reference image using the matching keypoints found by Bi-SIFT
4.3.2.3 Scatteredness

Three methods under consideration yield the same scatteredness for all of the cases in the face dataset. Since the Bi-SIFT and Tri-SIFT detects more keypoints than SIFT, their keypoints scattered more than those of SIFT as seen in Figure 59. Similarly, the scatteredness under scale transformation does not show a significant variation over the object dataset. However, scatteredness of the Tri-SIFT under the viewpoint transformation is greater than that of SIFT and Bi-SIFT as seen in Figure 60.

![Diagram showing scatteredness for SIFT, Bi-SIFT, and Tri-SIFT](image)

Figure 59. Variation of the scatteredness under scale transformation for SIFT, Bi-SIFT, and Tri-SIFT
Figure 60. Scatteredness under different viewpoint angles for the structured scene (graffiti images)

Figure 61. Average number of detected keypoints of the reference images in face dataset with respect to different octaves for SIFT, Bi-SIFT, and Tri-SIFT
4.3.2.4 Comparison of Gradient and Laplacian Kernels in Trilateral Filters

In this section, we compare two versions of the proposed Trilateral filter in terms of their noise performance based upon the signal-to-noise-ratio (SNR); the first version employs the gradient kernel and the second one employs the Laplacian kernel. Two sets of images are used to compare their low and high frequency behavior. The first set of images has low texture (low spatial frequency) content and the second set has high texture (high spatial frequency) content as shown in Figure 64 and Figure 65, respectively. The SNR performance is measured after adding white Gaussian noise and speckle noise to the original images. Note that the higher the value of SNR for a noise type, the better is the denoising performance of a filter. Figure 66 and Figure 67 show the results for low texture images with a Gaussian noise and speckle noise, respectively. The results for high texture images are presented in Figure 68 and Figure 69.

Figure 62. Average percentage of detected keypoints of the reference images in face dataset with respect to different octaves for SIFT, Bi-SIFT, and Tri-SIFT
As seen from the results presented in the figures, the Laplacian kernel of the Trilateral filter is not effective in low texture images compared to the gradient kernel, which is slightly more effective. On the other hand, in high texture images, the Laplacian kernel shows its performance in terms of denoising. This results is expected as we pointed out earlier in this paper; the Laplacian makes improvements at high variation since it measures the speed of variation whereas the gradient measures the value (slope) of variation. Within a low frequency neighbor, the speed of variation almost vanishes; therefore, the weights produced by the Laplacian kernel will be almost unity. This behaviour makes the Trilateral filter respond as a Bilateral filter. On the contrary, the weights of a gradient filter are still able to penalize variation around edges; this behaviour of the Trilateral filter with gradient kernel produces sharper images in the presence of noise. Figure 63 illustrates the behaviour of the Laplacian and gradient kernels over high and low frequency regions.

Figure 63. The effect of Laplacian and gradient kernels over a high and low frequency regions
Figure 64. Low texture images (left to right): Bird [49], duck [50], and kids [51]

Figure 65. High texture images (left to right): Rock [52], wall [53], and slate [54]

Figure 66. Comparison of spatial, Bilateral, and Trilateral filters over low texture images with Gaussian noise
Figure 67. Comparison of spatial, Bilateral, and Trilateral filters over low texture images with speckle noise.

Figure 68. Comparison of spatial, Bilateral, and Trilateral filters over high texture images with Gaussian noise.
Figure 69. Comparison of spatial, Bilateral, and Trilateral filters over high texture images with speckle noise.

Figure 70 compares the effectiveness of the Laplacian Trilateral filter versus the spatial-only (Gaussian) filter, Bilateral filter, and Gradient Trilateral Filter in the presence of speckle noise blended to a high texture image. The Gaussian filter has a high blurring effect that results in removal of the high frequency content. On the other hand, the Bilateral and Trilateral filters with gradient kernel preserve some of the high frequency contents; whereas, the Laplacian Trilateral filter gives the sharper image preserving the high frequency content more than the Gradient Trilateral Filter.
Figure 70. Effect of Laplacian Trilateral filter over a high textured image. (a) Input image, (b) input image with speckle noise, (c) Gaussian filtered image, (d) Bilateral filtered image, (e) Gradient Trilateral filtered image, (f) Laplacian Trilateral filtered image

4.4 Conclusion

In this study, we analyze frequency spectrum of the SIFT, Bi-SIFT, and Tri-SIFT and propose the Laplacian Trilateral filter. These two versions of SIFT have a nonlinear character alternative to SIFT’s linear Gaussian scale-space. The proposed filter that has a nonlinear characteristic adaptively adjusts itself according to the structure of the texture using two range (intensity) filters.

We compare these three versions of the SIFT with respect to the number of matching keypoints, warping intensity error, and scatteredness of detected keypoint locations. The frequency spectrum analysis shows that Tri-SIFT operates at higher frequency bands than Bi-SIFT and SIFT; the SIFT has the lowest band-pass filter range. Consequently, Tri-SIFT is able to detect more keypoints, hence more matching, than the Bi-SIFT and SIFT. According to experimental results, the SIFT,
Bi-SIFT, and Tri-SIFT algorithms can extract sufficient enough matching keypoints to compute a homography matrix that can register a reference image to its test image under the conditions of scale, rotation, and viewpoint angle up to 40 degrees, lighting, and jpeg compression. We found that SIFT and Bi-SIFT fail to find good matching keypoints when there is a change in viewpoint angle of more than 40 degrees. However, the Tri-SIFT algorithm can find a good number of matching keypoints under change in viewpoint angle up to 50 degrees. Based on the experimental results, the warping intensity error is almost the same for all three filters for all of the cases that we have experimented except the change in viewpoint angle. Tri-SIFT has less intensity error than that of the Bi-SIFT and SIFT in the case of change in viewpoint angle. Hence, one can conclude that the Tri-SIFT algorithm is more robust to the change in viewpoint angle than the SIFT and Bi-SIFT.

As a future work, the comparison can be extended to measuring classification accuracy of these three SIFT versions under variety of conditions. In addition, the effect of anisotropic smoothing can be explored for image enhancement. In this paper, we formulated the frequency spectrum of the Bilateral and Trilateral filters for an isotropic region. A similar analysis would be conducted for an anisotropic neighbor. The Bilateral and Trilateral filtering takes longer time than the Gaussian filtering. As an example, obtaining features from a 200x200 face image takes 0.90, 52.3, 62.6 seconds by the SIFT, Bi-SIFT, and Tri-SIFT, respectively. Accordingly, the computation time is significantly higher for Bi-SIFT and Tri-SIFT than that of the SIFT. This drawback of the Bi-SIFT and Tri-SIFT can be compensated by utilizing parallel computation to benefit their performance in terms of accuracy. Parallel implementation of these filters can be studied as an open problem.
CHAPTER 5. EXPERIMENTS AND RESULTS

5.1 Data Preparation

5.1.1 Data

Publicly available lung CT scan database created by The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) [55] is used to test the proposed CAD framework. LIDC/IDRI database contains 1010 CT scans which has the annotations for the nodules and the non-nodules which has diameter $\geq 3$mm. Annotations made by the radiologists belong to one of these three groups; nodule $\geq 3$mm, nodule $\leq 3$mm, or non-nodule $\geq 3$mm. The annotations of the CT scans are done by 4 expert radiologists in 2 phases, blinded-read phase and unblinded-read phase. In the initial blinded-read phase, each of the radiologists examined the scans independently without knowing the opinion of the others and the second unblinded-read phase, they examined the CT scans while knowing the annotations made by 3 other radiologists. While the surrounding boundary for the nodules $\geq 3$mm are annotated, the nodules $\leq 3$mm or non-nodules $\geq 3$mm has only their volume center is annotated. In the experiments, 100 CT scans were used from LIDC/IDRI dataset.

5.1.2 Extraction of Volume of Interest and Slice Selection

In the annotation of nodules $\geq 3$mm, since each radiologist marks the surrounding boundary of the nodules, the volume center of the same nodule might differ from one radiologist to another. Therefore, as an initial step of extraction of VOI, the volume center of each annotated nodule is computed based on the provided annotations by each radiologist. If the center coordinates of nodules, annotated by different radiologists, are closer than the threshold, they are assumed to be the same nodule. Hence, at the next step, the average volume center for each nodule with the number of radiologists’ approval is found. Similar approach is used for detecting the average
volume center and the number of radiologists’ approval for the non-nodules ≥ 3mm. There is a possibility that some of the objects might be annotated as nodule by one radiologist and non-nodule by the other(s) or vice versa. To overcome this problem, once the average volume centers are computed for nodules and non-nodules, if the volume centers of nodule and the non-nodule are closer than the threshold, they are eliminated from the dataset. After the volume centers of the objects are determined, 30x30x30 mm³ region around the volume center is extracted as the volume of interest. The reason for using 30x30x30 mm³ bounding cube is in the dataset the longest axis of the annotated largest nodule can be 30mm as provided in [55]. In LIDC-IDRI dataset, CT scans are collected from different CT scanners. Although all slices from all scans are 512x512 pixels, the physical size of a single pixel is not the same for all scans. Thus, 30x30x30 mm³ bounding cube corresponds to different size of pixel resolution. However, the input data for training and testing the proposed MPF model should be the same size. Therefore, all extracted 30x30x30 mm³ are normalized to the maximum resolution of 56x56x56 pixels.

Figure 71. 3D extracted volume of interest is given on the left and 2D transverse seen of the slice is given on the right
Since the bounding box is used and the nodules are not segmented out, within the extracted nodule volume of interest, there would be some slices which are not belong to the nodule and they should be removed from the nodule VOI. However, removing the slices which are not belong to the nodule may result different size of the input data. For instance, one volume of interest can have 10 slices not belong to the nodule and on the other hand the other volume of interest can have 20 slices not belong to the nodule and removing these slices will cause the different size of the input problem. To overcome this problem, the smallest nodule found within the data set and the number of slices belong to that nodule is found. So, if we select the same number of slices form each nodule VOI as the number of slices belong to the nodule from the smallest nodule VOI, this guarantees that we will end up having the same number of slices in each nodule VOI, and all selected slices will belong to the nodule. In the dataset used in this dissertation, the smallest number of slices belong to the nodule found as 6. Hence, from each nodule VOI, 6 slices are selected from each perspective. These 6 slices can be selected in a different way. One way could be selecting 6 slices from the center of the VOI. However, in this approach, if the nodule size is big, then there is a high chance of ending up selecting the similar slices, and this might be a disadvantage because they are not going to give any distinct information from slice to slice. Another approach that is used in this dissertation is selecting the slices from starting of the nodule to the end of the nodule by the equal intervals. So that as much as distinct information from slice to slice is preserved.

5.2 Performance Assessment Measures

Performance of the binary classifier can be measured by the confusion matrix given in Table 1 composed of four components; true positives (TP) which are the positive samples predicted as positive, false positives (FP) which are negative samples predicted as positive, true negatives (TN) which are negative samples predicted as negative and false negatives (FN) which are positive
samples predicted as negative. FP are also known as Type I error and FN are known as Type II error. Components of confusion matrix are used to obtain different types of performance measurement such as accuracy (ACC), true positive rate (TPR), false positive rate (FPR), true negative rate (TNR), precision, F score, receiver operating characteristic (ROC) curve etc. In this dissertation, ACC, area under the ROC curve, F-1 score, sensitivity, and specificity are used for the performance assessment of the proposed method.

Table 1. Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>Actual Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
</tr>
<tr>
<td>Predicted Class</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>True Positive (TP)</td>
</tr>
<tr>
<td>Negative</td>
<td>False Negative (FN) (Type II error)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>$P = TP + FN$</td>
</tr>
</tbody>
</table>

ACC, given in (45), is the total number of correctly classified samples out of total number of samples. If the sample data is unbalanced, then the accuracy might not be sufficient to measure the performance of the classifier.

$$ACC = \frac{TP + TN}{P + N} \tag{45}$$

TPR also known as sensitivity or recall measures how many positive samples are predicted as positives out of all positive samples. So, FP are not considered in recall measurement. However, precision measures how many positive samples are predicted as positive out of all samples predicted as positive. Therefore, using recall with precision might give a better idea about the
performance of the classifier, F-1 score which is the harmonic mean of the precision and the recall employs both the precision and the recall for measuring the performance of the classifier. Recall, precision and F-1 score are defined as

\[ \text{Recall} = \frac{TP}{P} \]  
\[ \text{Precision} = \frac{TP}{TP + FP} \]  
\[ F - 1 \text{ Score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \]

Similar to TRP, TNR also known as specificity measures how many negative samples are predicted as negative out of all negative samples. Specificity is defined as

\[ \text{Specificity} = \frac{TN}{N} \]

Another performance measurement is FPR which measures how many negative samples are predicted as positive out of all negative samples and it is defined as

\[ FPR = \frac{FP}{N} \]

Plotting the sensitivity versus 1-sensitivity by setting a multiple threshold for the output of the classifier is called ROC curve and the area under the ROC curve (AUC) is an effective method for measuring the performance of the diagnostic test. In other words, ROC curve provides the relation between the TPR and FPR for the classification. When the classification is perfect, AUC equals to 1. Therefore, obtaining an AUC close to 1 can be interpreted as good classification performance.

5.3 Experimental Results of MPF Model

The dataset used to train and to test the model is created using 100 CT scans from LIDC/IDRI database, and it contains the nodules and non-nodules approved by at least one radiologist. Data set is balanced and there are total of 604 nodule and non-nodule objects. Dataset is split into 2
parts, 70% for the training and 30% for the testing. Therefore, training data has total of 422 nodules and non-nodules, and the testing data has total of 182 nodules and non-nodules.

Figure 72 shows the change of slice level classification performances across different perspectives for MPF model which uses the raw slices from the extracted volume of interest. Although the slices from YZ-perspective gives the highest ACC, AUC, F1-score and sensitivity, specificity of the model created using slices from YZ-perspective is the smallest. On the other hand, ACC, AUC, F1-score and sensitivity of the model created using XY-slices is smallest among all 3 models. But the specificity of the model created using XY-slices is highest among all 3 models. These results also can be interpreted as the model uses the slices from YZ-perspective has higher tendency towards type-I error and has higher FP. Alternatively, the model uses the slices from XY-perspective has higher tendency towards type-II error and has higher FN. ROC curves across different perspectives for slice level classification for MPF model is shown in Figure 73.

Figure 72. MPF Model – change of slice level classification performances across different perspectives
Figure 73. MPF Model – ROC curves across different perspectives for slice level classification

Figure 74. MPF Model - missed nodules (FN)
After fusing the class scores from the slice level classification and obtaining the perspective level classification, except the specificity of the model uses the slices from XZ-perspective, all of the performance scores for all perspectives are increasing. At the perspective level classification, while the model uses slices from XY-perspective still has the lowest type-I error and the highest type-II error, the tendency toward type-I error of the model that uses YZ-perspectives is decreasing. At the perspective level classification, still the model which uses the slices from YZ-perspective has the highest performance score of ACC, AUC, F1-Score and sensitivity. ROC curves across different perspectives for perspective level classification for MPF model is shown in Figure 77.
Figure 76. MPF Model – change of perspective level classification performances across different perspectives

![Bar Chart]

<table>
<thead>
<tr>
<th>Perspective</th>
<th>ACC</th>
<th>AUC</th>
<th>F1-Score</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>XY</td>
<td>79%</td>
<td>81%</td>
<td>80%</td>
<td>80%</td>
<td>78%</td>
</tr>
<tr>
<td>XZ</td>
<td>76%</td>
<td>85%</td>
<td>80%</td>
<td>89%</td>
<td>62%</td>
</tr>
<tr>
<td>YZ</td>
<td>82%</td>
<td>88%</td>
<td>84%</td>
<td>90%</td>
<td>74%</td>
</tr>
</tbody>
</table>

Figure 77. MPF Model – ROC curves across different perspectives for perspective level classification

![ROC Curve]

Change of classification performance for slice, perspective and volume level classifications for each perspective for MPF model is shown in Figure 78. Increase of the classification performance
from slice level classification to perspective level classification and volume level classification can be seen in Figure 78. Slices level classification gives the highest ACC as 75% using the slices from YZ-perspective. When the class scores from multiple slices are fused at the perspective level, the highest classification ACC is increasing from 75% to 82%. Finally adding another hierarchical fusion level which fuses the class scores from all perspectives increasing the highest classification ACC from 82% to 87%. Similarly, AUC, F1-score, sensitivity and specificity scores are also increasing from slice level classification to perspective level classification and volume level classification. At the volume level classification both the tendency toward type-I error and type-II error are the same while having the 87% sensitivity and specificity. Comparison of ROC curves for slice, perspective and volume level classifications for the slices from XY-perspective, XZ-perspective and YZ-perspectives are shown in Figure 79, Figure 80 and Figure 81.
Figure 79. MPF Model – Comparison of ROC curve for slice, perspective and volume level classifications for the slices from XY-perspective

Figure 80. MPF Model – Comparison of ROC curve for slice, perspective and volume level classifications for the slices from XZ-perspective
Figure 81. MPF Model – Comparison of ROC curve for slice, perspective and volume level classifications for the slices from YZ-perspective

### 5.4 Experimental Results of SFMPF Models

In this study, four different SFMPF models based on Bilateral, Trilateral, Gabor and LOG filters are experimented using the same dataset which is used for MPF model. In the SFMPF model, featured images are created by filtering the raw slices from the extracted volume of interest by the aforementioned filters. Once the featured image dataset is obtained for each proposed SFMPF model, the same approach as MPF is taken to create the model for slice level classification, perspective level classification and the volume level classification.

#### 5.4.1 SFMPF Model Based on Bilateral Image

Similar to MPF model, increase of the classification performance from slice level classification to perspective level classification and volume level classification can be seen in Figure 82 for the Bilateral image based SFMPF model. At the slice level classification, the highest performance
score of ACC, AUC, F1-score and sensitivity achieved by the model uses the slices from YZ-perspective. Compared to MPF model, Bilateral image based SFMPF model accomplishes a slight improvement with respect to the highest ACC, AUC, F1-Socre, and specificity for the slice level classification. However, at the perspective and the volume level classifications, MPF model achieves slightly better performance than the Bilateral image based SFMPF for all of the performance measures except AUC. Since the highest AUR performance obtained from the model uses slices from YZ-perspective for the slice and the perspective level classifications, comparison of ROC curves for slice, perspective and volume level classifications for the slices from YZ-perspective is provided in Figure 83.

![Figure 82. SFMPF Model using Bilateral Image – Change of classification performance for slice, perspective and volume level classifications for each perspective](image-url)
Figure 83. SFMPF Model using Bilateral Image – Comparison of ROC curve for slice, perspective and volume level classifications for the slices from YZ-perspective

Figure 84. SFMPF Model based on Bilateral Image - missed nodules (FN)
Figure 85. SFMPF Model based on Bilateral Image - missed non-nodules (FP)

5.4.2 SFMPF Model Based on Trilateral Image

Change of the classification performance for slice, perspective and volume level classifications for each perspective is provided in Figure 86 for the Trilateral image based SFMPF model. Throughout the hierarchical fusion, the highest performance scores are increasing from slice to volume level classification such as ACC increases from 75% to 85%, AUC increases from 83% to 91%, F1-score increases from 76% to 86%, and sensitivity increases from 76% to 87%. The performance improvement from slice to volume level classification also can be seen in Figure 87 which shows the comparison of ROC curve for slice, perspective and volume level classifications. Although the proposed hierarchical fusion approach works well with the Trilateral image based SFMPF model, overall performance of MPF model achieves slightly better performance than the SFMPF model based on Trilateral image. Tuning the parameters of the Trilateral filter such as spatial, range and the Laplacian kernels’ standard deviation might improve the classification performance of the Trilateral image based SFMPF model.
Figure 86. SFMPF Model using Trilateral Image – Change of classification performance for slice, perspective and volume level classifications for each perspective.

<table>
<thead>
<tr>
<th></th>
<th>XY-Perspective</th>
<th>XZ-Perspective</th>
<th>YZ-Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Slice</strong></td>
<td>74% 79% 72%</td>
<td>77% 73% 71%</td>
<td>75% 76% 76% 75%</td>
</tr>
<tr>
<td><strong>Perspective</strong></td>
<td>81% 82% 81%</td>
<td>84% 87% 87%</td>
<td>83% 80% 86%</td>
</tr>
<tr>
<td><strong>Volume</strong></td>
<td>85% 85% 85%</td>
<td>84% 84% 84%</td>
<td>85% 86% 87%</td>
</tr>
</tbody>
</table>

Figure 87. SFMPF Model using Trilateral Image – Comparison of ROC curve for slice, perspective and volume level classifications for the slices from YZ-perspective.
5.4.3 SFMPF Model Based on Gabor Image:

Classification performance improvement from the hierarchical fusion approach in Gabor image based SFMPF model can be seen in Figure 92 and the change of ROC curve for slice, perspective and volume level classifications can be seen in Figure 93. The highest ACC in slice level classification increases from 77\% to 92\% at the volume level classification. While the highest sensitivity in slice level increases from 78\% to 92\%, the highest specificity is increased from 76\% to 79\% at the volume level. Hence, the Gabor image based SFMPF model has a higher tendency toward type-I error compared to MPF model. On the other hand, Gabor image based SFMPF model has higher sensitivity of 92\% compared to sensitivity of MPF model which is 87\%. Although, in
the literature for texture extraction, Gabor filter is used as a filter bank which composed of multiple Gabor filters in different frequencies and angles, in the proposed Gabor Image based SFMPF model, single Gabor filter is used to create the feature image. Using multiple Gabor filters with different frequencies and angles and then fusing them at the volume level may increase the performance of proposed Gabor image based SFMPF model.

Figure 90. SFMPF Model based on Gabor Image - missed nodules (FN)

Figure 91. SFMPF Model based on Gabor Image - missed non-nodules (FP)
Figure 92. SFMPF Model using Gabor Image – Change of classification performance for slice, perspective and volume level classifications for each perspective

Figure 93. SFMPF Model using Gabor Image – Comparison of ROC curve for slice, perspective and volume level classifications for the slices from YZ-perspective

5.4.4 SFMPF Model Based on LOG Image

Similar to other proposed SFMPF models and MPF model, proposed hierarchical fusion based deep learning approach significantly increases the performance of the classification result for the
LOG image based SFMPF. Change of classification performance from slice to perspective and volume level classifications for each perspective is shown in Figure 94 and Figure 95. The highest ACC at slice level classification increases from 78% to 85%, the highest AUC increases from 85% to 95% and the highest sensitivity increases from 79% to 94% by hierarchically fusing the class scores from all perspectives at the volume level classification. LOG image based SFMPF model achieves the sensitivity of 94% and the specificity of 80% at the volume level. Compared to MPF model, it has higher sensitivity as well as higher tendency toward type-I error. Missed nodules which are predicted as non-nodules (FN) by the LOG image based SFMPF model is provided in Figure 96 and missed non-nodules which are classified as nodules (FP) are provided in Figure 97.

---

**Figure 94. SFMPF Model using LOG Image – Change of classification performance for slice, perspective and volume level classifications for each perspective**
Figure 95. SFMPF Model using LOG Image – Comparison of ROC curve for slice, perspective and volume level classifications for the slices from YZ-perspective

Figure 96. SFMPF Model based on LOG Image - missed nodules (FN)
5.4.5 Classification Performance Comparison of SFMPF Models and MPF Model

In this section, the classification performances of the proposed feature image based SFMPF models and the MPF model are compared with respect to ACC, AUC, F1-score, sensitivity and specificity. Firstly, the comparison between the proposed SFMPF models and then the comparison between the MPF model and SFMPF models is done.

Change of the average slice level classification performance of the proposed classifiers over three perspectives is given in Figure 98. Classification performance with respect to ACC, F1-score and sensitivity, Trilateral image based SFMPF model has slightly lower performance compared to other feature image based SFMPF models. On the other hand, LOG image based SFMPF model has the highest ACC, AUC, F1-score and sensitivity among all. Whereas, the Bilateral image based SFMPF model has the highest specificity and lowest FPR, both Bilateral and Trilateral image based SFMPF models has the higher tendency toward type-II error and higher false negative rate (FNR) compared to Gabor and LOG image based SFMPF models.
As shown in Figure 98, proposed feature image based SFMPF models improve the classification performance compared to MPF model in terms of ACC, AUC, F1-score, sensitivity and specificity. Particularly, LOG image based SFMPF model, while it increases the sensitivity compared to MPF model, it keeps the specificity same. This shows that, whereas the LOG image based SFMPF model increases the TPR, FPR remains the same. For more detailed comparison, change of the slice level classification performances of the proposed models for each perspective is given in Figure 99.

![Figure 98](image)

Figure 98. Change of average slice level classification performance of the proposed classifiers over three perspectives

At the second level, perspective level, in the proposed hierarchical fusion scheme, again the proposed feature image based SFMPF models outperform the MPF model in terms of all performance measurements as depicted in Figure 100. At the slice level classification, except the LOG image based SFMPF model, all other proposed models either has the same sensitivity and specificity performances or higher specificity performances. In contrast to slice level classification, all proposed models including the LOG image based SFMPF model achieve the higher sensitivity compared to specificity, and they all have higher tendency toward type-II error at the perspective
level classification. This means that the nodule prediction performances of the proposed models are better than the non-nodule prediction performances at perspective level classification. Particularly, Bilateral image based SFMPF model increases the sensitivity of 73% to 83% and specificity of 77% to 78% at the perspective level and LOG image based SFMPF model increases the sensitivity of 77% to 87% and specificity of 74% to 76% at the perspective level. Whereas, the proposed MPF model improves the sensitivity of 74% to 86% at and decreases the specificity from 74% to 71% the perspective level. Therefore, one can conclude that the proposed feature image based SFMPF models not only improves the sensitivity but also improves the specificity while MPF model increases the sensitivity and decreases the specificity at the perspective level classification. For more detailed comparison, change of the perspective level classification performances of the proposed models for each perspective is given in Figure 101.

![Figure 99. Change of slice level classification performances for proposed SFMPF and MPF Models for each perspective](image-url)
Figure 100. Change of average perspective level classification performance of the proposed classifiers

Figure 101. Change of perspective level classification performances for each SFMPF Model for each perspective

At the final level, volume level, classification, the classification performances of all measures are increasing for all the proposed methods. MPF model achieves the sensitivity and specificity of 87% at the volume level. If we compare the sensitivity and specificity performance of the first (slice) level and the last (volume) level classifications, MPF model increases the sensitivity and
specificity of 74% to 87%, Bilateral image based SFMPF model increases the sensitivity of 73% to 85% and specificity of 77% to 87%, Trilateral image based SFMPF model increases the sensitivity of 72% to 87% and specificity of 76% to 84%, Gabor image based SFMPF model increases the sensitivity of 75% to 92% and specificity of 75% to 79%, and LOG image based SFMPF model increases the sensitivity of 77% to 94% and specificity of 74% to 80%. As seen from the results, at the final level of classification, while LOG image based SFMPF model predicts with the highest sensitivity, MPF model predicts with the highest specificity. Similar to slice and perspective level classification, LOG image based SFMPF model outperforms the other feature image based SFMPF models as well as MPF model with respect to AUC, F1-score, and sensitivity. Whereas the ACC, AUC, F1-score, and sensitivity increases through the proposed hierarchical fusion scheme, tendency toward type-I and type-II errors of the proposed models vary from layer to layer.

<table>
<thead>
<tr>
<th></th>
<th>ACC</th>
<th>AUC</th>
<th>F1-Score</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPF</td>
<td>87%</td>
<td>92%</td>
<td>88%</td>
<td>87%</td>
<td>87%</td>
</tr>
<tr>
<td>Bilateral</td>
<td>86%</td>
<td>93%</td>
<td>86%</td>
<td>85%</td>
<td>87%</td>
</tr>
<tr>
<td>Trilateral</td>
<td>85%</td>
<td>91%</td>
<td>86%</td>
<td>87%</td>
<td>84%</td>
</tr>
<tr>
<td>Gabor</td>
<td>86%</td>
<td>93%</td>
<td>87%</td>
<td>92%</td>
<td>79%</td>
</tr>
<tr>
<td>LOG</td>
<td>87%</td>
<td>95%</td>
<td>89%</td>
<td>94%</td>
<td>80%</td>
</tr>
</tbody>
</table>

Figure 102. Change of volume level classification performances for the proposed classifiers
5.5 Experimental Results of MFMPF Model

The idea behind the MFMPF model is first making multiple decisions for an object using different type of features by looking from different perspectives. Then fusing each of the decisions made based on different features to make the final decision. Therefore, all previously proposed feature image based SFMFP models and the basic MPF model which uses raw slices are fused to obtain the MFMPF model. By adding another of hierarchy, the class scores obtained at the final (volume) layer of each SFMPF models are in MFMPF model. Results from MFMPF model with the MPF model and the best performing SFMPF models based on Gabor and LOG images are given in Figure 106. MFMPF model outperforms all proposed feature image based SFMFP models as well as MPF model with respect all performance measures except specificity. Although proposed MFMPF model doesn’t perform better than the MPF model for detecting non-nodules, since it has higher sensitivity and same specificity compared to MPF. We can conclude that while
the TPR increases, FPR remains the same in the proposed MFMPF model and it performs better than all other proposed models.

![Figure 104. MFMPF Model based on Fusion of all - missed nodules (FN)](image)

![Figure 105. MFMPF Model based on Fusion of all - missed non-nodules (FP)](image)

![Figure 106. Change of final layer classification performance of MPF, SFMPF based on Gabor and LOG, and MFMPF models](image)

<table>
<thead>
<tr>
<th></th>
<th>ACC</th>
<th>AUC</th>
<th>F1-Score</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPF</td>
<td>87%</td>
<td>92%</td>
<td>88%</td>
<td>87%</td>
<td>87%</td>
</tr>
<tr>
<td>Gabor</td>
<td>86%</td>
<td>93%</td>
<td>87%</td>
<td>92%</td>
<td>79%</td>
</tr>
<tr>
<td>LOG</td>
<td>87%</td>
<td>95%</td>
<td>89%</td>
<td>94%</td>
<td>80%</td>
</tr>
<tr>
<td>Fusion of all</td>
<td>91%</td>
<td>96%</td>
<td>92%</td>
<td>95%</td>
<td>87%</td>
</tr>
</tbody>
</table>
CHAPTER 6. CONCLUSION AND FUTURE WORK

Lung cancer is the leading cancer type in terms of causing the mortality in both men and women. As reported in previous studies, screening the lung cancer using CT scans is very common and effective method. However, detecting pulmonary nodules in CT scans is a very challenging problem, particularly for the nodules in their early stages. CAD systems can be used by the radiologists during the examination of CT scans to increase the nodule detection rate as well as to decrease the false positives.

6.1 Contributions and Findings

In this research, multi-perspective hierarchical fusion based deep learning model is proposed for lung nodule detection from CT scans. Three different types of multi-perspective hierarchical fusion based deep learning models namely, multi-perspective fusion based deep learning (MPF) model, single-feature multi-perspective fusion based deep learning (SFMPF) model, and multi-feature multi-perspective fusion based deep learning (MFMPF) model, are proposed.

Contributions of this research can be summarized as following;

1- Synthesizing the decisions made by multiple deep learning models by looking at multiple slices from multiple perspectives in a hierarchical manner:

The proposed multi-perspective hierarchical fusion based deep learning model, MPF, employs three level of multi-perspective hierarchical fusion based classification. The idea behind the proposed model is first deciding based on each slice from each perspective. Then, the decisions made for each slice are fused to obtain a perspective level prediction. This approach is applied for all perspectives and a single class scores are obtained for each perspective. At the final layer, the decisions made for each perspective are fused to obtain a single volume level prediction for each extracted 3D VOI. To test the classification performance of the proposed
MPF model, total of 604 nodule and non-nodule objects are extracted from 100 CT scans and 70% of the data is used to train the proposed model and 30% of the data is used to test the proposed model. Experimental results show that the proposed hierarchical fusion based deep learning model achieved the ACC of 74%, AUC of 81%, sensitivity of 74%, and specificity of 74% at the first, slice, level classification, ACC of 79%, AUC of 85%, sensitivity of 86% and specificity of 71% at the second, perspective, level classification, and ACC of 87%, AUC of 92%, sensitivity of 87% and specificity of 87% at the final, volume, level classification. As seen from the results, proposed multi-perspective hierarchical fusion approach increases all the classification performance measures significantly from slice level to volume level.

2- Fusing the decisions of multiple deep learning models based on supervised learning:

At the first layer of MPF model, once the class scores are predicted for each slice, the class scores of the slices coming from the same object are concatenated and a new feature vector is created for that object in the size of the number of slices of an object. Then, these new set of feature vectors are used to train another classifier which is a one hidden layer ANN to obtain a class score for the second, perspective, level. The same process is applied for each perspective. Therefore, at the perspective level, three different class scores are predicted for each object. After that, these three class scores coming from three different perspectives are concatenated to create a new feature. Similarly, another one hidden layer ANN classifier is trained to obtain a single class score for each object at the final, volume, level. Based on the experimental results, proposed supervised learning based hierarchical fusion of deep learning models increases the classification performance considerably.

3- Feature image based multi-perspective hierarchical fusion based deep learning model:
Four different feature images are created using Bilateral, Trilateral, Gabor and LOG filters. Then each type of feature image is used with the proposed MPF model. According to experimental results, using feature images instead of raw slices increases the classification performance at all levels in terms of certain performance measures. Particularly, LOG image based SFMPF model increases the AUC of 92% to 95%, and sensitivity of 87% to 94% compared to MPF model at the volume level classification.

4- Adding another level of hierarchy to proposed SFMPF model by fusing the decisions made by multiple SFMPF models using proposed supervised learning based fusion:

Another level of classification is added to the proposed hierarchical model by fusing the volume level class scores obtained from each of four SFMPF models and MPF model with an additional ANN classifier. This additional level of classification increases the ACC of 87% to 91%, AUC of 92% to 96%, F1-score of 88% to 92% and sensitivity of 87% to 95% compared to MPF model.

5- Utilizing newly proposed Trilateral filter to obtain feature images and using them in proposed SFMPF and MFMPF models:

Trilateral image based SFMPF model which utilizes the newly proposed Trilateral filter is proposed. Despite the classification performance of MPF model and the Trilateral image based SFMPF model are almost same, tuning the parameters of the Trilateral filter might help to increase the classification performance of the SFMPF model based on Trilateral image.

6- Using newly proposed Trilateral filter to create anisotropic version of SIFT called Tri-SIFT:

Trilateral filter is formed by adding a Laplacian kernel as a second range kernel in addition to spatial and range kernels in the Bilateral filter. Trilateral filter that has a nonlinear characteristic adaptively adjusts itself according to the structure of the texture using two range (intensity)
filters. Thus, it is more effective at higher frequency structures. Tri-SIFT is created by replacing the isotropic Gaussian scale-space in SIFT with the proposed anisotropic Trilateral scale-space.

7- Exploring the frequency behavior of isotropic (regular SIFT) and anisotropic (Bi-SIFT and Tri-SIFT) versions of the scale-space keypoint detection algorithm SIFT: The behavior of the difference of Bilateral and the difference of Trilateral filters in low and high frequency regions are analyzed by using their frequency spectrum obtained with the Fourier Transform. The frequency spectrum analysis shows that Tri-SIFT operates at higher frequency bands than Bi-SIFT and SIFT; the SIFT has the lowest band-pass filter range. Consequently, Tri-SIFT is able to detect more keypoints, hence more matching, than the Bi-SIFT and SIFT.

8- Exploring the invariance of SIFT, Bi-SIFT, and Tri-SIFT under the change of scale, rotation, viewpoint, lighting, and jpeg compression:

According to experimental results, other than the change in viewpoint angle, SIFT, Bi-SIFT, and Tri-SIFT algorithms can extract sufficient enough matching keypoints to compute a homography matrix that can register a reference image to its test image. We found that SIFT and Bi-SIFT fail to find good matching keypoints when there is a change in viewpoint angle of more than 40 degrees. However, the Tri-SIFT algorithm can find a good number of matching keypoints under change in viewpoint angle up to 50 degrees. Therefore, we concluded that the Tri-SIFT algorithm is more robust to the change in viewpoint angle than the SIFT and Bi-SIFT.
6.2 Future Work

Although there is a significant classification performance improvement from slice level to perspective level and perspective level to volume level classification and also MPF model to SFMPF model and MFMPF model in the proposed hierarchical fusion based deep learning approach, we noticed that there are open problems and improvements are waiting to be explored as such:

1- 3D Rotation invariance:

If the nodule is not circular and elongated toward one of the axis, and if the rotated version of the similar nodule exists in another sample, most likely it is not classified as nodule since DCNN is not rotation invariant. Therefore, making the proposed model robust to change in ration can increase the classification performance. One way of making the proposed model invariant to rotation is normalizing the orientation of the input data to the same angle. This can be done by fitting an ellipse to each input sample and finding the orientation of the elongated axis and normalizing them to the same angle. However, this proposed method requires a segmentation of the nodule and the non-nodule objects to be able to fit an ellipse to find the initial orientation of the object.

2- Fusion of feature images at the slice or perspective level instead of at the volume level:

Proposed MFMPF model synthesis the class score at the volume level such that once the final volume level class scores are obtained for each SFMPF model, then with an additional ANN, these volume level class scores are fused. Another way of synthesizing the predictions from different feature images can be as follow; first the class scores can be obtained for each slice and for each perspective using different types of feature images.
Then, the class scores from each slice and from each of the feature images for each perspective can be fused at the slice level.

3- Using multiple Gabor filters with different orientation and scale:

In the proposed SFMPF model based on Gabor images, only single scale and orientation Gabor filter is used. To cover more structures with different orientations and scales, multiple Gabor filters can be used to create multiple Gabor image based SFMPF models. Then, the final class scores from each model can be fused with an additional layer of ANN classifier.

4- DoB and DoT image based SFMPF models:

DoG is an approximation to LOG, and according to the experimental results, LOG image based SFMPF models performs better than the other proposed SFMPF models. As shown in CHAPTER-4, behavior of DoB and DoT is similar to DoG. Therefore, DoB and DoT image based SFMPF models can be explored.

5- Using Long-Short Term Memory (LSTM) for classification after extracting the features with DCNN:

LSMT is a type of recurrent neural network (RNN) which is used for prediction of the time dependent sequence data such as video. For video content recognition, DCNN+LSTM based deep learning has been explored. In the CT scans, although there is no time dependency between the slices, there is a spatial correlation between the slices. Therefore, DCNN+LSTM network can be explored for CT scans and it can be used in the proposed hierarchical fusion based deep learning approach.
REFERENCES


[52] [Online]. Available: https://c1.staticflickr.com/5/4001/4648662998_5ef1a56fa3_z.jpg.


APPENDIX: PERMISSIONS

Reprint Permission Letter from Springer is given below.

This Agreement between Louisiana State University -- Kazim Sekeroğlu ("You") and Springer ("Springer") consists of your license details and the terms and conditions provided by Springer and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>License Number</th>
<th>4175181188918</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Aug 24, 2017</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>Springer</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Machine Vision and Applications</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Comparison of SIFT, Bi-SIFT, and Tri-SIFT and their frequency spectrum analysis</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>Kazim Şekerogiğli</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Jan 1, 2017</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Portion</td>
<td>Full text</td>
</tr>
<tr>
<td>Number of copies</td>
<td>10</td>
</tr>
<tr>
<td>Author of this Springer article</td>
<td>Yes and you are a contributor of the new work</td>
</tr>
<tr>
<td>Order reference number</td>
<td></td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>HIERARCHICAL FUSION BASED DEEP LEARNING FRAMEWORK FOR LUNG NODULE CLASSIFICATION</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Dec 2017</td>
</tr>
<tr>
<td>Estimated size(pages)</td>
<td>120</td>
</tr>
<tr>
<td>Requestor Location</td>
<td>Louisiana State University</td>
</tr>
<tr>
<td></td>
<td>Louisiana State University</td>
</tr>
<tr>
<td></td>
<td>BATON ROUGE, LA 70803</td>
</tr>
<tr>
<td></td>
<td>United States</td>
</tr>
<tr>
<td></td>
<td>Attn: Louisiana State University</td>
</tr>
<tr>
<td>Billing Type</td>
<td>Invoice</td>
</tr>
<tr>
<td>Billing Address</td>
<td>Louisiana State University</td>
</tr>
<tr>
<td></td>
<td>Louisiana State University</td>
</tr>
<tr>
<td></td>
<td>BATON ROUGE, LA 70803</td>
</tr>
<tr>
<td></td>
<td>United States</td>
</tr>
<tr>
<td></td>
<td>Attn: Louisiana State University</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>

Introduction
The publisher for this copyrighted material is Springer. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction along with the Billing and Payment terms and conditions.

https://s100.copyright.com/AppDispatchServlet
established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com). Limited License

With reference to your request to reuse material on which Springer controls the copyright, permission is granted for the use indicated in your enquiry under the following conditions:

- Licenses are for one-time use only with a maximum distribution equal to the number stated in your request.
- Springer material represents original material which does not carry references to other sources. If the material in question appears with a credit to another source, this permission is not valid and authorization has to be obtained from the original copyright holder.
- This permission
  • is non-exclusive
  • is only valid if no personal rights, trademarks, or competitive products are infringed.
  • explicitly excludes the right for derivatives.
- Springer does not supply original artwork or content.
- According to the format which you have selected, the following conditions apply accordingly:

  **Print and Electronic:** This License include use in electronic form provided it is password protected, on intranet, or CD-Rom/DVD or E-book/E-journal. It may not be republished in electronic open access.
  **Print:** This License excludes use in electronic form.
  **Electronic:** This License only pertains to use in electronic form provided it is password protected, on intranet, or CD-Rom/DVD or E-book/E-journal. It may not be republished in electronic open access.

For any electronic use not mentioned, please contact Springer at permissions.springer@spiglobal.com.

- Although Springer controls the copyright to the material and is entitled to negotiate on rights, this license is only valid subject to courtesy information to the author (address is given in the article/chapter).
- If you are an STM Signatory or your work will be published by an STM Signatory and you are requesting to reuse figures/tables/illustrations or single text extracts, permission is granted according to STM Permissions Guidelines: http://www.stm-assoc.org/permissions-guidelines/

For any electronic use not mentioned in the Guidelines, please contact Springer at permissions.springer@spiglobal.com. If you request to reuse more content than stipulated in the STM Permissions Guidelines, you will be charged a permission fee for the excess content.

Permission is valid upon payment of the fee as indicated in the licensing process. If permission is granted free of charge on this occasion, that does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

- If your request is for reuse in a Thesis, permission is granted free of charge under the following conditions:
  This license is valid for one-time use only for the purpose of defending your thesis and with a maximum of 100 extra copies in paper. If the thesis is going to be published, permission needs to be reobtained.
  - Includes use in an electronic form, provided it is an author-created version of the thesis on his/her own website and his/her university’s repository, including UMI (according to the definition on the Sherpa website: http://www.sherpa.ac.uk/romeo/);
  - is subject to courtesy information to the co-author or corresponding author.

Geographic Rights: Scope
Licenses may be exercised anywhere in the world.

Altering/Modifying Material: Not Permitted
Figures, tables, and illustrations may be altered minimally to serve your work. You may not
alter or modify text in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s).

Reservation of Rights
Springer reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction and (ii) these terms and conditions and (iii) CCC’s Billing and Payment terms and conditions.

License Contingent on Payment
While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Springer or by CCC) as provided in CCC’s Billing and Payment terms and conditions. If full payment is not received by the date due, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC’s Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unretracted license, may constitute copyright infringement and Springer reserves the right to take any and all action to protect its copyright in the materials.

Copyright Notice: Disclaimer
You must include the following copyright and permission notice in connection with any reproduction of the licensed material:
“Springer book/journal title, chapter/article title, volume, year of publication, page, name(s) of author(s), (original copyright notice as given in the publication in which the material was originally published) ”With permission of Springer”
In case of use of a graph or illustration, the caption of the graph or illustration must be included, as it is indicated in the original publication.

Warranties: None
Springer makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you without Springer’s written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer, by CCC on Springer’s behalf).

Objection to Contrary Terms
Springer hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC’s Billing and Payment terms and conditions. These terms and conditions, together with CCC’s Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC’s Billing and Payment terms and conditions, these terms and conditions shall control.

 Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof,
shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other conditions:
V 12AUG2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
VITA

Kazim Sekeroglu was born in Sivas, Turkey. He received B.S. degree in Electronics and Computer Education from Gazi University, Ankara, Turkey, in 2007, and M.S. degree in Electrical Engineering from Southern University and A&M College, Baton Rouge, Louisiana, in 2011. He has been working towards the Ph.D. degree in the department of Electrical Engineering, Louisiana State University, Baton Rouge, Louisiana since 2012. He is now a full-time instructor at the Department of Computer Science and Industrial Technology, Southeastern Louisiana University, Hammond, Louisiana. His research interests are computer vision and image processing.