ISTANBUL TECHNICAL UNIVERSITY ★ INFORMATICS INSTITUTE

REDUCED DIMENSIONAL FEATURES FOR OBJECT RECOGNITION

M.Sc. THESIS

Reyhan Kevser KESER

Department of Applied Informatics

Applied Informatics Programme

Thesis Advisor: Assoc. Prof. Dr. Behçet Uğur TÖREYİN

JULY 2018

ISTANBUL TECHNICAL UNIVERSITY ★ INFORMATICS INSTITUTE

REDUCED DIMENSIONAL FEATURES FOR OBJECT RECOGNITION

M.Sc. THESIS

Reyhan Kevser KESER
(708161014)

Department of Applied Informatics

Applied Informatics Programme

Thesis Advisor: Assoc. Prof. Dr. Behçet Uğur TÖREYİN

JULY 2018

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ BİLİŞİM ENSTİTÜSÜ

NESNE TANIMA İÇİN BOYUTU İNDİRGENMİŞ ÖZNİTELİK VEKTÖRLERİ

YÜKSEK LİSANS TEZİ

Reyhan Kevser KESER
(708161014)

Bilişim Uygulamaları Anabilim Dalı

Bilişim Uygulamaları Programı

Tez Danışmanı: Doç. Dr. Behçet Uğur TÖREYİN

TEMMUZ 2018

“Reyhan Kevser KESER”, a M.Sc. student of ITU Informatics Institute student ID 708161014, successfully defended the thesis entitled “REDUCED DIMENSIONAL FEATURES FOR OBJECT RECOGNITION”, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Assoc. Prof. Dr. Behçet Uğur TÖREYİN** ....
İstanbul Technical University

Jury Members : **Prof. Dr. Lütfiye DURAK ATA** ....
İstanbul Technical University

Assoc. Prof. Dr. Gökhan Bora ESMER ....
Marmara University

Date of Submission : 2 July 2018
Date of Defense : 27 July 2018

To my family,

FOREWORD

I want to thank my supervisor Assoc. Prof. Dr. Behçet Uğur TÖREYİN for his support on my academical journey. And, I want to thank my family for supporting me throughout my life. This thesis work is supported in part by the ITU BAP project, titled "Yeni Nesil Kent Güvenlik Sistemleri için Özgün Veri Edinimi, Veri Sıkıştırılması ve Anlamlandırılması ile Veri/Karar Birleştirme Yöntemlerinin Geliştirilmesi (Data Acquisition, Compression, Analysis and Fusion Methods for new generation surveillance systems)" and number MGA-2017-40964.

July 2018 Reyhan Kevser KESER

TABLE OF CONTENTS

Page

ABBREVIATIONS

SIFT	**:** Scale Invariant Feature Transform
SURF	**:** Speeded-Up Robust Features
MSE	**:** Mean Squared Error
HOG	**:** Histogram of Oriented Gradients
PLS	**:** Partial Least Squares
PCA	**:** Principal Component Analysis
LDA	**:** Linear Discriminant Analysis
RP	**:** Random Projection
SVD	**:** Singular Value Decomposition
LPP	**:** Locality Preserving Projection
KLD	**:** Kullback-Leibler Distance

LIST OF TABLES

LIST OF FIGURES

Page

REDUCED DIMENSIONAL FEATURES FOR OBJECT RECOGNITION

SUMMARY

Object recognition is one of the substantial problems of computer vision area. Traditional solutions consist of feature based object recognition techniques. Hence, there are many studies which are proposed feature detection and description methods. Object recognition can be performed with high accuracy thanks to these robust features. However, these features suffer from their high dimensional structure, in other words “curse of dimensionality”. Hence, dimensionality reduction of the feature vectors is quite studied and methods that reduce computational load are proposed, in the literature.

In this thesis, dimensionality reduction of visual features using autoencoders is proposed. And, the effect of dimensionality reduction of visual features are investigated on object recognition task. For this purpose, three well-known feature vectors are selected which are Histogram of Oriented Gradients (HOG), Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF).

To conduct experiments, three subsets of Caltech-256 dataset images are designed and HOG, SIFT and SURF feature vectors are obtained from these subsets. Dimensionality of these feature vectors are reduced to half using autoencoders. Then, object recognition is tested with original and reduced dimensional vectors with three different distance measures.

Autoencoders which are unsupervised neural network algorithms, are selected for dimensionality reduction of feature vectors since autoencoders can capture nonlinear relationship in data, provide trained model for new inputs and do not need labels. Also, Principal Component Analysis (PCA) is used for dimensionality reduction of these feature vectors for comparison, since PCA is commonly used for dimensionality reduction of these vectors in the literature. Moreover, experiments using the proposed method and PCA, are repeated on images with noise and results are reported.

The results show that object recognition accuracies are improved owing to dimensionality reduction. This shows that unnecessary features and noise are eliminated by dimensionality reduction. In addition to this, dimensionality reduction provides memory and time efficiency.

NESNE TANIMA İÇİN BOYUTU İNDİRGENMİŞ ÖZNİTELİK VEKTÖRLERİ

ÖZET

Gerçek verilerin çoğunluğu büyük boyutlu verilerdir. Ancak büyük boyutlu verilerin işlenmesi bazı nedenlerden ötürü zordur. Büyük boyutlu veriler büyük hesapsal yük oluşturur, daha çok bellek ve zamana ihtiyaç duyar ve görselleştirme açısından zorluklar barındırır. Büyük boyutlu verilerin işlenmesindeki bu zorluk literatürde "Boyutsallık laneti" olarak isimlendirilmiştir. Bu nedenle boyut indirgeme, verilerin işlenmesinde önemli bir basamak olarak karşımıza çıkmaktadır. Literatürde, boyut indirgeme üzerine birçok teknik önerilmiştir. Bu teknikler kullandıkları yöntemler açısından öznitelik seçme ve öznitelik çıkarma metodları olarak iki başlıkta toplanabilir. Veriyi dönüştürme biçimleri açısından ise doğrusal ve doğrusal olmayanlar olarak sınıflandırılabilir.

Nesne tanıma ise bilgisayarla görü alanının, üzerinde çokça çalışılmış ve çalışılmaya devam edilen problemlerinden bir tanesidir. Nesne tanıma, görüntüdeki cisim veya cisimlerin anlamlandırılması anlamına gelmektedir. Nesne tanıma için kullanılan yöntemlerden bir kısmı öznitelik vektörlerinden faydalanmaktadır. Literatürde, bu amaçla tanımlanmış birçok öznitelik vektörü elde etme metodları önerilmiştir. Bu vektörlerden popüler üç tanesi, bu tez çalışması için seçilmiştir. Bunlar HOG (Histogram of Oriented Gradients - Yönlü Gradyanların Histogramı), SIFT (Scale - Invariant Feature Transform - Ölçekten Bağımsız Öznitelik Dönüşümü) ve SURF (Speeded - Up Robust Features - Hızlandırılmış Gürbüz Öznitelikler) vektörleridir.

Bu tezde boyut indirgeme etkisi nesne tanıma problemi üzerinde incelenecektir ve öznitelik vektörlerinin nesne sınıflandırma başarımının artması hedeflenmektedir. Bu durumda farklı sınıflara ait veri noktaları arasındaki mesafeyi arttıracak yöntemler düşünülebilir. Ancak SIFT ve SURF vektörleri için böyle bir yöntem kullanılamaz. Çünkü bu vektörler tüm görüntüyü değil, görüntüdeki önemli noktaları betimleyen vektörlerdir. Görüntülerdeki önemli noktalar her zaman resme özgü olmak zorunda değildir, aynı önemli nokta birçok görüntüde birden bulunabilir. Bu nedenle, görüntülerin sınıflandırılmasında ve aynı zamanda nesne tanımada kullanılan yöntemler, sorgulanan görüntü ile en çok ortak veya benzer önemli nokta içeren resmin seçilmesiyle gerçekleşmektedir.

Literatürde denetimli ve denetimsiz algoritmalar kullanılarak bu konuya ilişkin çalışmalar yapılmıştır. Denetimli algoritmalar SIFT ve SURF gibi yerel görüntü tanımlayıcıları için iki şekilde etiket bilgisi kullanmışlardır. Bunlar öznitelik vektörünün içinde bulunduğu görüntünün sınıf etiketini kullanmak veya öznitelik vektörlerini gruplandırarak etiket bilgisi elde etmek şeklindedir. Ancak iki yöntem de sakıncalıdır. Bir öznitelik vektörü yerel bilgiye dayalı olduğundan sadece bir nesne sınıfıyla bağdaştırılamaz ve yukarıda da belirtildiği gibi bu durum sınıf etiketi kullanmayı verimsiz kılmaktadır. Vektörleri gruplandırma stratejisinde ise kullanıcının belirlediği sayıda grup oluşturulmakta ve grup etiketleri vektör etiketi

olarak kullanılmaktadır. Benzer şekilde kullanıcıya bağlı olan bu etiketleme aşaması da verimsizdir. Bu nedenle denetimsiz öğrenen algoritmalardan faydalanılmalıdır. Bu amaçla, bu çalışmada denetimsiz öğrenen bir yapay sinir ağı modeli olan otokodlayıcı kullanılmıştır. Otokodlayıcı temelde, özdeşlik fonksiyonunu öğrenmeye çalışmaktadır. Çünkü sistem çıktısı olarak girdiyi mümkün olduğu kadar tekrar çatması beklenmektedir.

Otokodlayıcılar kodlayıcı ve kod çözücü iki bölümden oluşmaktadır. Kodlayıcı verilen girdiyi "kod"a dönüştürür, kod çözücü ise "kod"u çıktıya dönüştürmektedir. Kod bölümünün boyutu girdi boyutundan küçük seçilerek boyut indirgeme işlemi sağlanmaktadır. Çünkü bu şekilde tasarlanan sistemler çıktıyı daha küçük boyutlu olan koddan elde etmeye çalışmaktadır. Çıktıyı en iyi şekilde elde etmek için kodun girdiyi en iyi şekilde temsil ediyor olması gerekmektedir. Girdinin istenen boyutta en iyi şekildeki temsili olan kod, girdinin boyut indirgenmiş sonucu olarak karşımıza çıkmaktadır.

Boyut indirgeme işlemi için otokodlayıcı kullanılmasının faydalarından biri otokodlayıcının yeni gelen veriye hazır bir model sunmasıdır. Buna ek olarak otokodlayıcı katmanlarında doğrusal olmayan fonksiyon kullanılmasıyla, verideki doğrusal olmayan ilişki yakalanabilir. Böylece doğrusal yöntemlere nazaran daha karmaşık verilerle baş edebilen bir çözüm sağlanmış olur.

Otokodlayıcılar, farklı kısıtlamalar getirerek farklı amaçlar için de kullanılabilirler. Gürültü giderme ve seyrek betimleme amaçları bunlardan ikisidir. Bu çalışmada kullanılan otokodlayıcılar ise 3 adet gizli katman içeren, "vanilya otokodlayıcı" yapılarıdır. Kullanılan otokodlayıcılarda kod bölümü haricinde tüm katmanların boyutu girdiyle aynı seçilmiştir. Kod bölümü için ise girdinin ½ katı boyut seçilmiştir. Boyutu indirgenmiş vektörleri elde etmek için otokodlayıcı eğitildikten sonra verilen girdiye ilişkin kod bölümü alınmaktadır.

Bu çalışmada obje tanıma problemi için Caltech-256 veri kümesinden 3 adet alt küme elde edilmiştir. Her bir küme 10 nesne sınıfına ait onbirer görüntüden meydana gelmektedir. Bu 11 görüntü ise kendi içinde 1+10 şeklinde ikiye ayrılmaktadır. Her nesne sınıfı için 1 görüntü, sınıf şablonu olarak kullanılmaktadır. Nesne sınıflarındaki kalan 10'ar görüntü ise obje tanıma işlemine sokulup, içindeki objenin belirlenmesi istenmektedir.

Her bir alt kümede farklı öznitelik vektörleri üzerinde çalışılmıştır. Her bir küme için öncelikle, kümedeki görüntülerden ilgili öznitelik vektörleri elde edilmiştir. Bu vektörler henüz orjinal boyutlarında iken obje tanıma testi yapılmıştır. Ardından otokodlayıcı kullanarak boyut indirgeme işlemi gerçekleştirilip, obje tanıma testi tekrarlanmıştır.

Her bir vektör grubu için 110 görüntü ile oluşturulan alt kümeler, otokodlayıcının test kümesi olarak kullanılmaktadır. Otokodlayıcının eğitimi için test kümesinin iki katı vektör içeren ve rastgele seçilmiş vektörlerden oluşan eğitim kümeleri kullanılmıştır. Sistem için oluşturulan doğrulama kümeleri ise test kümesindeki gibi 10 sınıfa ait 11'er görüntüden oluşmaktadır. Eğitim, test ve doğrulama kümelerinin farklı görüntüler kullanılarak oluşturulduğu not edilmelidir.

Her kümede, görüntülerden ilişkili öznitelik vektörleri çıkarılmıştır. Yapılan nesne tanıma ve boyut indirgeme işlemleri bu vektörler üzerinde gerçekleşmektedir. Otokodlayıcılar sistem yakınsayana kadar bu eğitim kümeleriyle eğitilmiş, ardından test kümeleri sisteme sokularak boyutu indirgenmiş vektörler elde edilmiştir.

Otokodlayıcı ile 1764 boyutlu HOG vektörleri 882, 128 boyutlu SIFT vektörleri 64 ve 64 boyutlu SURF vektörleri 32 boyuta indirgenmiştir. Kıyaslama amacıyla aynı boyut indirgeme işlemi, literatürde bu konu üzerine oldukça çalışılmış, Temel Bileşenler Analizi (PCA) ile de gerçekleştirilmiş ve sonuçlar sunulmuştur. Buna ilaveten oluşturulan kümelerdeki görüntüler gürültü ile bozularak gürültülü kümeler oluşturulmuş ve nesne tanıma işlemi gürültülü görüntülerden çıkarılan öznitelik vektörleri ile test edilmiştir. Ardından bu vektörler de otokodlayıcı ve Temel Bileşenler Analizi (PCA) kullanılarak boyut indirgeme işlemine tabi tutulmuştur. Boyutu indirgenmiş bu vektörlerin nesne tanıma başarımları ölçülmüştür

Elde edilen sonuçlar, otokodlayıcıya dayalı boyut indirgemenin, öznitelik vektörleri kullanılarak nesne tanıma işleminin hem orjinal hem gürültülü görüntülerde başarımını arttırdığını göstermektedir. Bunun sebebinin, boyut indirgemenin verideki fazlalık bilgiyi ve gürültüyü gidermesi olduğu düşünülmektedir. Boyut indirgeme, sadece başarımı arttırmamıştır, aynı zamanda vektörlerin saklanması için gereken bellek miktarını azaltmaktadır.

1. INTRODUCTION

1.1 Purpose of Thesis

Most of the real-world data are high dimensional. These data are hard to process in many aspects such as great computational load, big memory usage, difficulty of visualization and time inefficiency. To alleviate curse of dimensionality, which is a term describing the phenomena that processing data gets harder as the data has more dimensions, it is necessary to find ways to reduce dimensionality (Bishop, 2006).

In the literature, many methods are proposed such as PCA (Pearson, 1901), Factor Analysis (Spearmen, 1904), Classical Scaling (Torgerson, 1952), LLE (Roweis and Saul, 2000), LDA (Fisher, 1936), Kernel PCA (Schölkopf et al., 1998), mRMR (Ding and Peng, 2005), and ISOMAP (Tenenbaum et al., 2000) to reduce dimensionality. These methods are based on either one of the two principles, which are feature selection or future extraction and they use either of the two ways to transform the data, which are linear or non-linear transformation (Alpaydin, 2009).

In this thesis, autoencoders are used for dimensionality reduction. Autoencoders are neural network algorithms which are first presented by Rumelhart et al. (1986). Autoencoders learn in unsupervised manner, hence they do not need labeled data. In addition to this, an autoencoder can perform non-linear transformation. Thus, they can capture more complex relationship in data than linear methods.

In this thesis, effects of feature reduction on object recognition performance is investigated. For this purpose dimensionality of HOG, SIFT and SURF features are reduced to eliminate noise and unnecessary data using autoencoders, since autoencoders are suitable for dimensionality reduction of keypoint and image descriptors.

1.2 Literature Review

Many studies have been working on dimensionality reduction of SIFT features since these features are high dimensional vectors. Some of the studies proposed to binarize and quantize SIFT vectors to reduce computational load (Strecha et al., 2012; Yeo et al., 2008; Stommel and Herzog, 2009; Yeo et al., 2008; Tuytelaars and Schmid, 2007; Philbin et al., 2007), whereas some of the studies proposed linear projection methods for dimensionality reduction of SIFT vectors (Mikolajczyk and Matas, 2007; Valenzuela et al., 2014; Cai et al., 2011).

The most popular linear method used for dimensionality reduction of SIFT vectors is PCA (Watcharapinchai et al., 2009; Chandrasekhar et al., 2009; Asbach et al., 2008; Brown and Süsstrunk, 2011). A well-known method that benefits from PCA is PCA-SIFT (Ke and Sukthankar,2004) which is obtained by applying PCA to the normalized gradient patches instead of using smoothed weighted histograms in computation of SIFT descriptors. It was shown that PCA-SIFT gives better results than SIFT in image retrieval in the aspects of accuracy and time. Valenzuela et al. (2012) proposed to use PCA to reduce the dimensionality of SIFT and SURF features. The reduced vectors are named as Reduced-SIFT and Reduced-SURF. The authors stated that Reduced-SIFT is better than SIFT in image retrieval. However, Reduced-SURF could not achieve the success of SURF in image retrieval.

Ledwich and Williams (2004) changed some steps in SIFT computation in order to reduce dimensionality. The dimensionality is reduced well but there was a little loss in accuracy.

Some studies proposed supervised and semi-supervised methods to reduce dimensionality of SIFT vectors using such as PLS (Valenzuela et al. ,2013; Farquhar et al., 2005) and a derivation of LPP (Cevikalp et al., 2008).

Valenzuela et al. (2013) reported that supervised methods give better results than unsupervised methods. Image class labels are used as feature labels for implementation of supervised methods, in this study. However, SIFT features can be found in images which belong to different classes since these features are local descriptors. In other words, a feature cannot be related with one object class.

PCA is also used for dimensionality reduction of SURF and SURF based feature vectors (Asbach et al., 2008; Valenzuela et al., 2012; Boulkenafet et al., 2017).

Valenzuela et al. (2014) used four linear dimensionality reduction methods for SIFT and SURF vectors. The results show that RP (Random Projection), LDA (Linear Discriminant Analysis) and PLS (Partial Least Squares) methods could not achieve original SURF vectors' success while PCA outperforms original SURF vectors.

However, some studies showed that PCA is not the best method for dimensionality reduction of SURF vectors. There are studies demonstrating that SVD (Singular Value Decomposition) gives better results than PCA (Vinay et al., 2016) and that Reduced-SURF could not achieve the success of original SURF vectors (Valenzuela et al., 2012).

Valenzuela et al. (2013) used both of the supervised and unsupervised methods for dimensionality reduction of SIFT and SURF vectors. It is reported that unsupervised methods give acceptable results. However, supervised methods gave better results than unsupervised methods.

One way of using supervised methods can be constructing codebook vectors from feature vectors by clustering features and using cluster labels as feature labels. However, number of clusters must be defined by user and it can be different from intrinsic classes. Hence unsupervised methods are convenient for local image descriptors.

PCA is a popular dimensionality reduction technique for HOG features, too (Savakis et al., 2014; Felzenszwalb et al., 2010; Lu and Little, 2006; Kobayashi et al., 2007). However, better results than PCA are reported using different dimensionality techniques (Déniz et al., 2011; Misra et al., 2011; Monzo et al., 2011; Xiao et al., 2010). Another popular technique to dimensionally reduce HOG features is PLS (Partial Least Squares) which is a supervised method (Misra et al., 2011; Hussain and Triggs, 2010; Schwartz et al., 2009). Moreover, some studies reported the results of LPP and its derived versions (Wang and Zhang, 2008; Mathias et al., 2013), LDA and its derived versions (Déniz et al., 2011; Monzoet al., 2011; Mathias et al., 2013) and RP (Savakis et al., 2014) for dimensionality reduction of HOG features.

1.3 Previous Work

In our previous work (Keser et al., 2018b), dimensionality of SIFT features were reduced to half and quarter of the original size. For this purpose, a dataset (Keser et al., 2018a) is constructed which consist of cropped car logo images obtained from Medialab LPR dataset ("Medialab LPR database, " n.d.) and the Internet. This dataset consists of 90 images which is 10 images for each car brand (Figure 1.1). For further computations, SIFT vectors of the images in the dataset were extracted. Then dimensionality of the features was reduced using a 5-layered vanilla autoencoder. Mean squared error was used as loss function and Adam optimizer was used to minimize the loss function (Kingma and Ba, 2014).

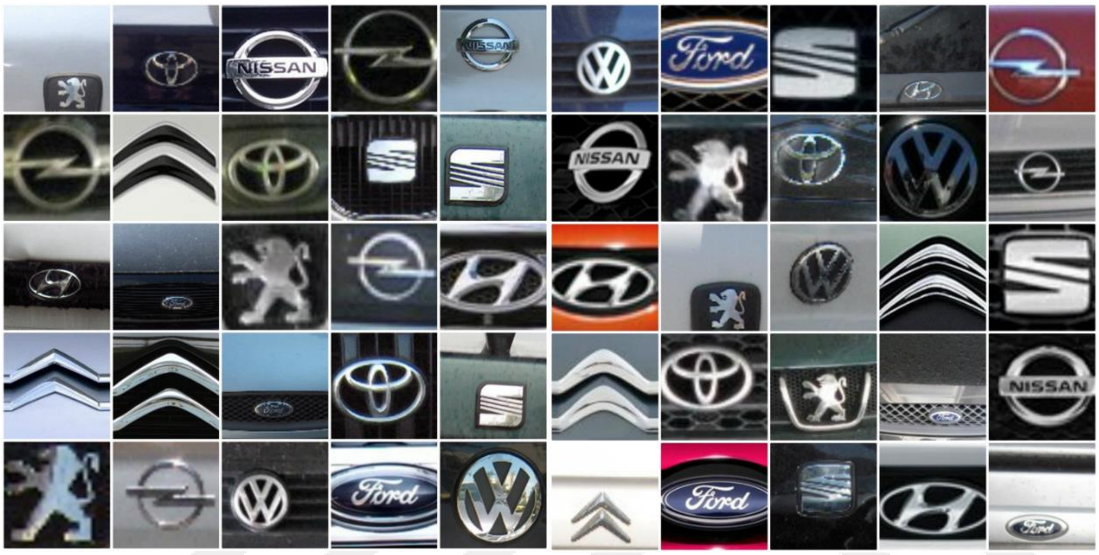

Figure 1.1 : Sample images of the vehicle logo dataset used in the previous work(Keser et al., 2018a).

In order to evaluate success of the autoencoder, original and reduced features were inputted into logo recognition process. When the half sized features were used in vehicle logo recognition, the accuracy of the recognition was decreased by 19% while memory saving of 50% and time saving of 8% was achieved. If the quarter sized features were used in vehicle logo recognition, the accuracy of vehicle logo recognition was decreased by 22% while memory saving of 75% and time saving of 21% was achieved (Keser et al., 2018b).

1.4 Hypothesis

In our previous work, autoencoders were used for dimensionality reduction of SIFT vectors (Keser et al., 2018b). The method provides memory saving in exchange for some loss in accuracy. In this thesis, the study is improved and expanded to SURF and HOG features, too.

In this thesis, it is aimed to improve object recognition accuracies obtained by HOG, SIFT and SURF feature vectors by reducing dimension. In other words, the goal of this thesis is removing unnecessary features and noise from feature vectors by dimensionality reduction.

In order to improve classification results, one can use methods which are aimed to maximize class separation. These methods require class labels, so they are supervised algorithms. However, class labels cannot be obtained for SIFT and SURF features, since these vectors have information about keypoints in images instead of the whole image. It is clear that, some keypoints can be found on the images too other than the query image and some of them can be found only on that particular query image. As in this thesis, object recognition can be accomplished by classifying the images which consist of one object and image classification is achieved by matching the query image with the image, which has the maximum number of similar keypoints. This scheme demonstrates too that a keypoint is not related with only one class.

To sum up, class labels cannot be obtained for keypoints, and, hence not for SIFT and SURF vectors, either. Thus, unsupervised methods should be used for SIFT and SURF vectors.

In addition to this, HOG vectors are inputted into dimensionality reduction task using the same method to evaluate the method's performance on both the keypoint descriptors and the image descriptors.

To achieve these goals, autoencoders, which are unsupervised neural network algorithms, are used in this thesis for dimensionality reduction of HOG, SIFT and SURF. Autoencoders are chosen to be used, because of their ability of capturing nonlinear relationship in the data, and providing a ready model for new input data.

2. FEATURE VECTORS FOR OBJECT RECOGNITION

In this thesis, effects of dimensionality reduction on feature based object recognition methods are investigated. To achieve this, three popular feature vectors for object recognition are selected. One of the selected vectors is HOG feature, which is an image descriptor and commonly used for human detection. Other selected vectors are two popular local image descriptors which are SIFT and SURF features. In this section, these three selected feature vectors to conduct experiments, are described.

2.1 Histogram of Oriented Gradients (HOG)

One of the commonly used feature vector for object recognition in the literature is Histogram of Oriented Gradients (HOG). It is proposed in 1986 without the term HOG. However, it was not popular until that it is shown that HOG features can be used for human detection (Dalal and Triggs, 2005).

In order to compute HOG features, firstly vertical and horizontal gradients of the image are obtained. These gradient values are converted into polar coordinate values which are the magnitude and direction values. If the image has three color channels, the maximum magnitude and direction values are selected among the magnitude and direction values related color channels for each pixel. According to these gradient magnitude and direction values, histogram of gradients vectors are obtained for each (typically 8x8) image blocks. These histogram values are computed according to 9 bins. Then these histogram vectors are normalized with a technique called 16x16 block normalization. It should be highlighted that the parameters can be changed according to the application. Finally, these vectors are concatenated to form a single HOG vector.

HOG features are robust to illumination changes thanks to the normalization step. However, they are not rotation and scale invariant. Also, it should be noted that the size of HOG feature changes according to the image size, due to the amount of the histogram vectors obtained from image blocks.

MATLAB's built-in function is used to extract HOG features from images, in this thesis. A sample image is given in Figure 2.1 to visualize HOG features.

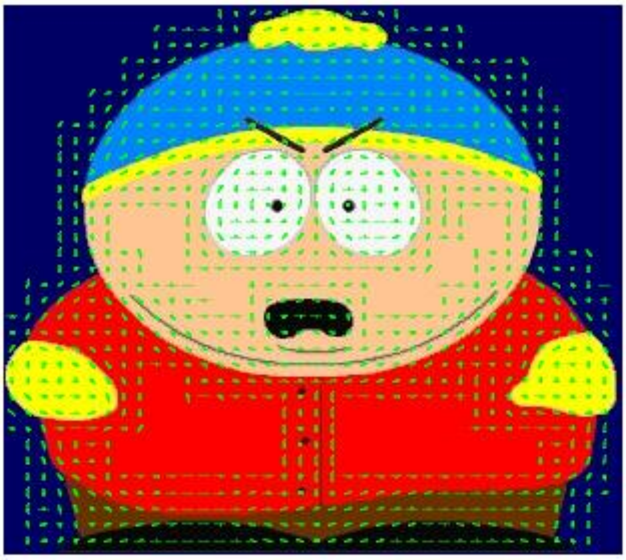

Figure 2.1 : A sample image with HOG features.

2.2 Scale Invariant Feature Transform (SIFT)

The Scale Invariant Feature Transform (SIFT) algorithm is a well-known method to detect interest points and form descriptors from them. This method is proposed by Lowe (2004a). It provides scale, rotation and translation invariant feature vectors for object detection and recognition.

In order to obtain the SIFT descriptors, firstly potential keypoints are detected by searching local extrema using different scales of the image. Difference of Gaussians (DoG) filters are used in this step. Then exact keypoints are determined by a thresholding system by eliminating low-contrast keypoints and poorly localized edge keypoints. After that, orientation(s) of the keypoints are determined. Finally, descriptors are computed for each keypoint using gradient orientations and magnitudes in the neighborhood of size 16x16. Thus, 128 dimensional feature vectors are obtained by concatenating 8 bins histograms, which are generated using gradient orientations and magnitudes, for each non-overlapping 4x4 windows on this neighborhood of the keypoint.

A sample image with SIFT vectors is given in Figure 2.2. The red vectors on the image, are SIFT vectors symbolizing keypoints location, scale and orientation. In this thesis, the demo software provided by Lowe (2004b) is used in order to extract SIFT vectors.

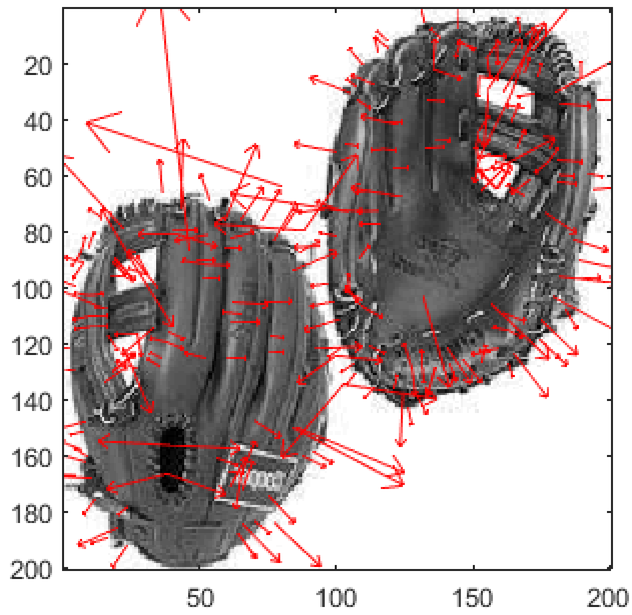

Figure 2.2 : SIFT vectors are shown on a sample image of Set 2.

2.3 Speeded-Up Robust Features (SURF)

Speeded-Up Robust Features (SURF) method is one of the commonly used feature detection and description method for object recognition in the literature. It is proposed by Bay et al. (2006).

Obtaining SURF feature vectors, consist of steps like in SIFT computation. However, these steps are improved in order to obtain features faster. In order to compute SURF features, box filters are used instead of Difference of Gaussians filters in SIFT process. The scale space is obtained using the box filters with different sizes. Blob detectors based on Hessian matrix are utilized to find the points of interest. The points where the determinant of the Hessian matrix is maximum, are selected. After locating the keypoints, keypoint orientations are computed. To determine the orientation of the keypoints, wavelet responses in horizontal and vertical axis are used. Then descriptors are computed.

SURF vectors are scale and rotation invariant features. However, to speed up the process, rotation variant SURF vectors which are called U-SURF, can be used.

An example image from Set 3 which is a subset generated from Caltech-256 dataset (Griffin et al., 2007) to demonstrate SURF vectors, is given in Figure 2.3 with 20 strongest SURF vectors on it.

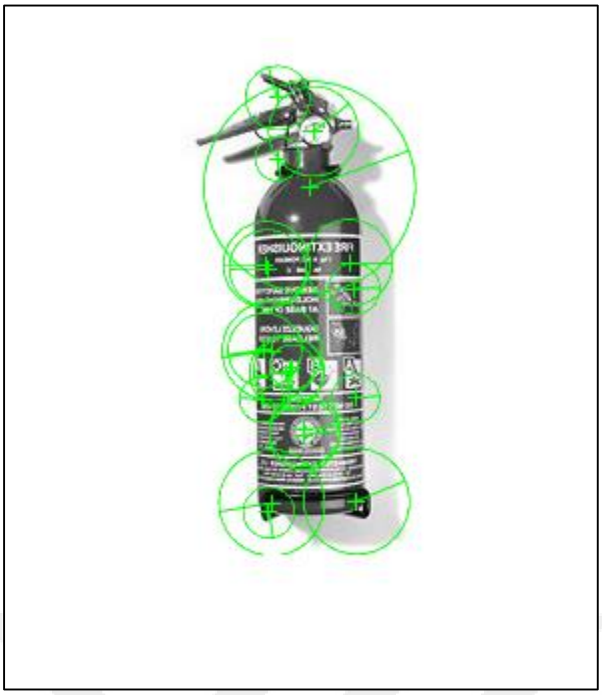

Figure 2.3 : 20 strongest SURF vectors are shown on an example image of Set 3.

3. OBJECT RECOGNITION

Object recognition is an easy task for human-beings, which consists of identifying objects in images or image sequences, even when objects are occluded, rotated, translated, scaled or illuminated. However, this task is troublesome for computers. Thus, object recognition is a substantial problem of computer vision field and many studies are carried out on this topic. Some of the studies proposed feature based techniques which rely on global or local feature extraction from the image and matching these features. In this thesis, dimensionality reduction effects on three feature based object recognition methods are investigated. To achieve this, three popular feature vectors for object recognition are selected and object recognition is tested with these vectors before and after the dimensionality reduction (Grauman and Leibe, 2011, p.65).

One of the chosen feature vectors is Histogram of Oriented Gradients (HOG). The details of HOG features are mentioned in Section 2. Figure 3.1 demonstrates the scheme of object recognition with HOG feature vectors.

Other chosen feature vectors are Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) vectors. The details about these features is mentioned in Section 2. Image matching process is similar for SIFT and SURF vectors and Figure 3.2 shows the scheme of object recognition with these vectors.

Image matching scheme is different for HOG features because one HOG feature is extracted per image. However, hundreds of SIFT and SURF features are extracted per image.

Related papers of SIFT and SURF, are proposed to use threshold values for vector matching (Lowe, 2004a; Bay, 2006). This threshold is used for the ratio of distance of the closest vector to distance of the second closest vector. If this ratio is higher than the threshold, the match is accepted, if else, the match is unaccepted. However, in this thesis any threshold is used for vector matching for both of the original and reduced dimensional vectors. Since the effect of the dimensionality reduction is examined, any scheme can be used for parameters if the conditions are the same.

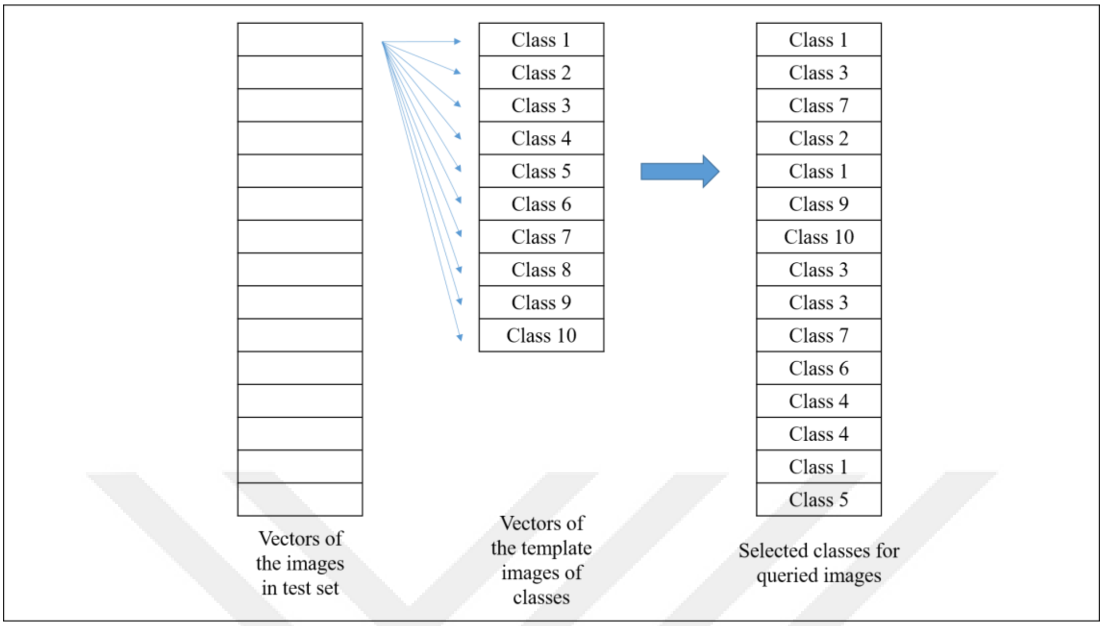

Figure 3.1 : The test set of object recognition method consists of 100 images, and so 100 HOG vectors.To match an image, related HOG vector is tried to match with the vectors of template images using a distance measure such as Cosine, Euclidean and Kullback-Leibler. The image is determined in the class whose vector has the minimum distance to the query image vector.

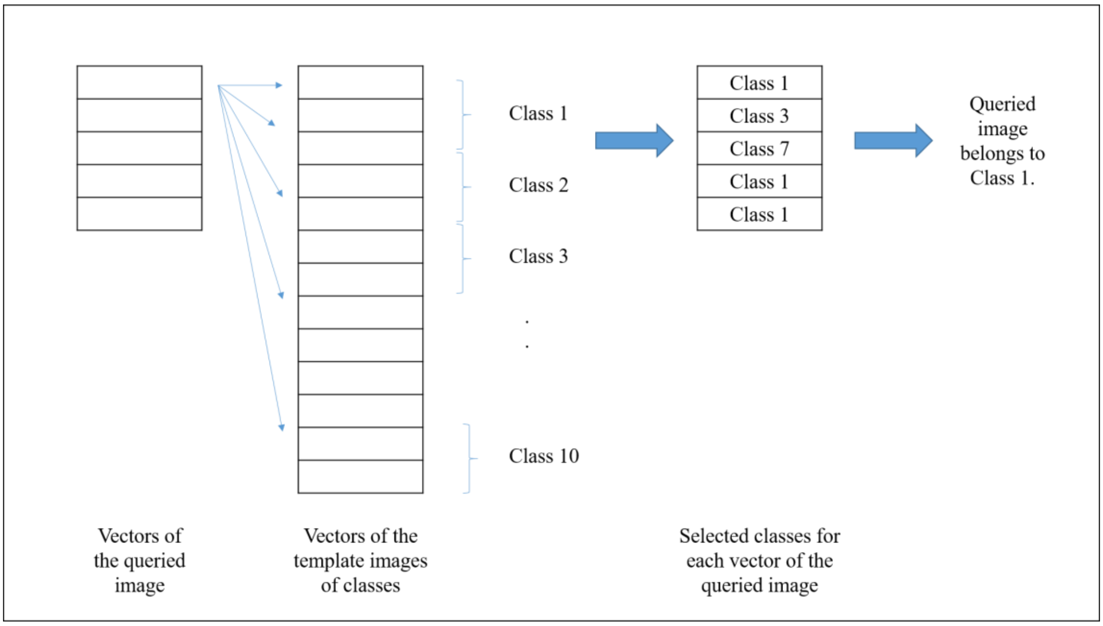

Figure 3.2 : Each feature vector of the image is compared with all of the feature vectors of class templates. Thus each query vector selects a class whose vector have the minimum distance with query vector. So, the most selected class shows the matching result of the query image.

In this thesis, subsets used for object recognition, consist of images of one object, like majority of images in Caltech-256 dataset. Hence images are tried to match with template class images in order to recognize objects in the images.

In order to match feature vectors, three distance measures which belong to different distance families, are used for each set (Prasath et al., 2017). The distance families used in this thesis are inner product, L_p Minkowski and Shannon entropy distance families.

One of the selected measures is the Cosine distance, which belongs to the inner product distance family. It is obtained from the Cosine similarity. There are versions of this distance. In this thesis, the version in Keras library (Chollet, 2015) is used as presented in Equation 3.1, where x and y represent the vectors and d is the distance.

$$d = -\left(\frac{x \cdot y}{\|x\|\|y\|}\right) \tag{3.1}$$

Another selected measure is the Euclidean distance, which belongs to L_p Minkowski distance family. It is obtained by setting p = 2, hence it is also known as L_2 norm. It is shown in Equation 3.2, where x and y represent the vectors and d is the distance.

$$d = \sqrt{\sum (x-y)^2} \tag{3.2}$$

The other selected measure is the Kullback-Leibler distance, which belongs to the Shannon entropy distance family. It is also called relative entropy or information deviation. It is not symmetric, so it is not a metric measure. Moreover, it is used to quantify the distance between two probability distributions, hence clipped versions of vectors are used for computation of this distance as shown in Equation 3.3, where x_c and y_c are the clipped vectors for limiting the vectors in [0, 1], log is the natural logarithm and d is the distance.

$$d = \sum y_c \log\left(\frac{y_c}{x_c}\right) \tag{3.3}$$

To assess the performance of object recognition, accuracy metric is computed as in Equation 3.4, where A represents the accuracy, T is the number of true matches and N is the number of all images in the set.

$$A = T / N \tag{3.4}$$

4. DIMENSIONALITY REDUCTION

In this thesis, effect of dimensionality reduction on object recognition is studied. Hence object recognition is performed before and after the dimensionality reduction. And scheme of object recognition based feature vectors is presented in the previous section. In this section, firstly autoencoders which are proposed for dimensionality reduction of feature vectors, are mentioned. Then implementation of the suggested method and dataset is presented. Finally results obtained by the proposed method and PCA on original and noisy images are given.

4.1 Autoencoders

Autoencoders are neural networks whose goal is to reconstruct the input. In other words, they try to learn the identity function, in an unsupervised manner.

While autoencoders try to reconstruct the input, they are limited by some restrictions. Hence, they cannot copy only the data, instead of it they must learn useful features in order to reconstruct the data. Hence, they are useful for applications such as dimensionality reduction, data denoising and feature learning (Goodfellow et al., 2016).

They mainly consist of two components which are encoder and decoder. The encoder maps the input to the code and the decoder maps the code to the output as demonstrated in Figure 4.1.

Figure 4.1 : General structure of an autoencoder.

They try to minimize the objective function in Equation 4.1**,** since they try to reconstruct the input.

$$\min L(x, g(f(x))) \tag{4.1}$$

where x represents the input, f and g are the encoder and decoder functions, respectively, L is a distance function, such as mean squared error (MSE).

If the code part of the autoencoder has smaller dimension than input, then the autoencoder is undercomplete. These autoencoders are limited by forcing the code being in smaller dimension and they can be used in order to learn the most useful features of the data. If the code part has greater dimension from the input, then the autoencoder is overcomplete. These autoencoders can be used with some regularization to prevent just copying the data instead of learning useful features. For example, they can be forced to learn sparse representations and to have small values of the derivatives of the representations (Goodfellow et al., 2016).

In this study, vanilla autoencoders are used. In other words, the autoencoders are only forced to have the code in smaller dimension of the input, which means that the autoencoders are undercomplete. However, there are versions of autoencoders in the literature such as denoising, sparse, variational and contractive autoencoders.

4.2 Method

In order to achieve dimensionality reduction, six autoencoders are used for different feature vector sets. Keras with Tensorflow backend is used for all autoencoder implementations (Chollet, 2015; Abadi et al., 2016).

Figure 4.2 shows the autoencoder structure which is used for original and noisy HOG features. This autoencoder has 5 layers which have 1764, 1764, 882, 1764 and 1764 units in order.

To reduce the dimensionality of original HOG features, hyperbolic tangent activation functions (Equation 4.2) are used for all layers except the output layer. Linear activation function is used in the output layer. All of the layers are densely-connected layers. In order to train this autoencoder, stochastic gradient descent optimizer with Nesterov momentum is used with the batch size of 32, learning rate of 0.3 and epoch of 90. Before each epoch, training data is shuffled and biases of the last layer are initialized with zero. Mean squared error is used as loss function in Equation 4.3.

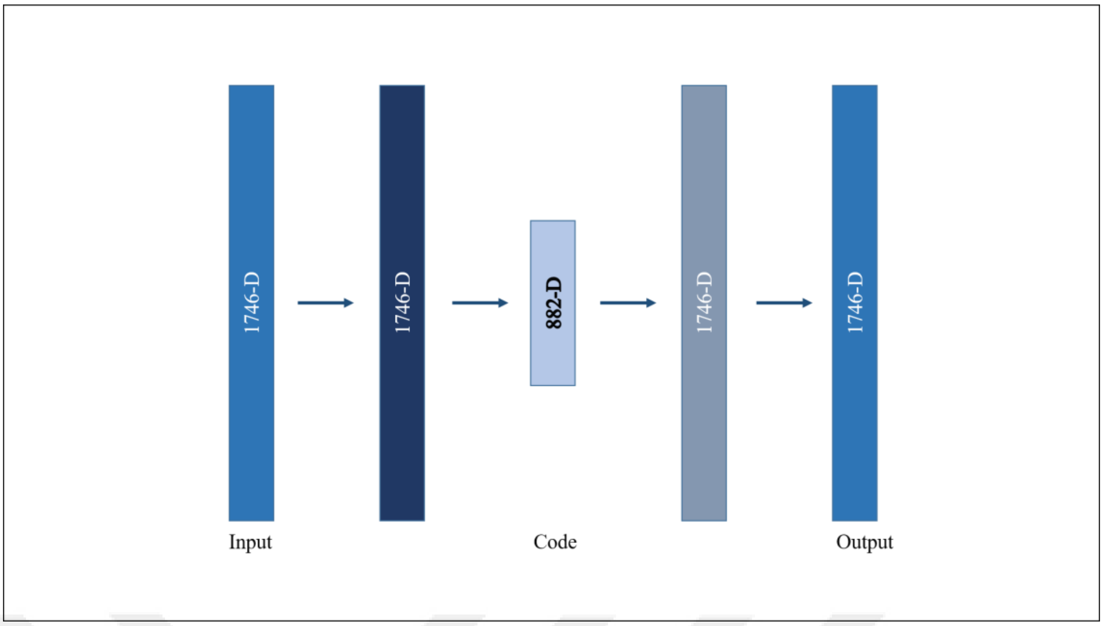

Figure 4.2 : Autoencoder structure used for HOG features.

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{4.2}$$

$$L = \frac{1}{N} \sum (x - y)^2 \tag{4.3}$$

To reduce the dimensionality of noisy HOG features, Leaky ReLU activation functions (alpha = 0.001) are used for all layers except the output layer (Equation 4.4). In the Equation 4.4, α is a small constant and 1(.) is the indicator function which is 1 if the condition inside is true, and 0 otherwise. Linear activation function is used in the output layer. All of the layers are densely-connected layers. In order to train this autoencoder, Adam optimizer (Kingma and Ba, 2014) with recommended hyperparameters is used with the batch size of 32, learning rate of 5×10^{-5} and epoch of 33. Before each epoch, training data is shuffled and all biases are initialized with zero. Mean squared error is used as loss function in Equation 4.3.

$$f(x) = 1(x < 0)(\alpha x) + 1(x \geq 0)(x) \tag{4.4}$$

Figure 4.3 shows the autoencoder structure which is used for original and noisy SIFT features. This autoencoder has 5 layers which have 128, 128, 64, 128 and 128 units in order.

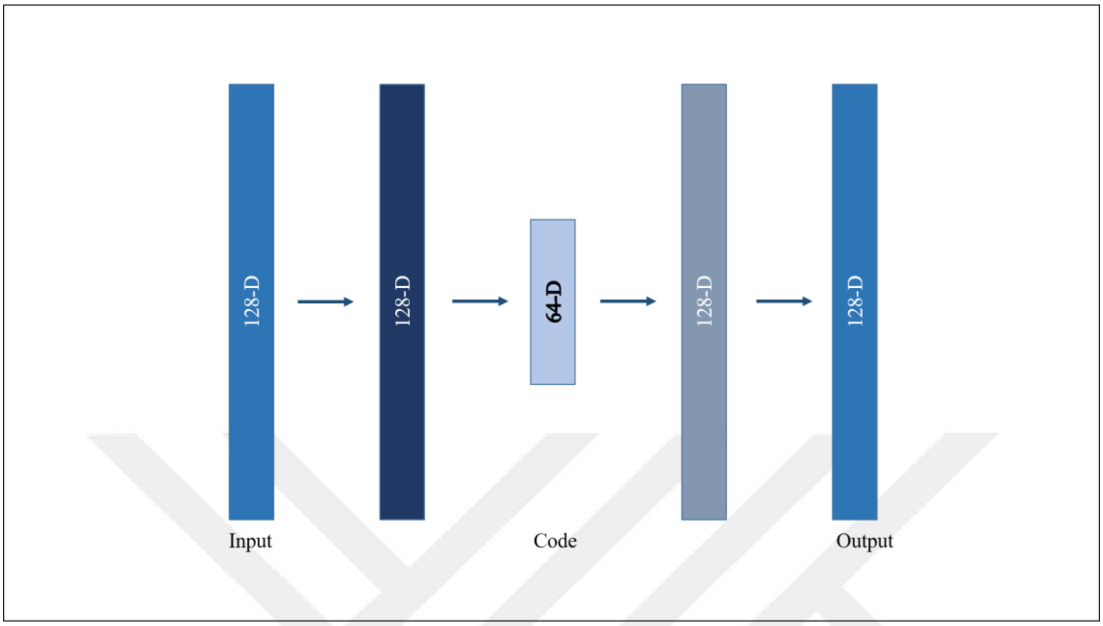

Figure 4.3 : Autoencoder structure used for SIFT features.

To reduce the dimensionality of original SIFT features, hyperbolic tangent activation functions are used for all layers except the output layer. Linear activation function is used in the output layer. All of the layers are densely-connected layers. In order to train this autoencoder, stochastic gradient descent optimizer with Nesterov momentum is used with the batch size of 512, learning rate of 0.4 and epoch of 53. Before each epoch, training data is shuffled. Cosine proximity is used as loss function in Equation 3.1.

To reduce the dimensionality of noisy SIFT features, hyperbolic tangent activation functions are used for all layers except the output layer. Linear activation function is used in the output layer. All of the layers are densely-connected layers. In order to train this autoencoder, Adam optimizer with recommended hyperparameters is used with the batch size of 128, learning rate of 5×10^{-4} and epoch of 90. Before each epoch, training data is shuffled. Cosine proximity is used as loss function in Equation 3.1.

Figure 4.4 shows the autoencoder structure which is used for original and noisy SURF features. This autoencoder has 5 layers which have 64, 64, 32, 64 and 64 units in order.

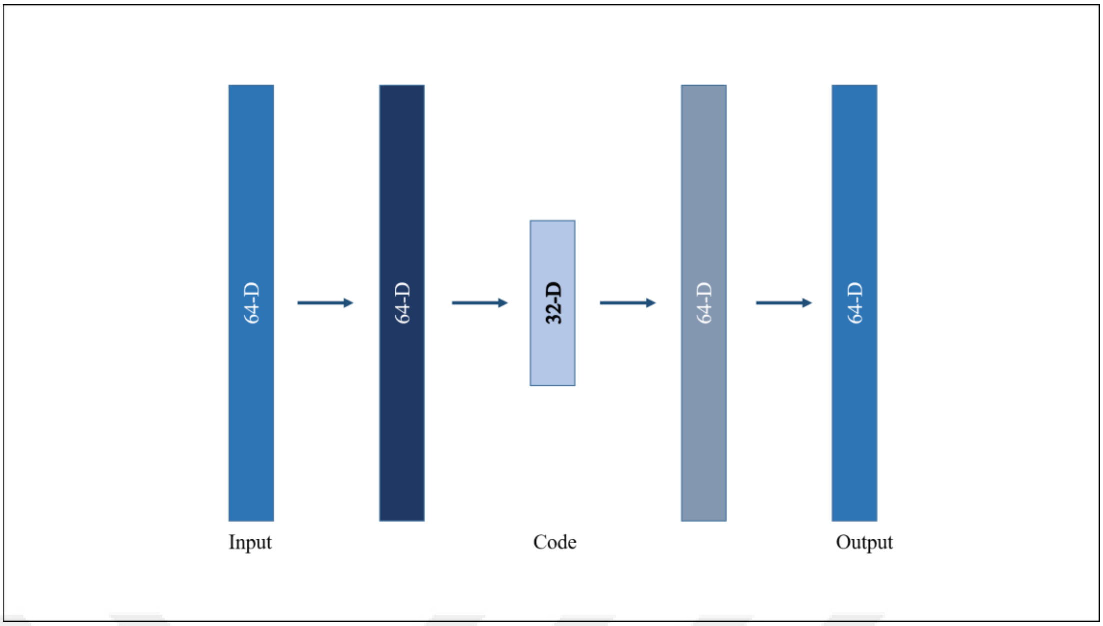

Figure 4.4 : Autoencoder structure used for SURF features.

To reduce the dimensionality of original SURF features, hyperbolic tangent activation functions are used for all layers except the output layer. Linear activation function is used in the output layer. All of the layers are densely-connected layers. In order to train this autoencoder, Adam optimizer with recommended hyperparameters is used with the batch size of 128, learning rate of 5×10^{-4} and epoch of 25. Before each epoch, training data is shuffled. Cosine proximity is used as loss function in Equation 3.1.

To reduce the dimensionality of noisy SURF features, hyperbolic tangent activation functions are used for all layers except the output layer. Linear activation function is used in the output layer. All of the layers are densely-connected layers. In order to train this autoencoder, Adam optimizer with recommended hyperparameters is used with the batch size of 128, learning rate of 1×10^{-5} and epoch of 90. Before each epoch, training data is shuffled. Cosine proximity is used as loss function in Equation 3.1.

4.3 Data Set

The experiments are conducted on three subset of Caltech-256 dataset. HOG, SIFT and SURF features are acquired from Set 1, Set 2 and Set 3, respectively. Object recognition task is tested on each set, as shown in Figure 4.5.

To form the subsets, ten classes are chosen from Caltech-256 dataset as presented in Table 4.1. 11 images are selected from each class to form the subsets. It should be

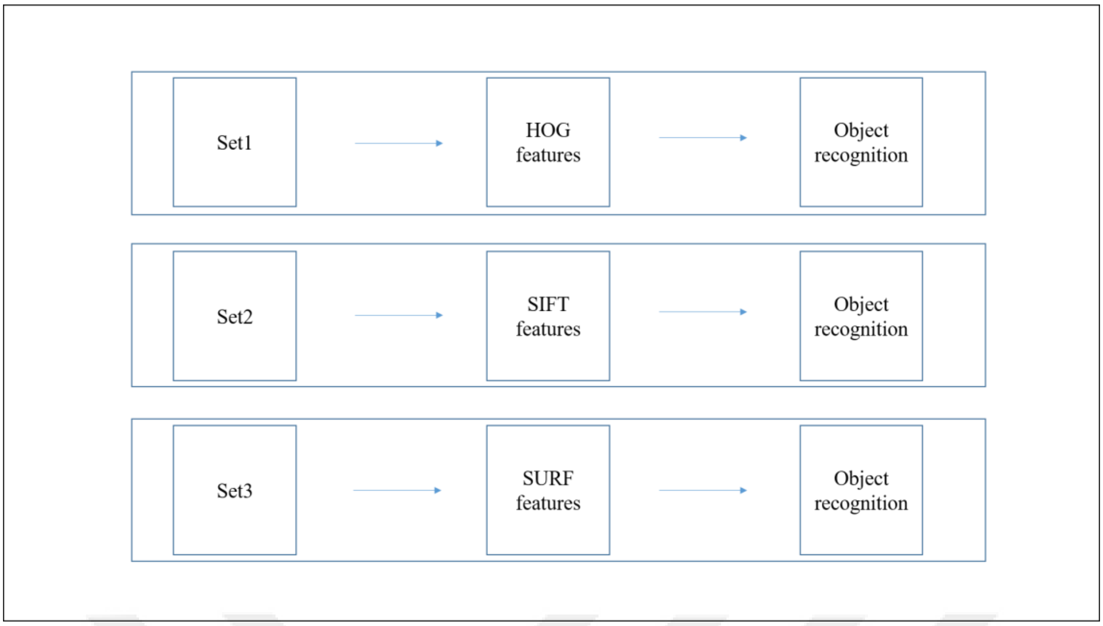

Figure 4.5 : Object recognition is tested for HOG, SIFT and SURF features on Set 1, Set 2 and Set 3, respectively.

noted that the features of the subsets are used as test set of the dimensionality reduction algorithm. Training sets of dimensionality reduction algorithms are twice the size of the test sets and consist of randomly selected different feature vectors for each run. And validation sets which consist of features of different 11 images per each selected class are used for the dimensionality reduction algorithm. Scheme of obtaining training, validation and test sets are shown in Figure 4.6. It should be noted that validation and test sets are fixed and they are used for both of the implementations of dimensionality reduction techniques.

Table 4.1 : For each set 10 object classes are selected from Caltech-256 dataset.

Set 1	Set 2	Set 3
Baseball glove	Baseball glove	Baseball glove
Bonsai-101	Bonsai-101	Brain-101
Bowling-pin	Brain-101	Fire extinguisher
Cartman	Calculator	French horn
Desk globe	Cartman	Frying pan
Electric guitar-101	Desk globe	Grand piano-101
Fire extinguisher	Fire extinguisher	Hamburger
Flashlight	Megaphone	House fly
French horn	Mountain bike	Megaphone
Frying pan	Paperclip	Video projector

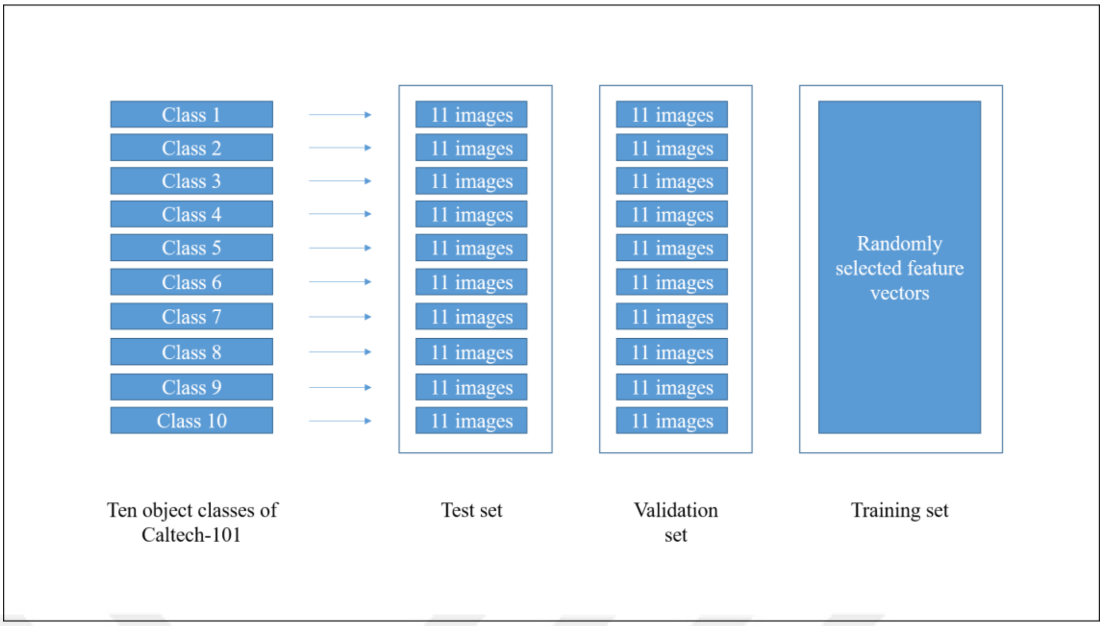

Figure 4.6 : Design of the test, training and validation sets of the dimensionality reduction algorithm. The test set of the system is used for object recognition tests. The validation set has the same amount of images with the test set and is used to validate the system. The training set of the system consists of randomly chosen vectors which are twice the size of vectors in test set.

Before computations, all images are converted to grayscale. It should be highlighted that Set 1 images are also resized to 64x64 pixels, in order to have fixed dimensional HOG features.

For object recognition, each subset is divided to two parts. One part, called as template images, consists of one image per class which are used for matching with query images. This part has 10 images since there are ten classes. The other part consists of the remaining 100 images which are 10 images per each class, as shown in Figure 4.7.

4.4 Results

In this section, details of the implementation of the proposed method are given. And object recognition results before and after the dimensionality reduction by the proposed method and PCA are presented, for comparison. Moreover, all tests are repeated on images with Gaussian noise with zero mean and 0.01 variance, and all results are shown in this section.

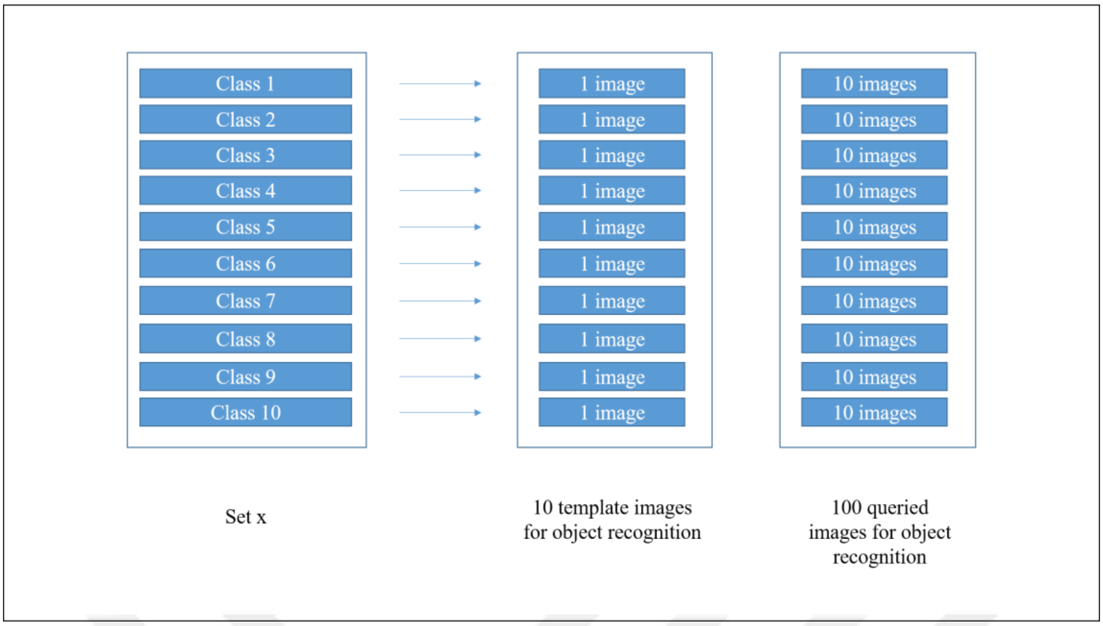

Figure 4.7 : Design of a subset for object recognition tests. 100 query images are compared to 10 template images with a distance measure.

4.4.1 Dimensionality reduction of HOG features

For dimensionality reduction of HOG features, a 5-layered autoencoder is used, which is shown in Figure 4.2. The autoencoder is trained with the training set of 220 vectors for 90 epochs and validated with the validation set. The loss graph of the model is shown in Figure 4.8. After the training step, vectors in the test set is inputted into the autoencoder. The vectors obtained in the code layer are stored, which are the reduced representations of these vectors. Then the low-dimensional vectors are used in object recognition task as mentioned in Section 3. To compare our results, the dimensionality of HOG features is reduced to the same size by PCA as stated by Felzenszwalb et al. (2010). For this purpose random training sets are used for PCA training, and then system is validated using object recognition accuracy of reduced vectors of validation set as shown in Figure 4.6. In this manner, reduced vectors of the Set 1 is obtained by PCA, too.

It should be highlighted that the maximum accuracy of object recognition by original sized HOG features is observed while using the Euclidean distance. Hence, it is aimed to improve object recognition results using Euclidean distance. So, in order to train the autoencoder, MSE is used as loss function and also object recognition accuracies of validation set using Euclidean distance is used to validate the PCA. Accuracy and memory usage results of object recognition using Euclidean distance, before and after the dimensionality reduction by both of the autoencoder and PCA is given in Table 4.2.

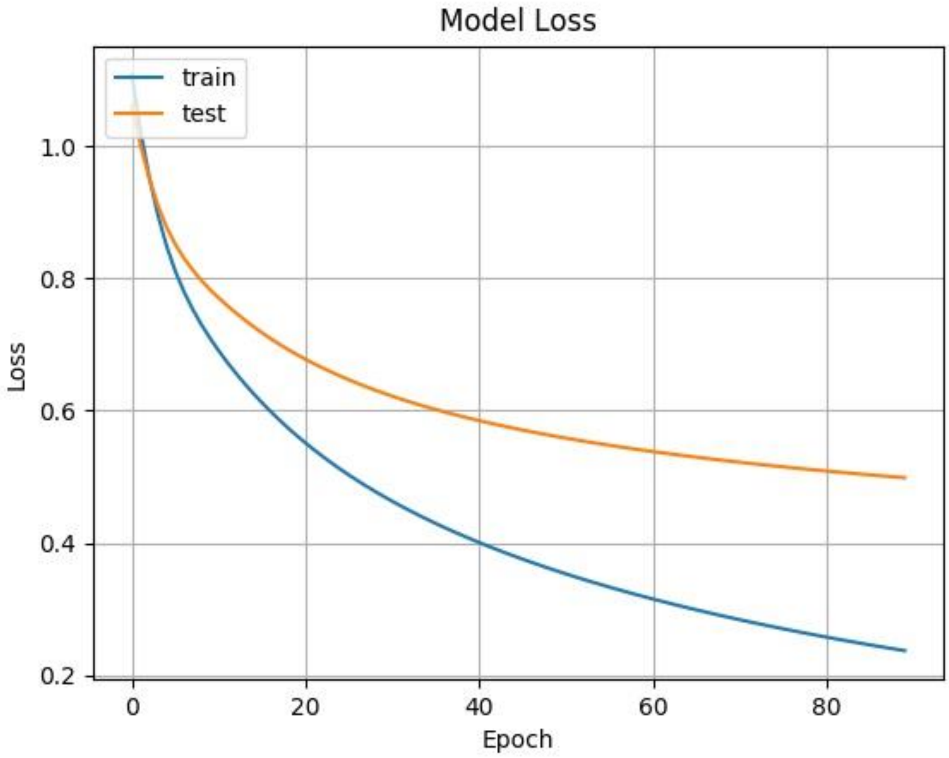

Figure 4.8 : Train and test losses versus epoch for HOG vectors. Test set consists of 110 HOG vectors obtained from Set 1 images.

Table 4.2 : Comparison of memory usage and accuracy results before and after dimensionality reduction by autoencoder (AE) and PCA. Memory presents the memory space occupied by HOG vectors of images in Set 1. Accuracy column shows the object recognition accuracy obtained using Euclidean distance measure.

	Memory	Accuracy
1764 D	1.55 MB	53 %
882 D (AE)	776.2 KB	58 %
882 D (PCA)	776.2 KB	55 %

After that, the experiments are repeated on images with noise. First, object recognition is tested on the noisy images. Then, the dimensionality of noisy features is reduced with proposed method and PCA for comparison and object recognition with reduced features is tested. The loss graph of the autoencoder which is trained with the training set of 220 vectors for 33 epochs, is shown in Figure 4.9. The results are shown in Table 4.3.

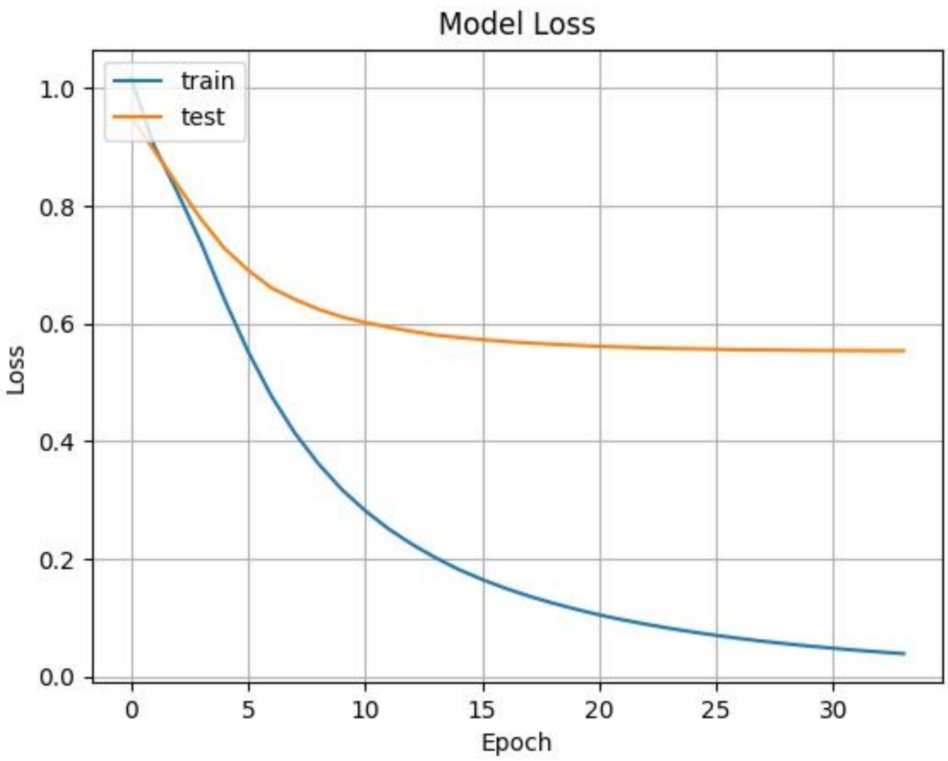

Figure 4.9 : Train and test losses versus epoch for noisy HOG vectors. Test set consists of 110 noisy HOG vectors obtained from Set 1 images.

Table 4.3 : Comparison of memory usage and accuracy results before and after dimensionality reduction by autoencoder (AE) and PCA. Memory presents the memory space occupied by HOG vectors of noisy images in Set 1. Accuracy column shows the object recognition accuracy obtained using Euclidean distance measure.

	Memory	Accuracy
1764 D	1.55 MB	54 %
882 D (AE)	776.2 KB	61 %
882 D (PCA)	776.2 KB	59 %

4.4.2 Dimensionality reduction of SIFT features

For dimensionality reduction of SIFT features, a 5-layered autoencoder is used, which is shown in Figure 4.3. The autoencoder is trained with the training set of 76102 vectors for 53 epochs and validated with the validation set. The loss graph of the model is shown in Figure 4.10. After the training step, vectors in the test set is inputted into the autoencoder. The vectors obtained in the code layer are stored, which are the reduced representations of these vectors. Then the low-dimensional vectors are used in object recognition task as mentioned in Section 3. To compare our results, the dimensionality of SIFT features is reduced to the same size by PCA as Valenzuela

(2012) stated. For this purpose random training sets are used for PCA training, and then system is validated using object recognition accuracy of reduced vectors of validation set as shown in Figure 4.6. In this manner, reduced vectors of the Set 2 is obtained by PCA, too.

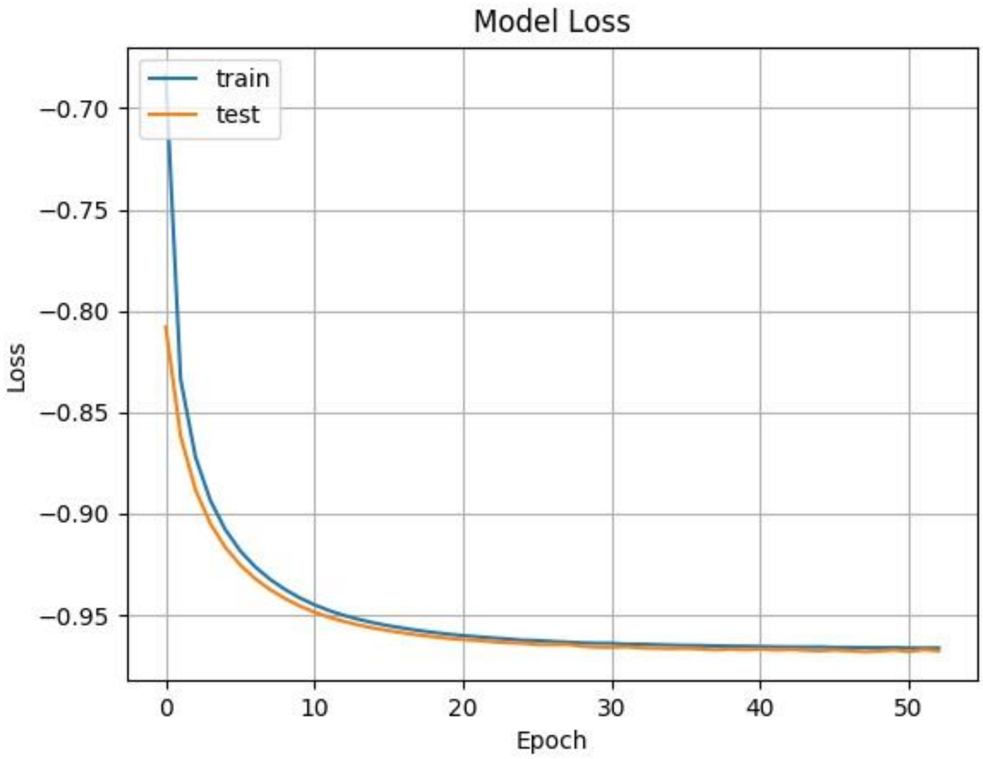

Figure 4.10 : Train and test losses versus epoch for SIFT vectors. Test set consists of 38051 SIFT vectors obtained from Set 2 images.

It should be highlighted that the maximum accuracy of object recognition by original sized SIFT features is observed while using the Euclidean and Cosine distance. Hence, it is aimed to improve object recognition results using Cosine distance because of being one of the most successful measures among the selected measures. So, in order to train the autoencoder, cosine proximity is used as loss function and also object recognition accuracies of validation set using Cosine distance is used to validate the PCA. Accuracy and memory usage results of object recognition using Cosine distance, before and after the dimensionality reduction by both of the autoencoder and PCA is given in Table 4.4.

After that, the experiments are repeated on images with noise. First, object recognition is tested on the noisy images. Then, the dimensionality of noisy features is reduced with proposed method and PCA for comparison and object recognition with reduced features is tested. The loss graph of the autoencoder which is trained with the training set of 92630 vectors for 90 epochs, is shown in Figure 4.11. The results are presented in Table 4.5.

Table 4.4 : Comparison of memory usage and accuracy results before and after dimensionality reduction by autoencoder (AE) and PCA. Memory presents the memory space occupied by SIFT vectors of images in Set 2. Accuracy column shows the object recognition accuracy obtained using Cosine distance measure.

	Memory	Accuracy
128 D	38.96 MB	47 %
64 D (AE)	19.48 MB	49 %
64 D (PCA)	19.48 MB	46 %

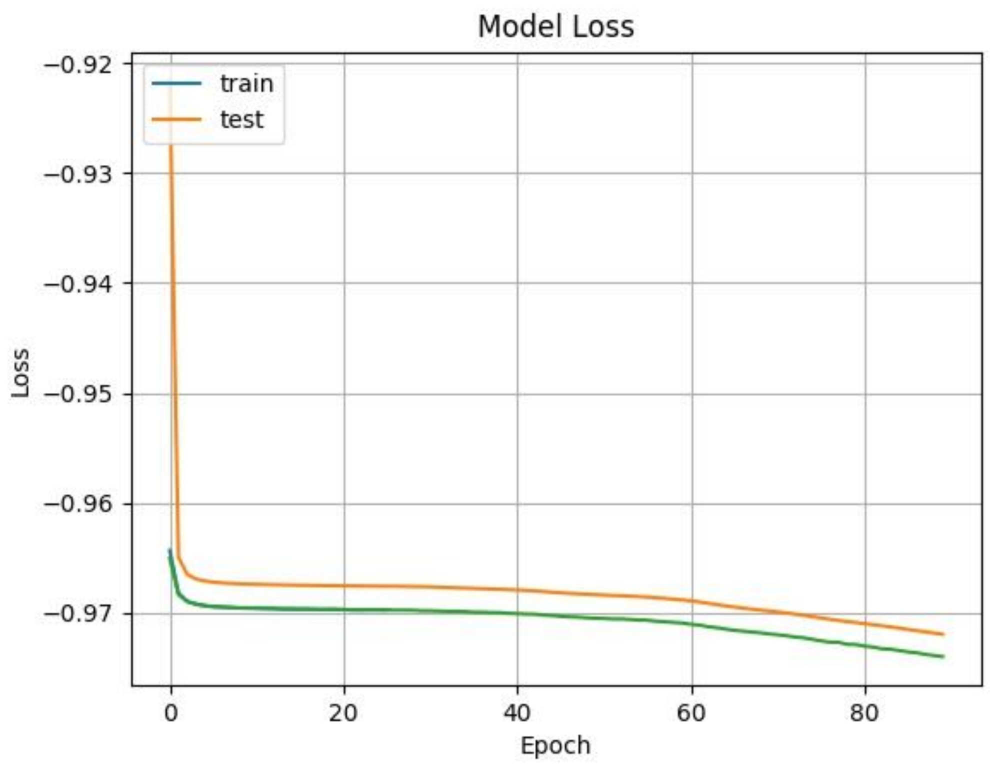

Figure 4.11 : Train and test losses versus epoch for noisy SIFT vectors. Test set consists of 46315 noisy SIFT vectors obtained from Set 2 images.

4.4.3 Dimensionality reduction of SURF features

For dimensionality reduction of SURF features, a 5-layered autoencoder is used, which is shown in Figure 4.4. The autoencoder is trained with the training set of 47790 vectors for 25 epochs and validated with the validation set. The loss graph of the model is shown in Figure 4.12. After the training step, vectors in the test set is inputted into the autoencoder. The vectors obtained in the code layer are stored, which are the reduced representations of these vectors. Then the low-dimensional vectors are used in object recognition task as mentioned in Section 3. To compare our results, the dimensionality of SURF features is reduced to the same size by PCA as Valenzuela (2012) stated. For this purpose random training sets are used for PCA training, and then system is validated using object recognition accuracy of reduced vectors of validation set as shown in Figure 4.6. In this manner, reduced vectors of the Set 3 is obtained by PCA , too.

Table 4.5 : Comparison of memory usage and accuracy results before and after dimensionality reduction by autoencoder (AE) and PCA. Memory presents the memory space occupied by SIFT vectors of noisy images in Set 2. Accuracy column shows the object recognition accuracy obtained using Cosine distance measure.

	Memory	Accuracy
128 D	47.43 MB	37 %
64 D (AE)	23.7 MB	35 %
64 D (PCA)	23.7 MB	35 %

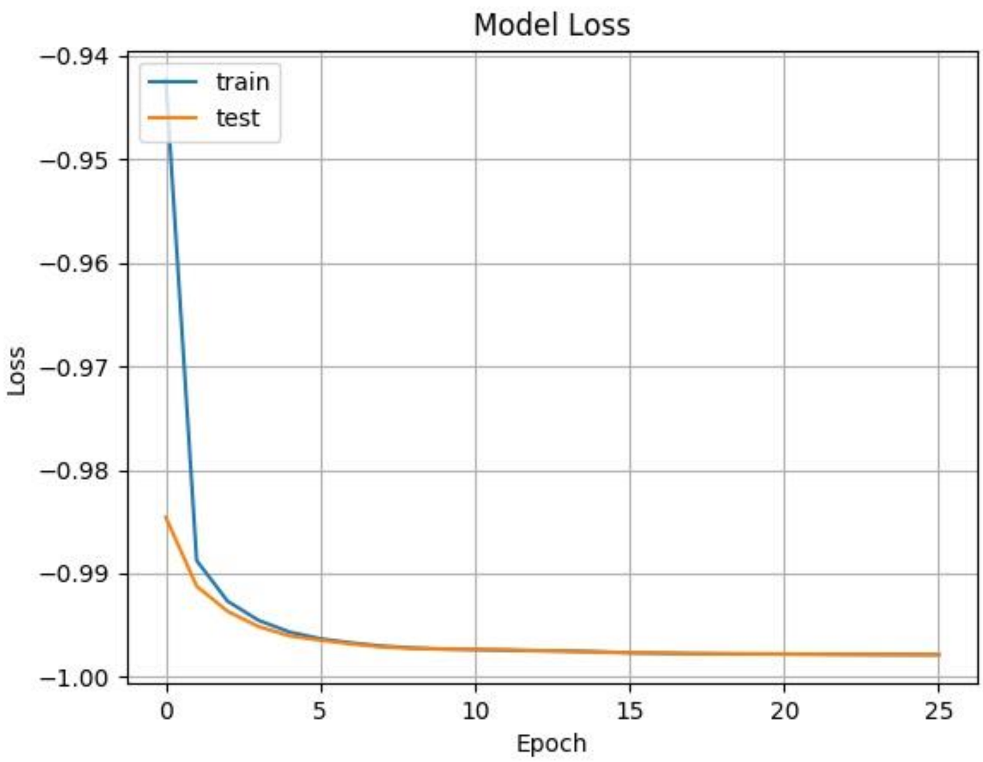

Figure 4.12 : Train and test losses versus epoch for SURF vectors. Test set consists of 23895 SURF vectors obtained from Set 3 images.

It should be noted that the maximum accuracy of object recognition by original sized SURF features is observed while using the Euclidean and Cosine distance. Hence, it is aimed to improve object recognition results using Cosine distance because of being one of the most successful measures among the selected measures. So, in order to train the autoencoder, cosine proximity is used as loss function and also object recognition accuracies of validation set using Cosine distance is used to validate the PCA. Accuracy and memory usage results of object recognition using Cosine distance, before and after the dimensionality reduction by both of the autoencoder and PCA is given in Table 4.6.

Table 4.6 : Comparison of memory usage and accuracy results before and after dimensionality reduction by autoencoder (AE) and PCA. Memory presents the memory space occupied by SURF vectors of images in Set 3. Accuracy column shows the object recognition accuracy obtained using Cosine distance measure.

	Memory	Accuracy
64 D	12.23 MB	40 %
32 D (AE)	6.12 MB	42 %
32 D (PCA)	6.12 MB	42 %

After that, the experiments are repeated on images with noise. First, object recognition is tested on the noisy images. Then, the dimensionality of noisy features is reduced with proposed method and PCA for comparison and object recognition with reduced features is tested. The loss graph of the autoencoder which is trained with the training set of 76400 vectors for 90 epochs, is shown in Figure 4.13. The results are shown in Table 4.7.

Table 4.7 : Comparison of memory usage and accuracy results before and after dimensionality reduction by autoencoder (AE) and PCA. Memory presents the memory space occupied by SURF vectors of noisy images in Set 3. Accuracy column shows the object recognition accuracy obtained using Cosine distance measure.

	Memory	Accuracy
64 D	19.56 MB	40 %
32 D (AE)	9.78 MB	44 %
32 D (PCA)	9.78 MB	43 %

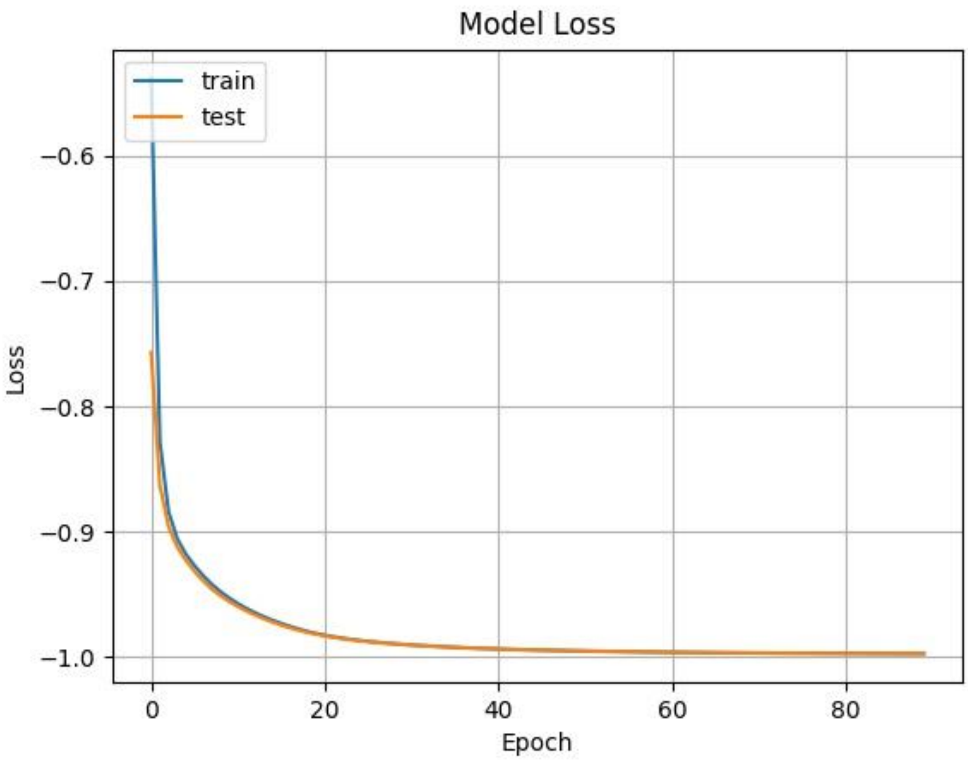

Figure 4.13 : Train and test losses versus epoch for noisy SURF vectors. Test set consists of 38200 noisy SURF vectors obtained from Set 3 images.

4.4.4 Comparison of the proposed method and PCA

To show the effect of the dimensionality reduction, the original features are inputted into object recognition firstly. The results are obtained for each set and for three distance measures and shown in Table 4.8.

Table 4.8 : Object recognition accuracies for each set and three distance measures before dimensionality reduction.

	HOG (1764-D) (Set 1)	SIFT (128-D) (Set 2)	SURF (64-D) (Set 3)
Cosine D.	51 %	47 %	40 %
Euclidean D.	53 %	47 %	40 %
KLD	26 %	15 %	23 %

By reducing the dimensionality of the feature vectors to the half, both of the useless dimensions and the noise are eliminated. Table 4.9 and Table 4.10 show the new object recognition results obtained with half dimensional feature vectors using proposed method and PCA, respectively.

Table 4.9 : Object recognition accuracies for each set and three distance measures after dimensionality reduction by proposed method.

	HOG (882-D) (Set 1)	SIFT (64-D) (Set 2)	SURF (32-D) (Set 3)
Cosine D.	57 %	49 %	42 %
Euclidean D.	58 %	39 %	39 %
KLD	58 %	35 %	13 %

Table 4.10 : Object recognition accuracies for each set and three distance measures after dimensionality reduction by PCA.

	HOG (882-D) (Set 1)	SIFT (64-D) (Set 2)	SURF (32-D) (Set 3)
Cosine D.	55 %	46 %	42 %
Euclidean D.	55 %	47 %	37 %
KLD	42 %	15 %	15 %

After obtaining these results, noisy sets of images are constructed. The noisy features are inputted into object recognition process and the results are presented in Table 4.11.

Table 4.11 : Object recognition accuracies for each noisy set of images and three distance measures before dimensionality reduction.

	HOG (1764-D) (Set 1)	SIFT (128-D) (Set 2)	SURF (64-D) (Set 3)
Cosine D.	54 %	37 %	40 %
Euclidean D.	54 %	37 %	40 %
KLD	53 %	13 %	12 %

By reducing the dimensionality of the feature vectors to the half, both of the useless dimensions and the noise are eliminated. Table 4.12 and Table 4.13 show the new object recognition results obtained with half dimensional feature vectors using proposed method and PCA, respectively.

Table 4.12 : Object recognition accuracies for each noisy set of images and three distance measures after dimensionality reduction by proposed method.

	HOG (882-D) (Set 1)	SIFT (64-D) (Set 2)	SURF (32-D) (Set 3)
Cosine D.	58 %	35 %	44 %
Euclidean D.	61 %	33 %	42 %
KLD	53 %	24 %	28 %

Table 4.13 : Object recognition accuracies for each noisy set of images and three distance measures after dimensionality reduction by PCA.

	HOG (882-D) (Set 1)	SIFT (64-D) (Set 2)	SURF (32-D) (Set 3)
Cosine D.	54 %	35 %	43 %
Euclidean D.	59 %	36 %	37 %
KLD	40 %	13 %	23 %

5. CONCLUSIONS

In this thesis, effect of dimensionality reduction on object recognition problem using autoencoders is studied.

For this purpose, three well-known feature vectors for object recognition are selected, which are HOG, SIFT and SURF features. The dimensionality reduction of these vectors is achieved using autoencoders which are neural networks that learn in unsupervised manner. Firstly, subsets for each feature type are constructed once and object recognition results are noted. Then features are reduced to half size using autoencoders. Finally, reduced dimensional vectors are inputted into object recognition and the results are compared with first results. Also, results of dimensionality reduction is obtained using PCA which is used mostly to reduce dimensionality of these feature vectors in the literature, for comparison. Moreover noisy image sets are constructed and all experiments are repeated on these images.

As the initial step, object recognition is tested on original sized features using three distance measures. The most successful distance measure is determined for each feature set considering the accuracies of object recognition. These distance measures were Euclidean distance for HOG features and Cosine distance for SIFT and SURF features. Thus, improving the performance of these measures for related feature sets is aimed. Moreover, the results of unselected distance measures are reported , too.

The results indicate that accuracy gain of 5% and memory saving of 50% are achieved after dimensionality reduction of original HOG features using the proposed method. Moreover, the proposed method outperforms PCA according to three distance measures, as shown in Figure 4.14.

The proposed method provide accuracy gain of 2% on object recognition using original SIFT features and memory saving of 50% as demonstrated in Figure 4.15. However, PCA could not achieve the success of object recognition using original sized features.

PCA and the proposed method showed the same performance on dimensionality reduction of SURF features as presented in Figure 4.16. By both of the methods, accuracy gain of 2% and memory saving of 50% are achieved for original SURF features.

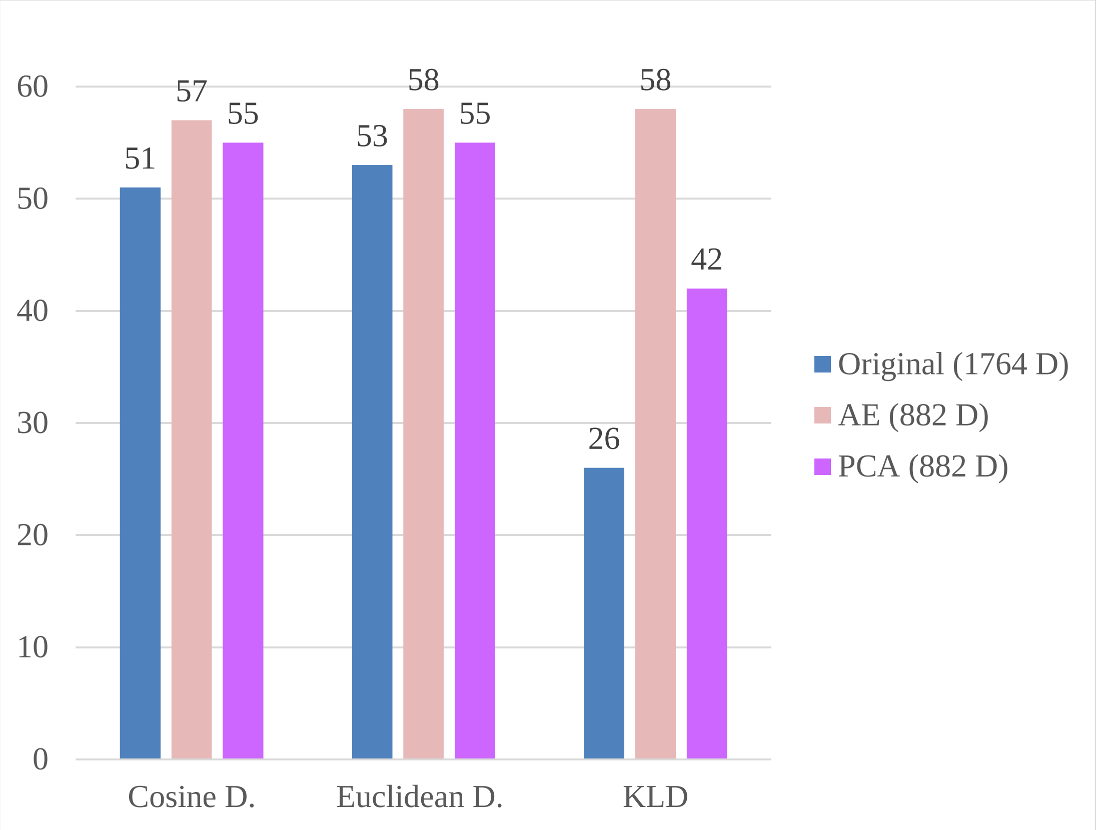

Figure 4.14 : Accuracies of object recognition using original HOG features.

The results demonstrate that object recognition performance of noisy HOG features is improved thanks to the dimensionality reduction using the proposed method. As demonstrated in Figure 4.17, accuracy gain of 7% is achieved by the proposed method, which outperforms PCA according to three distance measures.

On dimensionality reduction of noisy SIFT features, PCA and the proposed method showed the same performance on selected distance measure, as demonstrated in Figure 4.18. Both of the methods provide memory saving of 50% in exchange for loss in accuracy of 2%.

Accuracy gain of 4% is achieved using the proposed method for dimensionality reduction of noisy SURF features. Although both of the methods provide memory saving of 50%, the proposed method outperforms PCA on dimensionality reduction regardless which distance measure is used, as shown in Figure 4.19.

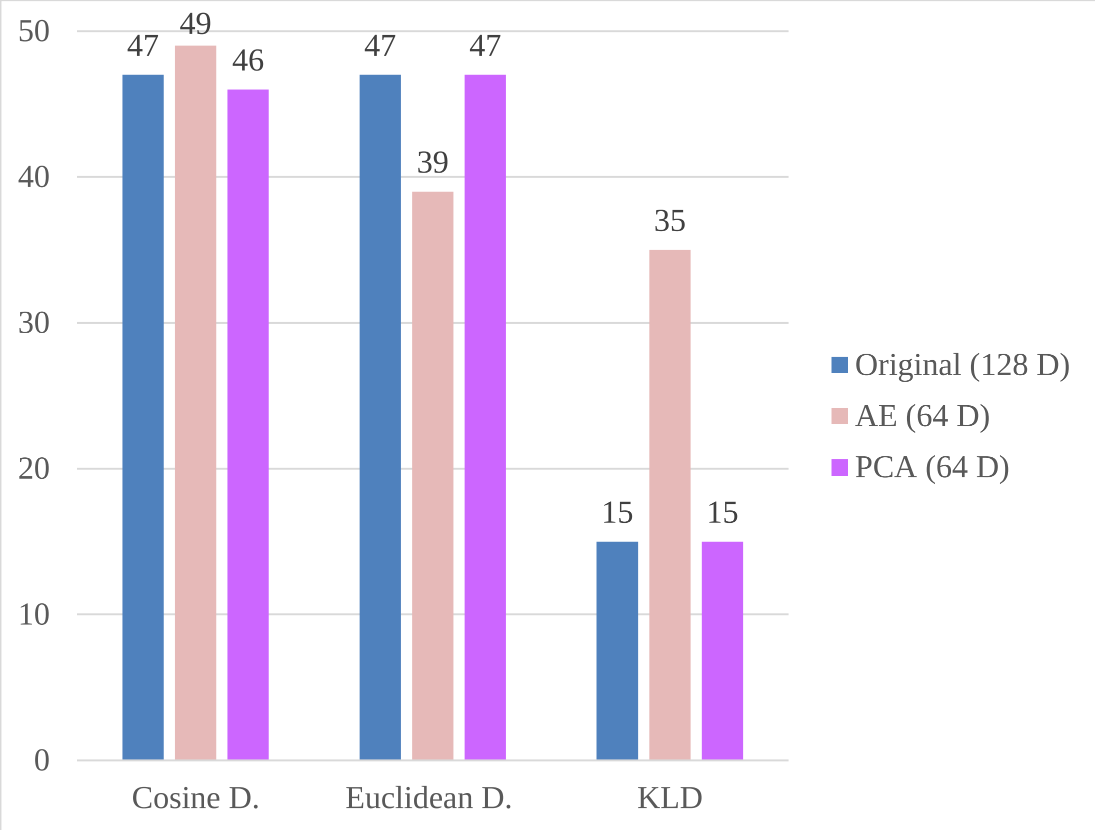

Figure 4.15 : Accuracies of object recognition using original SIFT features.

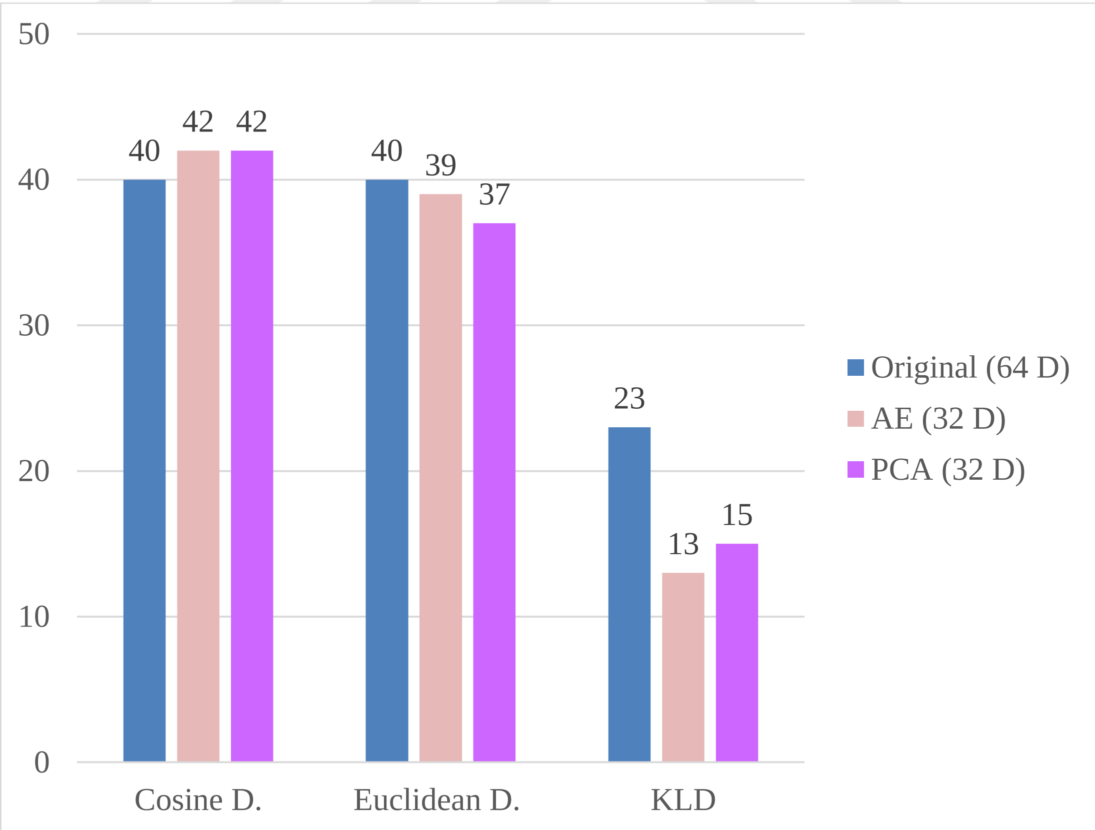

Figure 4.16 : Accuracies of object recognition using original SURF features.

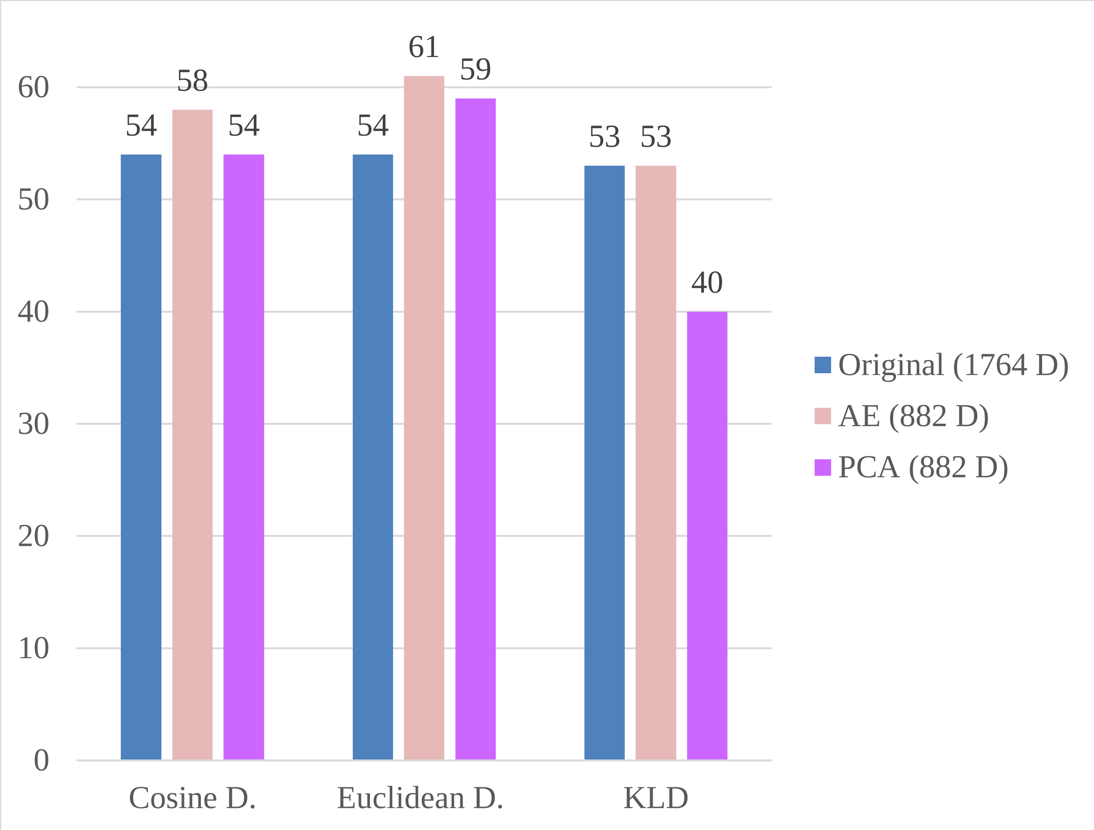

Figure 4.17 : Accuracies of object recognition using noisy HOG features.

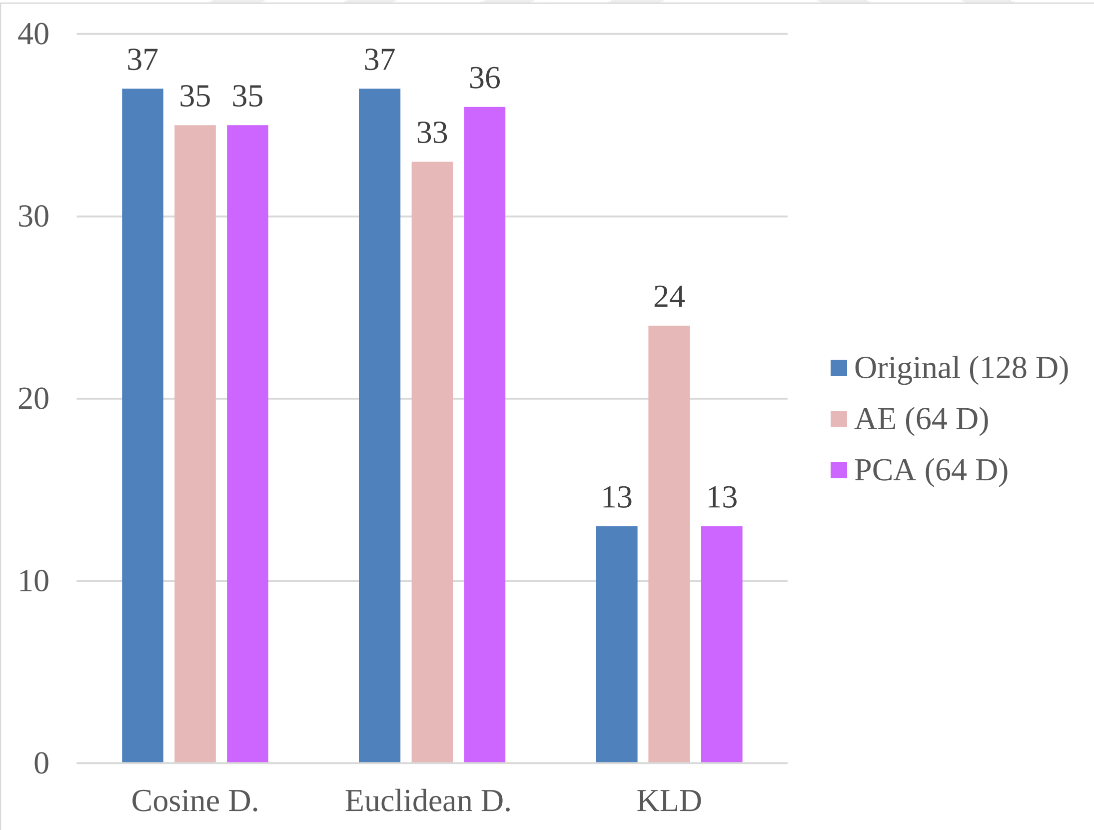

Figure 4.18 : Accuracies of object recognition using noisy SIFT features.

In conclusion, the results demonstrate that dimensionality reduction provides better or acceptable results, while memory saving of 50% is achieved thanks to the proposed method.

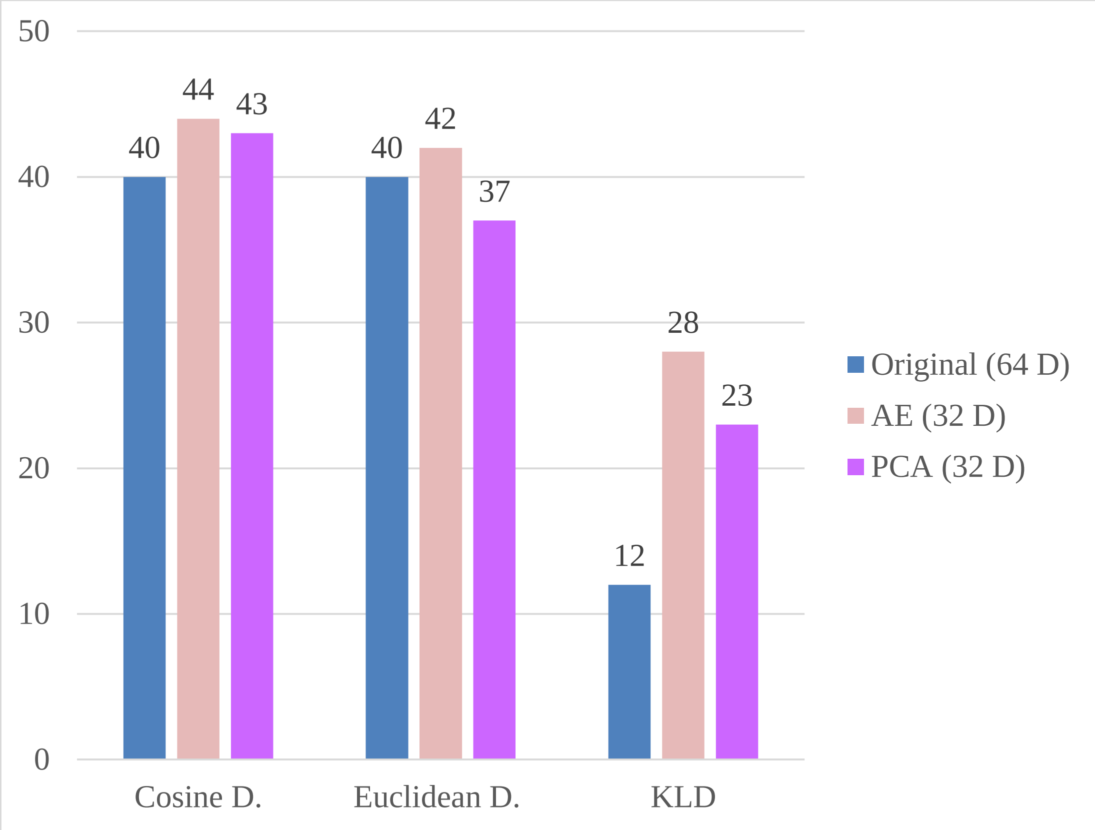

Figure 4.19 : Accuracies of object recognition using noisy SURF features.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Ghemawat, S. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

Alpaydin, E. (2009). *Introduction to machine learning*. MIT press, pp. 109-139.

Asbach, M., Hosten, P., & Unger, M. (2008). An evaluation of local features for face detection and localization. In Ninth International Workshop on Image Analysis for Multimedia Interactive Services (pp. 32-35). IEEE.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features. In *European conference on computer vision* (pp. 404-417). Springer, Berlin, Heidelberg.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (pp. 33-39). New York, NY: Springer-Verlag

Boulkenafet, Z., Komulainen, J., & Hadid, A. (2017). Face antispoofing using speeded-up robust features and fisher vector encoding. IEEE Signal Processing Letters, 24(2), 141-145.

Brown, M., & Süsstrunk, S. (2011). Multi-spectral SIFT for scene category recognition. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on (pp. 177-184). IEEE.

Cai, H., Mikolajczyk, K., & Matas, J. (2011). Learning linear discriminant projections for dimensionality reduction of image descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(2), 338-352.

Cevikalp, H., Verbeek, J., Jurie, F., & Klaser, A. (2008). Semi-supervised dimensionality reduction using pairwise equivalence constraints. In VISAPP'08-3rd International Conference on Computer Vision Theory and Applications (Vol. 1, pp. 489-496). INSTICC.

Chandrasekhar, V., Takacs, G., Chen, D., Tsai, S. S., Singh, J., & Girod, B. (2009). Transform coding of image feature descriptors. In Visual Communications and Image Processing 2009 (Vol. 7257, p. 725710). International Society for Optics and Photonics.

Chollet, F. (2015). Keras.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886-893). IEEE.

Déniz, O., Bueno, G., Salido, J., & De la Torre, F. (2011). Face recognition using histograms of oriented gradients. Pattern Recognition Letters, 32(12), 1598-1603.

Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. Journal of bioinformatics and computational biology, 3(02), 185-205.

Farquhar, J., Szedmak, S., Meng, H., & Shawe-Taylor, J. (2005). Improving" bag-of-keypoints" image categorisation: Generative models and pdf-kernels.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence, 32(9), 1627-1645.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of eugenics, 7(2), 179-188.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. http://www.deeplearningbook.org

Grauman, K., & Leibe, B. (2011). Visual object recognition. *Synthesis lectures on artificial intelligence and machine learning*, *5*(2), 1-181.

Griffin, G., Holub, A., & Perona, P. (2007). Caltech-256 object category dataset.

Hussain, S. U., & Triggs, W. (2010). Feature sets and dimensionality reduction for visual object detection. In BMVC 2010-British Machine Vision Conference (pp. 112-1). BMVA Press.

Ke, Y., & Sukthankar, R. (2004). PCA-SIFT: A more distinctive representation for local image descriptors. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on(Vol. 2, pp. II-II). IEEE.

Keser, R. K., Ergün, E., & Töreyin, B. U. (2018, January 9). ITU Vehicle Logo Dataset. Retrieved from https://kovan.itu.edu.tr/index.php/s/o36RNe6ac6cahli

Keser, R. K., Ergün, E., & Töreyin, B. U. (2018). Vehicle Logo Recognition with Reduced-Dimension SIFT Vectors Using Autoencoders. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 2, No. 2, p. 92).

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kobayashi, T., Hidaka, A., & Kurita, T. (2007). Selection of histograms of oriented gradients features for pedestrian detection. In International conference on neural information processing (pp. 598-607). Springer, Berlin, Heidelberg.

Ledwich, L., & Williams, S. (2004). Reduced SIFT features for image retrieval and indoor localisation. In Australian conference on robotics and automation (Vol. 322, p. 3).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

Lowe, D. G. (2004). U.S. Patent No. 6,711,293. Washington, DC: U.S. Patent and Trademark Office.

Lu, W., & Little, J.J. (2006). Simultaneous tracking and action recognition using the PCA-HOG descriptor. The 3rd Canadian Conference on Computer and Robot Vision (CRV'06), 6-6.

Mathias, M., Timofte, R., Benenson, R., & Van Gool, L. (2013). Traffic sign recognition-How far are we from the solution?. In Neural Networks (IJCNN), The 2013 International Joint Conference on (pp. 1-8). IEEE.

Medialab LPR database. (n.d.). Retrieved July 15, 2017, from http://www.medialab.ntua.gr/research/LPRdatabase.html

Mikolajczyk, K., & Matas, J. (2007). Improving descriptors for fast tree matching by optimal linear projection. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on (pp. 1-8). IEEE.

Misra, A., Abe, T., & Deguchi, K. (2011). Hand gesture recognition using histogram of oriented gradients and partial least squares regression. In MVA (pp. 479-482).

Monzo, D., Albiol, A., Albiol, A., & Mossi, J. M. (2011). Color HOG-EBGM for face recognition. In Image Processing (ICIP), 2011 18th IEEE International Conference on(pp. 785-788). IEEE.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11), 559-572.

Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. (2007). Object retrieval with large vocabularies and fast spatial matching. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on (pp. 1-8). IEEE.

Prasath, V. B., Alfeilat, H. A. A., Lasassmeh, O., & Hassanat, A. (2017). Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier-A Review. arXiv preprint arXiv:1708.04321.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. science, 290(5500), 2323-2326.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088), 533.

Savakis, A., Sharma, R., & Kumar, M. (2014). Efficient eye detection using HOG-PCA descriptor. In Imaging and Multimedia Analytics in a Web and Mobile World 2014 (Vol. 9027, p. 90270J). International Society for Optics and Photonics.

Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural computation, 10(5), 1299-1319.

Schwartz, W. R., Kembhavi, A., Harwood, D., & Davis, L. S. (2009). Human detection using partial least squares analysis. In Computer vision, 2009 IEEE 12th international conference on (pp. 24-31). IEEE.

Spearman, C. (1904). General intelligence objectively determined and measured. American Journal of Psychology, 15, 206-221.

Stommel, M., & Herzog, O. (2009). Binarising SIFT-descriptors to reduce the curse of dimensionality in histogram-based object recognition. In Signal

Processing, Image Processing and Pattern Recognition (pp. 320-327). Springer, Berlin, Heidelberg.

Strecha, C., Bronstein, A., Bronstein, M., & Fua, P. (2012). LDAHash: Improved matching with smaller descriptors. IEEE transactions on pattern analysis and machine intelligence, 34(1), 66-78.

Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. science, 290(5500), 2319-2323.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. Psychometrika, 17(4), 401-419.

Tuytelaars, T., & Schmid, C. (2007). Vector quantizing feature space with a regular lattice. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on (pp. 1-8). IEEE.

Valenzuela, R. E. G., Schwartz, W. R., & Pedrini, H. (2012). Dimensionality reduction through PCA over SIFT and SURF descriptors. In Cybernetic Intelligent Systems (CIS), 2012 IEEE 11th International Conference on (pp. 58-63). IEEE.

Valenzuela, R. E. G., Schwartz, W. R., & Pedrini, H. (2014). Linear dimensionality reduction applied to scale invariant feature transformation and speeded up robust feature descriptors. Journal of Electronic Imaging, 23(3), 033017.

Valenzuela, R., Pedrini, H., & Schwartz, W. (2013). Dimensionality reduction through LDA and Bag-of-features applied to image retrieval. In Proc. IV ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing (pp. 31-37).

Vinay, A., Vasuki, V., Bhat, S., Jayanth, K. S., Murthy, K. B., & Natarajan, S. (2016). Two dimensionality reduction techniques for surf based face recognition. Procedia Computer Science, 85, 241-248.

Wang, Q. J., & Zhang, R. B. (2008). LPP-HOG: A new local image descriptor for fast human detection. In Knowledge Acquisition and Modeling Workshop, 2008. KAM Workshop 2008. IEEE International Symposium on (pp. 640-643). IEEE.

Watcharapinchai, N., Aramvith, S., Siddhichai, S., & Marukatat, S. (2009). Dimensionality reduction of SIFT using PCA for object categorization. In Intelligent Signal Processing and Communications Systems, 2008. ISPACS 2008. International Symposium on (pp. 1-4). IEEE.

Xiao, X. Y., Hu, R., Zhang, S. W., & Wang, X. F. (2010). HOG-based approach for leaf classification. In Advanced intelligent computing theories and applications. with aspects of artificial intelligence (pp. 149-155). Springer, Berlin, Heidelberg.

Yeo, C., Ahammad, P., & Ramchandran, K. (2008). Rate-efficient visual correspondences using random projections. In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on (pp. 217-220). IEEE.

CURRICULUM VITAE

Name Surname : Reyhan Kevser Keser

Place and Date of Birth : İstanbul / Turkey, 1994

E-Mail : keserr@itu.edu.tr

B.Sc. : Electronics and Telecommunications Engineering, Istanbul Technical University, 2016

Professional Experience and Rewards:

Istanbul Technical University, Research Assistant (2017 - …)

PUBLICATIONS/PRESENTATIONS ON THE THESIS

- **Keser, R. K.,** Ergün, E., & Töreyin, B. U. (2018, January). Vehicle Logo Recognition with Reduced-Dimension SIFT Vectors Using Autoencoders. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 2, No. 2, p. 92).