Pós-Graduação em Ciência da Computação

Fidel Alejandro Guerrero Peña

A BAYESIAN FRAMEWORK FOR OBJECT RECOGNITION UNDER SEVERE OCCLUSION

Federal University of Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE
2017

Fidel Alejandro Guerrero Peña

A BAYESIAN FRAMEWORK FOR OBJECT RECOGNITION UNDER SEVERE OCCLUSION

A M.Sc. Dissertation presented to the Center for Informatics of Federal University of Pernambuco in partial fulfillment of the requirements for the degree of Master of Science in Computer Science.

Advisor: *Germano Crispim Vasconcelos*

RECIFE
2017

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

G934b Guerrero Peña, Fidel Alejandro
A Bayesian framework for object recognition under severe occlusion / Fidel Alejandro Guerrero Peña. – 2017.
107 f.: il., fig., tab.

Orientador: Germano Crispim Vasconcelos.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2017.
Inclui referências e apêndice.

1. Inteligência computacional. 2. Reconhecimento de objeto. I. Vasconcelos, Germano Crispim (orientador). II. Título.

006.3 CDD (23. ed.) UFPE- MEI 2017-108

Fidel Alejandro Guerrero Peña

A Bayesian Framework for Object Recognition under Severe Occlusion

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Aprovado em: 22/02/2017.

BANCA EXAMINADORA

_____

Prof. Dr. Tsang Ing Ren
Centro de Informática / UFPE

_____

Prof. Dr. Alexandre Magno Andrade Maciel
Escola Politécnica de Pernambuco/UPE

_____

Prof. Dr. Germano Crispim Vasconcelos
Centro de Informática / UFPE
(**Orientador**)

To my family.

Acknowledgements

I want to thanks my wife for being with me in good and bad, for bearing me, for understanding me and for her support and love. Thanks my parents and sister for their infinite love, for encouraging me to be better and for their help ALL the time. Thanks my advisor Prof. Germano Vasconcelos for accepting me as his student and his support both professionally and personally. Thanks to all the professors at informatics center for their excellent classes and help. Thanks to all my friends for all the good times.

Abstract

Shape classification has multiple applications. In real scenes, shapes may contain severe occlusions, hardening the identification of objects. In this work, a bayesian framework for object recognition under severe and varied conditions of occlusion is proposed. The proposed framework is capable of performing three main steps in object recognition: representation of parts, retrieval of the most probable objects and hypotheses validation for final object identification. Occlusion is dealt with separating shapes into parts through high curvature points, then tangent angle signature is found for each part and continuous wavelet transform is calculated for each signature in order to reduce noise. Next, the best matching object is retrieved for each part using Pearson's correlation coefficient as query prior, indicating the similarity between the part representation and of the most probable object in the database. For each probable class, an ensemble of Hidden Markov Model (HMM) is created through training with the one-class approach. A sort of search space retrieval is created using class posterior probability given by the ensemble. For occlusion likelihood, an area term that measure visual consistency between retrieved object and occlusion is proposed. For hypotheses validation, a area constraint is set to enhance recognition performance eliminating duplicated hypotheses. Experiments were carried out employing several real world images and synthetical generated occluded objects datasets using shapes of CMU_KO and MPEG-7 databases. The MPEG-7 dataset contains 1500 test shape instances with different scenarios of object occlusion with varied levels of object occlusion, different number of object classes in the problem, and different number of objects in the occlusion. For real images experimentation the CMU_KO challenge set contains 8 single view object classes with 100 occluded objects per class for testing and 1 non occluded object per class for training. Results showed the method not only was capable of identifying highly occluded shapes (60%-80% overlapping) but also present several advantages over previous methods. The minimum F-Measure obtained in MPEG-7 experiments was 0.67, 0.93 and 0.92, respectively and minimum AUROC of 0.87 for recognition in CMU_KO dataset, a very promising result due to complexity of the problem. Different amount of noise and varied amount of search space retrieval visited were also tested to measure framework robustness. Results provided an insight on capabilities and limitations of the method, demonstrating the use of HMMs for sorting search space retrieval improved efficiency over typical unsorted version. Also, wavelet filtering consistently outperformed the unfiltered and sampling noise reduction versions under high amount of noise.

Keywords: Severe occlusion. Hidden Markov Model. Wavelet Transform. Object Recognition.

Resumo

A classificação da forma tem múltiplas aplicações. Em cenas reais, as formas podem conter oclusões severas, tornando difícil a identificação de objetos. Neste trabalho, propõe-se uma abordagem bayesiana para o reconhecimento de objetos com oclusão severa e em condições variadas. O esquema proposto é capaz de realizar três etapas principais no reconhecimento de objetos: representação das partes, recuperação dos objetos mais prováveis e a validação de hipóteses para a identificação final dos objetos. A oclusão é tratada separando as formas em partes através de pontos de alta curvatura, então a assinatura do ângulo tangente é encontrada para cada parte e a transformada contínua de wavelet é calculada para cada assinatura reduzindo o ruído. Em seguida, o objeto mais semelhante é recuperado para cada parte usando o coeficiente de correlação de Pearson como *prior* da consulta, indicando a similaridade entre a representação da parte e o objeto mais provável no banco de dados. Para cada classe provável, um sistema de múltiplos classificadores com Modelos Escondido de Markov (HMM) é criado através de treinamento com a abordagem de uma classe. Um ordenamento do espaço de busca é criada usando a probabilidade *a posterior* da classe dada pelos classificadores. Como verosimilhança de oclusão, é proposto um termo de área que mede a consistência visual entre o objeto recuperado e a oclusão. Para a validação de hipóteses, uma restrição de área é definida para melhorar o desempenho do reconhecimento eliminando hipóteses duplicadas. Os experimentos foram realizados utilizando várias imagens do mundo real e conjuntos de dados de objetos oclusos gerados de forma sintética usando formas dos bancos de dados CMU_KO e MPEG-7. O conjunto de dados MPEG-7 contém 1500 instâncias de formas de teste com diferentes cenários de oclusão por exemplo, com vários níveis de oclusões de objetos, número diferente de classes de objeto no problema e diferentes números de objetos na oclusão. Para a experimentação de imagens reais, o desafiante conjunto CMU_KO contém 8 classes de objeto na mesma perspectiva com 100 objetos ocluídos por classe para teste e 1 objeto não ocluso por classe para treinamento. Os resultados mostraram que o método não só foi capaz de identificar formas altamente ocluídas (60% - 80% de sobreposição), mas também apresentar várias vantagens em relação aos métodos anteriores. A F-Measure mínima obtida em experimentos com MPEG-7 foi de 0.67, 0.93 e 0.92, respectivamente, e AUROC mínimo de 0.87 para o reconhecimento no conjunto de dados CMU_KO, um resultado muito promissor devido à complexidade do problema. Diferentes quantidades de ruído e quantidade variada de espaço de busca visitado também foram testadas para medir a robustez do método. Os resultados forneceram uma visão sobre as capacidades e limitações do método, demonstrando que o uso de HMMs para ordenar o espaço de busca melhorou a eficiência sobre a versão não ordenada típica. Além disso, a filtragem com wavelets superou consistentemente as versões de redução de ruído não filtradas e de amostragem sob grande quantidade de ruído.

Palavras-chave: Oclusão severa. Modelo Escondido de Markov. Transformada de Wavelet. Reconhecimento de Objetos.

List of Figures

List of Tables

List of Acronyms

AR	Autoregressive 26
AUC	Area Under Curve 71
AUROC	Area Under ROC 78
BRIEF	Binary Robust Independent Elementary Features 23
CSS	Curvature Scale Space 26
CV	Curvature 81
CWT	Continuous Wavelet Transform 48
DCE	Discrete Curve Evolution 31
DPM	Deformable Part Model 24
DTW	Dynamic Time Warping 30
FD	Fourier Descriptor 26
FT	Fourier Transform 43
HCP	High Curvature Points 57
HMM	Hidden Markov models 48
HOG	Histogram of Oriented Gradient 24
IDSC	Inner Distance Shape Context 30
MAP	Maximum A Posteriori 84
MCMC	Markov Chain Monte Carlo 30
MVM	Minimum Variance Matching 80
ORB	Oriented fast and Rotated BRIEF 23
PCA	Principal Component Analysis 24
RANSAC	Random Sample Consensus 32
ROC	Receiver Operating Characteristic 71
SDIM	Scale Invariant and Deformation tolerant partial shape Matching 80
SIFT	Scale Invariant Feature Transform
SURF	Speed-Up Robust Features 23
SVM	Support Vector Machines 24
SSR	Search Space Retrieval 65

STFT Short Time Fourier Transform 46

TAS Tangent Angle Signature 57

WD Wavelet Descriptor 26

WT Wavelet Transform 46

Contents

1 Introduction

The main purpose of an object recognition system is to detect and recognize objects of interest in a scene image taken in the real world, using object models which are known as priors. Object recognition has extensive applications in many areas, such as blood cell recognition and counting (GUERRERO-PENA et al., 2015), scene description (ZHOU et al., 2014), parts in assembly lines (ZHANG; XU; LIU, 2015), to mention just a few. Although humans perform object recognition almost effortlessly and instantaneously, implementation of this task in machines is very difficult. It is a major and also challenging task in computer vision. Many researchers have dedicated themselves into this area of research and made great contributions in the past few decades (TAUBIN; COOPER, 1991; TIENG; BOLES, 1997; YANG; LEE; LEE, 1998; BELONGIE; MALIK; PUZICHA, 2002; ZHANG; XU; LIU, 2015).

The object recognition problem can be defined as a labeling problem based on models of known objects. These model images for the individual objects are stored in the database and are identified with different labels. Given an image (scene image) containing one or more objects of interest and a set of labels corresponding to a set of models known to the recognition system, this system should match and assign correct label(s) to the objects in the scene image.

A vision recognition system can be thought of as an emulation of the human recognition process were objects can be detected and labeled if have seen before. Specifically, for object recognition, images are defined by features, representing and quantifying their distinguishable characteristics. To recognize objects of interest in a scene, features extracted from the objects are matched with the feature sets of all model images in the database. Those correctly matched features are called correspondences.

Object recognition can be challenging, because performances of recognition systems vary with vision applications, one of which is when objects are partially occluded. While humans can easily distinguish different objects in occlusion scenarios, object recognition methods (FRANSENS; STRECHA; VAN GOOL, 2006; HORÁČEK; KAMENICKÝ; FLUSSER, 2008; WANG; HAN; YAN, 2009; HSIAO; HEBERT, 2014) are far behind in this ability. General feature based recognition algorithms suffers from a major setback when applied to occluded object recognition, i.e., incomplete and/or deformed features.

However, the task of recognizing occluded objects is more challenging because dealing with the interactions of features from different objects in the scene. This problem is ill-posed, because there is no formal definition of what constitutes an object category.

While matching discriminative features work for the majority of feature-rich objects, many man-made objects contain repeated patterns from logos, text and printed graphics. Features extracted from those images have similar descriptors which, in extreme case, may be exactly the same. Current algorithms (LOWE, 2004; DALAL; TRIGGS, 2005; BAY et al., 2008) discard ambiguous matches because they are assumed to arise from background clutter. However, ignoring those matches often results in insufficient correspondences for reliable recognition.

Matching of non-deformable object shapes is an interesting as well as an important problem in computer vision since shape is one of the most distinguishing characteristics of an object, being unaffected by photometric changes and background variations. Furthermore, it has been found from human perception that, in the presence of challenges such as partial occlusions, local articulations, geometric distortions, intra-class variations and viewpoint changes, it is possible to identify and recognize an object simply from its shape. Thus, shape matching has been successfully used in various tasks such as object detection and classification (RAVISHANKAR; JAIN; MITTAL, 2008; LU et al., 2009; WANG et al., 2012; GUO et al., 2014), optical character recognition (BELONGIE; MALIK; PUZICHA, 2002), medical image registration (HUANG; PARAGIOS; METAXAS, 2006) and image retrieval (KUMAR et al., 2012).

In real scenarios, when an object is segmented out automatically using techniques such as Background Subtraction or Image Segmentation, the matching of the extracted contours should be robust to errors introduced by the segmentation process. For instance, the output of a Background Subtraction technique often misses out some portions of the object or adds some extra portions such as object shadows. Sometimes, two objects may be merged into one if they are close to each other. Similar problems exist due to the use of Image Segmentation techniques as well. A robust shape matching algorithm must deal with such variations and distractions in order to be useful in a practical scenario.

1.1 Motivation

In natural scenes, it is commonly found that objects are occluded by or occluding other objects or surfaces. Such situations result in partially visible objects in the scene, which is referred to as occlusion in the rest of this dissertation. For human beings, to recognize an object under severe occlusion could also be a difficult task.

Occlusion is a challenging problem, as it can severely compromise performances of many computer vision applications, such as path planning in autonomous robot navigation, target tracking and recognition in visual surveillance and segmentation from overlapping tissues in medical diagnosis.

Occlusion is also a major challenge for the accurate computation of visual correspon-

dence, such as stereo matching and image registration. In these cases, occluded pixels are visible in only one image, so there are no corresponding pixels in the other image. For 3-D reconstruction, it is particularly important to obtain good results at discontinuities, which are places where occlusions often occur. Ideally, a pixel in one image should correspond to at most one pixel in the other image, and a pixel that corresponds to no pixel in the other image should be labeled as occluded.

Object recognition is one of the most fundamental tasks in computer vision, still lacking a completely automatic solution. The main idea is to find a set of features that describes and discriminates the object of interest from the rest of the image. Object color is a low-level feature that can be used as such descriptor, although its discrimination capacity is often insufficient in real images. Using shape as a high-level feature is a common approach as an enhancement to such low-level features (LECUMBERRY; PARDO; SAPIRO, 2010).

Inherent to object boundaries, structural shape plays an important role in computer vision and object recognition. A fundamental problem is to match different shape instances and measure their shape similarity (or dissimilarity). In the 2D case, each shape instance can be represented by a closed contour that delineates the boundary of the structure of interest.

Partial shape matching is a challenging problem that aims to match an occluded shape contour to a template shape contour without occlusion by identifying the matched portions from the two contours. Different from full shape matching (SHARVIT et al., 1998; GDALYAHU; WEINSHALL, 1999; BELONGIE; MALIK; PUZICHA, 2002; MCNEILL; VIJAYAKUMAR, 2006), the number, the size and the location of the matched contour segments are usually all unknown in partial shape matching.

The study and improvement of techniques to partial shape matching is of great interest to both scientific community and industry. Existing methods generally needs quantity of objects that are present in occlusion to be given, because lacks of robust hypotheses validation steps (SABER; XU; Murat Tekalp, 2005; CHEN; FERIS; TURK, 2008; MICHEL; OIKONOMIDIS; ARGYROS, 2011). Also, some works needs occluded object categories to be given like in Ref. CAO et al. (2011) or part separation is manually made (LATECKI et al., 2007; MERHY et al., 2014). Therefore, occluded object recognition is developed in non-automatic way. All this methods were developed for partially occluded objects recognition, this means only one object's part is visible and the rest is occluded. However, several problems appears when occluded object has several non-consecutive visible parts, hereafter named as severe occluded object. Since new approaches that makes automatic and valid recognition for severe occluded objects are needed.

1.2 Problem Statement

This research has as main objective the investigation and develop an shape based object recognition method capable of dealing with severe occluded objects. Among other goals, the method aims to perform recognition of non-deformable severe occluded objects without any

information about (1) quantity of objects present in the occlusion, (2) which parts of the occluded contour belong to each object, and (3) the categories of occluded objects.

Specific objectives of this work are:

- Investigate the state of the art in object recognition, analyzing the current method capabilities for dealing with occlusions.
- Investigate the state of the art in shape matching and representation, and its applicability for occluded object recognition.
- Proposal and implementation of occluded object recognition method steps as feature extraction, feature grouping, hypotheses extraction and hypotheses validation.
- Validation and comparison of the proposed method with different occluded objects datasets.

1.3 Organization of the Dissertation

In the first part of the dissertation, its addressed an overview of state-of-the-art methods for object recognition, is conducted with respect to shape representation and partial shape matching. Next, background methods are presented for better understanding the method proposed in this work (Chapter 2). In Chapter 3 a new Bayesian method for severe occluded object recognition is detailed. A search space retrieval sorting based on an ensemble of Hidden Markov models and two new area constraints for occlusion likelihood in the bayesian method and hypothesis validation are introduced. Several experiments are carried out in Chapter 4 for determination of capabilities and limitations of proposed method and comparison to previous works. Finally, in Chapter 5 conclusions about method performance, limitations, contributions and future work are provided.

1.4 Publication Note

The publications which comprise this dissertation are listed below:

- Guerrero-Peña, F. A. and Vasconcelos, G. C. **Search-space sorting with hidden Markov models for occluded object recognition**. In *IEEE 8th International Conference on Intelligent Systems* (IS), 2016, pp. 47-52.
- Guerrero-Peña, F. A. and Vasconcelos, G. C. **Object Recognition Under Severe Occlusions with a Hidden Markov Model Approach**. *Pattern Recognition Letters*, 2016, 86, pp. 68-75.

Two other papers comprising other aspects and results developed in this work are under preparation for submission to other journal and conference.

- Guerrero-Peña, F. A. and Vasconcelos, G. C. **Wavelet Filtering Influence on Noisy Severe Occluded Object Recognition** (Draft title).
- Guerrero-Peña, F. A. and Vasconcelos, G. C. **A Bayesian Framework for Occluded Object Recognition with Dynamic Time Warping** (Draft title).

2

Overview of the Problem

Significant research in recognition of object instances from a single image has been made. In this Chapter, an overview of the literature approaches in this area and some necessary background are provided. First state of the art for object recognition, shape representation, partial shape matching and occlusion are presented. Then, segmentation and shape representation approaches used are discussed and wavelet transform and hidden Markov model are introduced for better understanding the proposed method.

2.1 State of the art

Literature review for object recognition, shape representations, partial shape matching and occluded object recognition applications is provided in this section. Main taxonomies and methods for each of them are discussed in each subsection.

2.1.1 Object Recognition

Object recognition consists of detecting and identifying all objects in a given image. This means to localize and label objects in particular categories of objects that depend on the database. Category recognition and object detection are two parts of the object recognition process. In typical object recognition approaches four components are used commonly: feature extraction, feature grouping, object hypothesis generation and object verification (ANDREOPOULOS; TSOTSOS, 2013) (see Figura 2.1). In this process, on of the most important step is feature extraction and extensive research has been dedicated to designing repeatable methods for this, making them scale and rotation invariant, and creating highly discriminative descriptors (MOKHTARIAN; MACKWORTH, 1986; ZAHN; ROSKIES, 1972; BELONGIE; MALIK; PUZICHA, 2002; ZHANG; LU, 2002, 2004).

Several methods have been proposed for object detection and recognition. They can be grouped in three categories, keypoint based; sliding windows based; and part based detectors (BRAHMBHATT, 2014).

Keypoint based algorithms try to find distinctive points in the test image and see if they

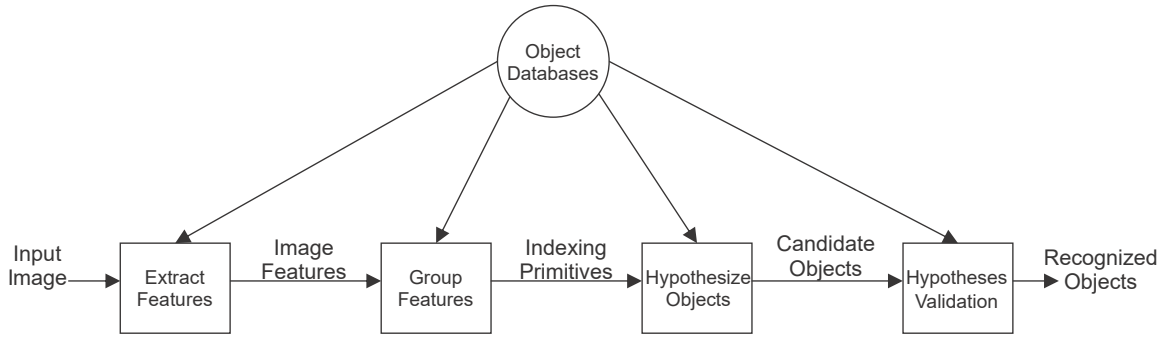


Figure 2.1: Components used in object recognition systems.

look like some stored keypoints extracted from training images in a feature space. If a lot of keypoint matches are founded in the test image region, the region gets a high score. Usually, the keypoints are extracted by maximizing some measure of distinctiveness and repeatability (HSIAO; HEBERT, 2014). Features are extracted from a region around the keypoint to capture information about the local appearance. By normalizing the feature descriptor extraction process with respect to scale and rotation the keypoint, descriptors gain invariance to changes in the size and rotation of the object. For example, popular Scale Invariant Feature Transform features (LOWE, 2004) define keypoints as extremes in the difference of Gaussian filter applied to the image pyramid (GONZALEZ; WOODS, 2008), and build the descriptor by stacking histograms of oriented gradients in a circular region around keypoints whose size is proportional to its scale in the pyramid. Speed-Up Robust Features (SURF) (BAY et al., 2008) also uses a histograms of oriented gradients based approach but uses integral images and box filter approximation of Gaussan filters to accelerate keypoint extraction. Oriented fast and Rotated BRIEF (ORB) (RUBLEE et al., 2011) reaches rotation invariance modifying Binary Robust Independent Elementary Features (BRIEF) (CALONDER et al., 2010), which uses a series of binary features extracted by comparing points intensity values around keypoints. The use of binary feature descriptors speeds up descriptor matching considerably. The advantage of keypoint based object detectors is that the search space is greatly reduced by the keypoint extraction step and partial occlusion can be handled because the object is not represented holistically (DICKINSON; PIZLO, 2015). On the other hand, to get good repeatable keypoints, both the keypoint extraction and description steps tend to be quite involved. Simpler descriptors tend to produce many false-positive matches at test time, leading to large high-dimensional matching problems. But the biggest drawback of keypoint based approaches is that since they are only interested in distinctive points, the object of interest needs to be well-textured. Additionally, since only sparse areas of the object are stored in the model, it is not possible to get a pixel level binary segmentation of the bounding box unless keypoint descriptors are extracted in a dense grid, which essentially reduces to a sliding window approach discussed next.

Sliding window based algorithms work densely on the whole image and often also at all image pyramid levels. Basically, the algorithms convert the image to some feature space and learn weights for a rectangular filter such that the convolution of the filter with the feature-image

will produce high scores at regions where the filter covers the object. The filter weights are often learned with Support Vector Machines (SVM) with a linear kernel. Recently, Histogram of Oriented Gradient (HOG) features (DALAL; TRIGGS, 2005) have become quite popular. HOG divides the image into small cells, usually 8x8 pixels, and constructs a histogram of gradient orientation from every pixel in the cell, weighted by the gradient magnitude at the cell. Histograms of blocks of cells are then L1 normalized to gain contrast invariance.

A Principal Component Analysis (PCA)-reduced and slightly modified version of HOG is used as the main feature in the highly successful Deformable Part Model (DPM) object detection system (FELZENSZWALB et al., 2010). The advantage of sliding window based methods is that detection is as simple as convolving a filter with the feature-image and picking the best after a non-maximum suppression step. However, the convolution needs to be done over the whole image pyramid. In addition, algorithms using simple features like HOG cannot be rotation invariant, a problem that has to be handled by learning different filters for different object poses and picking up the maximum response. Occlusion decreases the filter response. Hence, if a tight threshold is used on the score to increase precision, occlusion cases will not be handled, causing a sharp drop in recall on a challenging dataset.

All those methods work really well on book covers, paintings and cereal boxes, but are unable to obtain good matches with smooth objects (HSIAO; HEBERT, 2014). Yet, for many of these smooth objects, it is not because there is a lack of keypoints. The main difference is that the majority of keypoints in smooth objects are on the boundary at the intersection of background clutter with the object or at corners due to specularities and lighting effects. These types of keypoints are not repeatable, since they will not fire at the same locations if the background or lighting is slightly different. The only repeatable keypoints are corners of the object and are also not very informative. These corners are on the object boundary, resulting in a large portion of the descriptor capturing the background clutter. Since descriptors are designed to be highly discriminative, the same feature extracted in different backgrounds will have very different values. Thus, they will not be matched unless the background statistics around the keypoint are very similar as well, which is rarely the case.

Stochastic textures are also a problem for discriminative features. These textures have the same statistical properties between objects, but do not have the same exact appearance. Many objects have stochastic textures such as a wooden bowl or a fur coat. While many keypoints can be extracted from these textures, the features are usually not useful. For the same exact model of the object, the textures will not be exactly the same for two physical objects, leading to different descriptors. In addition, the keypoints are often difficult to localize in these regions. As an alternative, textureless objects are characterized by their shapes, although often ambiguous in many situations even without occlusions.

Object color and textures are low-level features that can be used as descriptor, although its discrimination capacity is often insufficient in real images. Using shape as a high-level feature is a common approach applied to improve performance (LECUMBERRY; PARDO; SAPIRO,

2010).

2.1.2 Shape Representation

The first task for any shape matching technique is to come up with a representation of shapes such that they can be matched efficiently and accurately. In the occluded object recognition problem, the inputs are assumed to be outer shape contours, that may be obtained from automatic techniques such as background subtraction and image segmentation or, in some cases, may be manually drawn.

Shape contours have often been represented by decomposition into small parts so that local transformations can be determined for each part. While decomposition of a shape is important, it is, however, a very challenging task, especially under occlusions or noise.

Shape representation and description techniques can generally be classified into two class of methods: (1) contour-based methods and (2) region-based methods (ZHANG; LU, 2004). The classification is based on whether shape features are extracted from the contour only or are extracted from the whole shape region. In each class, the different methods are further divided into structural approaches and global approaches. This sub-class is based on whether the shape is represented as a whole or represented by segments/sections (primitives). Such approaches can be further distinguished into space domain and transform domain, based on whether the shape features are derived from the spatial domain or from the transformed domain.

Contour shape techniques only exploit shape boundary information. There are generally two types of very different approaches for contour shape modeling: continuous approach (global) and discrete approach (local or structural). Continuous approaches do not divide shape into sub-parts, usually a feature vector derived from the integral boundary is used to describe the shape. The measure of shape similarity is usually a metric distance between the acquired feature vectors. Discrete approaches break the shape boundary into segments, called primitives using a particular criterion. The final representation is usually a string or a graph, the similarity measure is done by string matching or graph matching.

Common simple global descriptors are area, circularity ($perimeter^2/area$), eccentricity (length of major axis divided by length of minor axis), major axis orientation, and bending energy (YOUNG; WALKER; BOWIE, 1974). These simple global descriptors usually can only discriminate shapes with large differences and therefore, are usually used as filters to eliminate false hits or are combined with other shape descriptors to discriminate shapes.

Belongie et al. proposed a correspondence-based shape matching method using shape contexts (BELONGIE; MALIK; PUZICHA, 2002). To extract the shape context at a point p, the vectors of p to all the other boundary points are found. The length r and orientation θ of the vectors are quantized to create a histogram map which is used to represent the point p. The histogram of each point is flattened and concatenated to form the context of the shape. To make the histogram more sensitive to positions of nearby points than to those of points farther away, the vectors are put into log-polar space.

A shape signature represents a shape by a one dimensional function derived from shape boundary points. Many shape signatures exist, they include centroidal profile, complex coordinates, centroid distance, tangent angle, cumulative angle, curvature, triangle area and chord-length (COSTA; CESAR JR, 2000; ZHANG; LU et al., 2002). Shape signatures are sensitive to noise, and slight changes in the boundary can cause large errors in matching. These functions are mostly translation, rotation and scale invariants and correspond to local shape representations.

Boundary moments can be used to reduce the dimensions of the boundary representation. Assuming the shape boundary has been represented as a shape signature $f(t)$, the rth moment m_r and central moment μ_r can be estimated as (SONKA; HLAVAC; BOYLE, 2014)

$$m_r = \frac{1}{N}\sum_{j=1}^{N}[f(j)]^r \text{ and} \qquad \mu_r = \frac{1}{N}\sum_{j=1}^{N}[f(j)-m_1]^r$$

where N is the number of boundary points. The normalized moments are invariant to euclidian transformations and corresponds with a global shape descriptors. However, it is difficult to associate higher order moments with physical interpretation.

Time-series models and especially Autoregressive (AR) modeling has been used for calculating shape descriptors (DUBOIS; GLANZ, 1986; DAS; PAULIK; LOH, 1990; SEKITA; KURITA; OTSU, 1992). Methods in this class are based on the stochastic modeling of a 1-D function f obtained from the shape signature. A linear autoregressive model expresses a value of a function as the linear combination of a certain number of preceding values. Specifically, each function value in the sequence has some correlation with previous function values and can therefore be predicted through a number of m observations of previous function values. The disadvantage of the AR method is that in the case of complex boundaries, a small number of AR parameters is not sufficient for an adequate description. The choice of m is a complicated problem and is usually decided empirically.

The scale space representation of a shape is created by tracking the position of inflection points in a shape boundary filtered by low-pass Gaussian filters of variable widths. As the width (σ) of Gaussian filter increases, insignificant inflections are eliminated from the boundary and the shape becomes smoother. The inflection points that remain present in the representation are expected to be significant object characteristics. The result of this smoothing process is an interval tree, called fingerprint, consisting of inflection points. This approach is called a Curvature Scale Space (CSS) contour image and matching proves to be very complex and expensive (MOKHTARIAN; MACKWORTH, 1986).

Spectral descriptors overcome the problem of noise sensitivity and boundary variations by analyzing shape in spectral domain. Spectral descriptors include Fourier Descriptor (FD) and Wavelet Descriptor (WD), they are derived from spectral transforms on 1-D shape signatures.

The Fourier coefficients are derived from Fourier reconstructed shape boundary rather than from original boundary. This is not different from FD derived from a smoothed boundary. FD and WD are backed by the well-developed and well-understood Fourier theory. The advantages of these over many other shape descriptors are (1) simple to compute; (2) each descriptor has specific physical meaning; (3) simple to do normalization, making shape matching a simple task; (4) captures both global (FD) and local (WD) features. With sufficient features for selection, they overcomes the weak discrimination ability of those simple global descriptors. Also overcomes the noise sensitivity (CHELLAPPA; BAGDAZIAN, 1984; VAN OTTERLOO, 1991; TIENG; BOLES, 1997; YANG; LEE; LEE, 1998).

Chain code describes an object by a sequence of unit-size line segments with a given orientation. In this approach, an arbitrary curve is represented by a sequence of small vectors of unit length and a limited set of possible directions, thus termed the unit-vector method. From a selected starting point, a chain code can be generated by using 4-directional or 8-directional chain code. Alternatively, the boundary can be represented by the differences in the successive directions in the chain code instead of representing the boundary by relative directions. This can be computed by subtracting each element of the chain code from the previous one and taking the result modulo n, where n is the connectivity. After these operations, a rotationally invariant chain code is obtained by a cyclic permutation which produces the smallest number. Such a normalized differential chain code is called the shape number. The chain code usually has high dimensions and is sensitive to noise. It is often used as an input to a higher level analysis. For example, it can be used for polygon approximation and for finding boundary curvature which is a important perceptual feature (COSTA; CESAR JR, 2000).

Another region-based shape representations as Hu's geometric moment (HU, 1962), algebraic moment (TAUBIN; COOPER, 1991), Zernike's orthogonal moment (TEAGUE, 1980), generic Fourier descriptor (ZHANG; LU, 2002) and grid based method (LU; SAJJANHAR, 1999) measures pixel distribution of the shape region, which are less likely affected by noise and variations. As a result, they usually can cope well with shape of significant detection. Particularly popular region methods are moment methods. The lower order moments or moment invariants carry physical meanings associated with region pixel distribution. However, it is difficult to associate higher order moments with physical interpretation. Grid methods are subject to noise due to the use of the major axis for normalization, and it is not rotation invariant for region-based shapes.

Region structural methods as convex hull (DAVIES, 2004; GONZALEZ; WOODS, 2008; SONKA; HLAVAC; BOYLE, 2014) and medial axis (COSTA; CESAR JR, 2000) have similar problems to contour structural approach. Apart from their complex computation and implementation, the graph matching is also an issue which needs to be solved itself.

2.1.3 Partial Shape Matching

After shape representation, a robust matching technique must be used for object recognition in order to obtain object hypotheses. For occluded object recognition task, a partial shape matching needs to be done.

Several algorithms have been considered in the past for the problem of shape matching. As in (ZHANG; LU, 2004), most of these can be broadly classified into two categories based on the types of features used; (1) methods that treat the shape as a blob in order to come up with an approximate representation; and (2) methods that use the contour boundary information directly.

One of the most popular blob-based approaches is to use a shock graph or a medial-axis transform for shape representation. Techniques that use these (SIDDIQI et al., 1999; SEBASTIAN; KLEIN; KIMIA, 2004; ALAJLAN; KAMEL; FREEMAN, 2008; MACRINI et al., 2011), first build a graph that models the skeleton of a shape. Topological similarity between the graphs helps in identifying the global shape structure, whereas geometric similarity at every node helps to capture the local shape information. These methods perform well in the presence of deformations. However, they build the skeleton a priori and can only match shapes when there is an overall global similarity between them and may fail in the presence of articulations, occlusions or noise in shape extraction.

To deal with such challenges some methods (FELZENSZWALB, 2005; BRONSTEIN et al., 2008) segment the shape into regions or parts that are then used for the task of matching. These methods capture the local shape variation much better by allowing articulations of such portions about each other. Felzenszwalb (FELZENSZWALB, 2005) proposed a technique for shape representation based on triangulated polygons and used Dynamic Programming to match such representations. However, this method can lead to errors in the presence of occlusions. To handle partial matching of shapes, Bronstein et al. (BRONSTEIN et al., 2008) proposed a pareto framework to determine an optimal tradeoff between part similarity and part decomposition. This method is computationally expensive due to working on shape blobs. Furthermore, the optimization method proposed to search over the parts is computationally very expensive and gives only an approximate solution.

While it may be claimed that blob-based methods are more robust due to the consideration of the entire 2D space, such methods tend to be computationally very expensive due to the processing of all the pixels enclosed by a shape. Thus, it is much more efficient to use the boundary information alone in order to match shapes.

Methods that use only contour boundary information for shape representation can further be classified into (1) global and (2) part-based methods.

Many global methods exist such as shape descriptors (LATECKI; LAKAMPER; ECKHARDT, 2000; BELONGIE; MALIK; PUZICHA, 2002; MORI; BELONGIE; MALIK, 2005), shape distances (HUTTENLOCHER; KLANDERMAN; RUCKLIDGE, 1993; LIU et al., 2010) and contour matching techniques (BASRI et al., 1998; LATECKI; LAKÄMPER, 2000; LATE-

CKI et al., 2007; DALIRI; TORRE, 2008; CAO et al., 2011), some of which also estimate the affine or projective transformation required to match the shapes (BELONGIE; MALIK; PUZICHA, 2002; BRYNER et al., 2014). In shape context the dissimilarity between two shapes is a weighted sum of the matching errors, computed using a maximum bipartite algorithms and a measure on the transformation required to match the two shape contours. The method works well to match shapes invariant to rigid transformations and deals with small deformations present in the contour boundary. One of the popular extensions of shape context (proposed by Mori et al. (MORI; BELONGIE; MALIK, 2005) solves the problem of shape matching very efficiently using multi-stage pruning techniques. The first stage is called representative shape contexts that matches very few shape contexts and identifies the outliers very fast, whereas the second stage matches the shapes in more details based on vector quantization in the space of shape contexts that involves clustering of the vectors, called as shapemes. All these methods rely on global features and hence fail in the presence of articulations, partial occlusions and noise present in the contour boundary.

To address such problems, Hong et al. (HONG et al., 2006) and Adamek and O'Connor (ADAMEK; O'CONNOR, 2004) represent shapes in terms of local features such as concave or convex portions of a contour to preserve the local geometry. Even though these more local methods are robust to some deformations, articulations and noise, they do not preserve sufficient contour information for a very discriminative matching. To model shapes better, the techniques mentioned in (FELZENSZWALB; SCHWARTZ, 2007) and (XU; LIU; TANG, 2009) combine local and global features. Felzenszwalb and Schwartz (FELZENSZWALB; SCHWARTZ, 2007) proposed a hierarchical matching technique for deformable shapes even in the presence of a cluttered background wherein a tree is built whose leaf nodes capture the local information and nodes close to the root capture global information. A Dynamic Programming based matching technique is used to match the two shape trees. On the other hand, Xu et al. (XU; LIU; TANG, 2009) proposed a Contour Flexibility descriptor that gives a deformation potential to each contour point so as to deal with deformations. The similarity between shapes is calculated by considering a linear combination of the local and the global measures. Although these methods use both local and global features and perform well against deformations, they do not consider partial matching of shapes and so may fail in the presence of occlusions.

In order to handle occlusions, Latecki et al. (LATECKI et al., 2007) developed an elastic partial shape matching algorithm to model a possible non-rigid shape deformation. In this approach, the problem is formulated as identifying a continuous segment from a target shape contour such that this identified contour segment best matches a query contour segment. This method identifies the outliers that may be present in the query shapes by allowing skips while matching. Non-rigid shape deformation is considered in the contour segment matching and a shortest-path-like algorithm developed to efficiently search for an optimal solution.

In Chen, Feris and Turk method (CHEN; FERIS; TURK, 2008), two shape contours are first sampled at a sequence of points and various features are extracted from each sampled

point. This way, the partial shape matching problem is reduced to the molecular subsequence matching problem studied in computational biology. The Smith-Waterman algorithm (SMITH; WATERMAN, 1981), which is widely used for molecular subsequence matching, is used to achieve efficient partial shape matching.

The approach proposed in (HORÁČEK; KAMENICKÝ; FLUSSER, 2008) divides an object into affine-invariant parts and uses modified radial vector for the description of parts. Object recognition is performed via string matching in the space of radial vectors. The part division is made by the zero curvature points and results in convex parts some of which are not present in any object. This leads to wrong hypotheses in the matching process.

Some works in the state of the art study the problem of establishing the best match between an open contour and a part of a closed one. For this purpose, they used in (MICHEL; OIKONOMIDIS; ARGYROS, 2011) a Dynamic Time Warping (DTW) matching technique with a scale descriptor. In (CUI et al., 2009), the same matching method was reused with just changing the representation mode to a new scale invariant signature. The authors in (SABER; XU; Murat Tekalp, 2005) proposed a sub-matrix matching to evaluate the partial similarity between two given shapes. These approaches do not support the impact of the difference between relative lengths of parts even in the case of elastic matching.

There have been numerous research efforts to deal with the local shape variations of an object shape and solve the problem of articulations. Cao (CAO et al., 2011) propose a new Markov Chain Monte Carlo (MCMC) based algorithm to handle partial shape matching with mildly non-rigid deformations. Represent each shape contour by a set of ordered landmark points and a selection of a subset of these landmark points is evaluated and updated into the shape matching by a posterior distribution, which is composed of a matching likelihood and a prior distribution. This method has the drawback that is computationally quite expensive due to the consideration of individual point-point matchings across the shapes.

Ma et al. (MA; LATECKI, 2011) proposed a technique for partial matching using geometric relations of shape context as shape descriptor followed by maximal clique inference based hypothesis used to identify the best possible part correspondences. Although, this method handles the problem of partial occlusions, it may fail in the presence of articulations as the local descriptors are not restricted to capturing information only within a part. Furthermore, the method is computationally very expensive due to the use of sampling methods.

Another popular approach for handling articulations is the Inner Distance Shape Context (IDSC) proposed by Ling and Jacobs (LING; JACOBS, 2007) that solves the problem of articulation in certain scenarios and can be considered as an improvement over Shape Context (BELONGIE; MALIK; PUZICHA, 2002). This method builds a descriptor based on the relative spatial distribution of the contour points using the Inner Distance instead of the Euclidean distance, and the Inner Angle instead of the regular angle. The Inner Distance (ID) between a pair of contour points is defined as the length of the shortest path between them while totally remaining within the shape and the Inner Angle is the angle from one point to the other that is in

the direction of this shortest inner path. A Dynamic Programming-based algorithm instead of a bipartite matching approach was also introduced by taking advantage of the ordering constraint in order to solve the point correspondence problem. This method is invariant to the 2D- articulations of a shape as it captures the part structure effectively. However, it is not invariant to affine changes of individual parts and also fails under partial occlusions as all the contour points are considered while building the descriptor and while matching.

In order to handle local affine changes, Gopalan et al. (GOPALAN; TURAGA; CHELLAPPA, 2010) proposed a shape-decomposition technique that divides a shape into convex parts using Normalized Cuts. These parts are then individually affine normalized and combined into a single shape that is matched using IDSC. As a result, this method is able to capture more deformations of local portions, such as a 3D part articulation that may be modeled by a 2D affine transformation of its projection. This yields a significant improvement over IDSC in many cases. It nevertheless assumes an a priori shape decomposition from a single shape that may be inconsistent in the presence of occlusions or noise in shape extraction. Furthermore, the matching is still global and hence one will be unable to handle partial occlusions of the shapes.

MERHY et al. (2014) presents a planar curve matching framework based on computing similarities between shape parts. They propose an elastic similarity measure issued from shape geodesics in the shape space. As the partial matching leads to additional difficulties, a shape decomposition process based on the Discrete Curve Evolution (DCE) is made. This approach is good for the partial shape task matching but an extension for occluded objects classification has not be proposed.

ZHANG; XU; LIU (2015) propose a new method for the recognition of partially occluded objects with affine distortion. The objects are divided into affine-invariant parts through a set of feature points. The matching scheme is based on the match of the feature points in the model and the query, and the hypotheses for each part are obtained calculating the extremum for a Hausdorff distance based metric. This new partial shape matching algorithm is capable of recognizing occluded non-deformable objects with affine transformation but is very sensitive to noise.

In MARVANIYA; GUPTA; MITTAL (2015) a new method for matching deformable objects is proposed. This method uses the High Curvature points and Opposite points for the division of the contour in segments and the division in parts is made grouping some segments until its proposed complexity metric is in a certain range. For the matching of the parts between the query and a target is used Dynamic Programming for the minimization of a defined energy function. The method is tested with some partially occluded objects but the results are given in terms of the retrieval problem, this means the k most probable objects in the occlusion without an object recognition.

2.1.4 Occlusion

Occlusions are common in real world scenes and are a major obstacle to robust object recognition. For feature-rich objects, discriminative keypoint features can be used to match

unique local patterns on the object even under severe occlusions. For feature-poor objects, however, occlusions further increase the shape ambiguity. While many shape matching approaches work really well when objects are entirely visible, their performance decrease rapidly with occlusions. When objects are under heavy occlusions, the score of false positives begin to overwhelm the scores of true detections, resulting in the inability to recognize objects robustly in these scenarios.

Occlusions are commonly modeled as regions that are inconsistent with object statistics. Girshick et al. (GIRSHICK; FELZENSZWALB; MCALLESTER, 2011) use an occluder part in their grammar model when all parts cannot be placed. Wang et al. (WANG; HAN; YAN, 2009) use the scores of individual HOG filter cells, while Meger et al. (MEGER et al., 2011) use depth inconsistency from 3D sensor data to classify occlusions. Local coherency of occlusions are often enforced with a Markov Random Field (FRANSENS; STRECHA; VAN GOOL, 2006) to reduce noise in these classifications. Li et al. (LI; GU; KANADE, 2011) use Random Sample Consensus (RANSAC) to generate a large set of hypotheses and hallucinate points at positions where there is high error. These approaches, however, assume that occlusions can happen randomly on an object. While this is true for some cases, in real world environments, objects are usually occluded by other objects resting on the same surface. It is thus often more likely for the bottom of an object to be occluded than the top of an object (DOLLAR et al., 2012).

Recently, researchers have attempted to learn the structure of occlusions from data (GAO; PACKER; KOLLER, 2011; KWAK et al., 2011). With enough data, these methods can learn an accurate model of occlusions. However, obtaining a broad sampling of occluder objects is usually difficult, resulting in biases to the occlusions of a particular dataset. In general occlusion reasoning has primarily been used to separate regions which belong to the object from those that do not. This allows the detector to ignore occluded regions which would otherwise corrupt the overall score.

2.2 Background

In this section are presented background methods further used in development of this work. Segmentation, principals shape representations, integral transformations and Hidden Markov models are explained in details in subsequent subsections.

2.2.1 Segmentation

Image segmentation is a critical component in many machine vision and information retrieval systems. It is typically used to partition images into regions that are in some sense homogeneous, or have some semantic significance, thus providing subsequent processing stages high-level information about scene structure. Recently, there has been an increasing interest in graph based segmentation algorithms (BOYKOV; FUNKA-LEA, 2006), (VU; MANJUNATH, 2008). These variational segmentation methods addresses the challenge of separating object

from background in a colored image, given certain constraints. In interactive segmentation the user is requested to supply the high-level information needed to detect and extract semantic objects through a series of interactions. Typically, operators mark areas of the image as object or background, and the algorithm updates the segmentation using the new information. These constraints are used as initial solution to the problem, leading to an iterative method which in conclusion aims to assign each pixel in the image its label (background or foreground).

Graph based segmentation algorithms starts with defining an optimization problem (energy cost function) which can be solved by creating a specific graph model (vertices and edges with weights) and running the graph cut algorithm with it as input.

The most basic object segmentation energy (UNGER et al., 2008) combines boundary regularization with log-likelihood ratios for fixed foreground and background appearance models, e.g. color probability densities functions, Θ^1 and Θ^0 and considers appearance models is not known a priori. Main objective of energy optimization is to find segmentation S of an image I, referring as I_p to pixel p in I and S_p segmentation result to pixel p in S (ROTHER; KOLMOGOROV; BLAKE, 2004).

$$E(S,\Theta^1,\Theta^0) = -\sum_{p:S_p=0} \ln P(I_p|\Theta^0) - \sum_{p:S_p=1} \ln P(I_p|\Theta^1) + |\partial S| \tag{2.1}$$

where regularization term is defined

$$|\partial S| = \sum_{\{p,q\}\in N} \omega_{pq}|S_p - S_q|$$

with similarity between pixels p and q

$$\omega_{pq} = \frac{1}{\|p-q\|} e^{\frac{-\Delta I^2}{2\sigma^2}} \tag{2.2}$$

The term $lnP(I_p|\Theta^i)$ in Equação 2.1 refers to probability of pixel p belongs to model Θ^i with $i = 0$ for background and $i = 1$ for foreground. $S_p \in \{0,1\}$ are binary indicator variables for membership of pixel p to background $(S_p = 0)$ or foreground $(S_p = 1)$ and $S = \{S_p|p \in \Omega\}$ where Ω refers to image domain (every image pixel) .

Given a functional like Equação 2.1 the objective is determine the unknown parameters S, Θ^0 and Θ^1 such that minimize the above energy where variable S represents all possible segmentations of input image.

$$(S,\Theta^0,\Theta^1)^* = \arg\min_{S,\Theta^0,\Theta^1} E(S,\Theta^0,\Theta^1) \tag{2.3}$$

This energy function holds that the evaluation of the best segmentation S^* for optimal models Θ^0 and Θ^1 is a global minimum. Given that this is NP-hard problem (VICENTE; KOLMOGOROV; ROTHER, 2009), Tang et. al (TANG et al., 2013) propose equivalent energy

representation by non-parametric color histograms.

$$E(S) = |S|H(\Theta^1) + |\bar{S}|H(\Theta^0) + |\partial S| \tag{2.4}$$

where $H(\Theta^x) = \sum_i p_i^x \cdot \ln p_i$ is the cross entropy of model Θ^x; p_i is probability of pixel i occurs in S and p_i^x is probability of pixel i occurs in model Θ^x.

In order to solve this theoretic optimization problem a graph cut approach is used. First, graph creation is made by taking N vertices (N=number of image pixels) and linking neighboring vertices with edges whose weights are defined by the vertices pixels color similarity defined in Equação 2.2. Next are added two more vertices to the graph representing foreground and background labels, each of which linked to all of the N pixels with an edge whose weight is defined by the probability of the pixel to match a color distribution of the background $D(0) = |\bar{S}|H(\Theta^0)$ or foreground $D(1) = |S|H(\Theta^1)$ (Equação 2.4). Figura 2.2 shows graph creation.

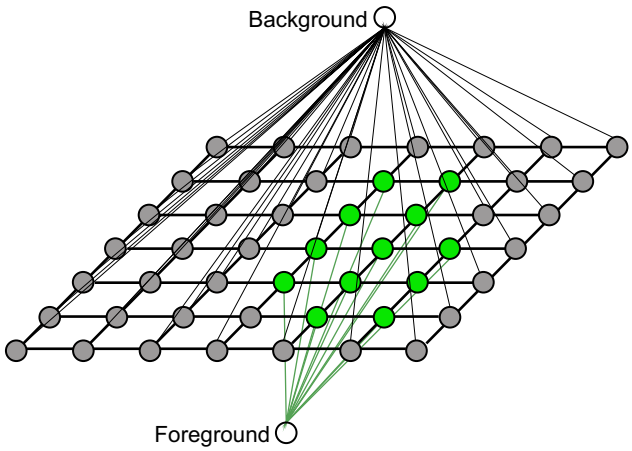


Figure 2.2: Graph construction for OneCut segmentation.

After graph creation is desired find a cut (a graph vertices partition into two disjoint subsets that are joined by at least one edge) that separate source(foreground) and sink(background) nodes such that minimize the connected components volume. Reminding subgraph V_{SG} volume is defined as

$$Vol(V_{SG}) = \sum_{i \in V_{SG}} \sum_{j \in V} E_{ij}$$

been $V_{SG} \subseteq V$ with V graph vertices set and E_{ij} the edge weight between node i and j.

Graph cut leads to partition vertices associated with images pixels into foreground and background subsets which is segmentation objective. Boykov-Kolmogorov algorithm (BOYKOV; JOLLY, 2001) is an efficient way to compute graph minimum cut and available source code from the authors was used in this work.

An interactive segmentation method based on graph cut named One Cut (TANG et al., 2013) was employed on the real image dataset studied in the experiments. The corresponding

energy is defined in Equação 2.4 and the scribbles given by the users serves to define weights for $D(0)$ and $D(1)$ in corresponding scribbled pixels. Weights for foreground scribbled pixels are $D(1) = \infty$ and $D(0) = 0$ meanwhile background scribbled pixels weights are $D(1) = 0$ and $D(0) = \infty$. The rest of the weights are determined according Equação 2.4 as explained before.

Examples of segmentation is shown in Figura 2.3 with obtained foreground and background shown in blue and red respectively (Figura 2.3(b)). The resulting contour (show in green) corresponds to the occlusion contour analyzed later by the proposed method in order to perform occluded object recognition.

(a) (b)

Figure 2.3: (a) Original images; and (b) segmentation results (foreground in blue).

2.2.2 Shape Representation

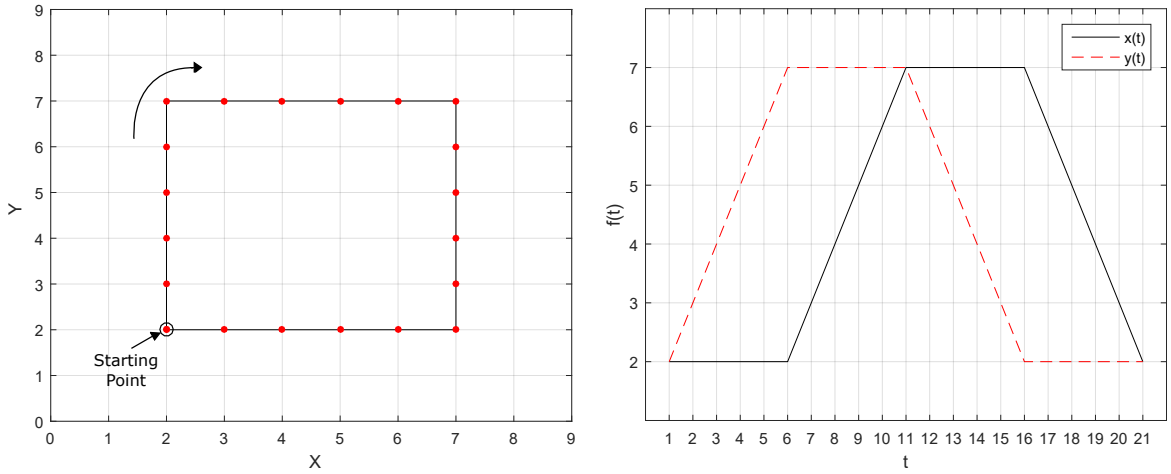
According to (ZHANG; LU, 2004), a taxonomy of computational shape representation can be divided into contour-based and region-based. Contour-based approaches are more popular than region-based approaches in the literature. This is because human beings are thought of to discriminate shapes mainly by their contour features. Another reason is because in many of the shape applications, like occluded object recognition, the shape contour is the only interest, whilst the shape interior content is not important.

It is important to note that contours of spatially quantized shapes can be represented directly in a binary image. Nevertheless, representing shapes directly by binary images implies some drawbacks (e.g. demands large storage space and does not explicitly identify the shape elements) that motivate the creation of alternative schemes. An important way of representing a contour is the so called parametrized or parametric representation, which is analogous to parametric representation of curves in differential geometry (COSTA; CESAR JR, 2000).

In order to better understand the concept of parametric contours, refer to Figura 2.4, which presents the contour of a square shape in a hypothetical simple image. The parametric representation of this contour is obtained by initially defining an arbitrary starting point and traversing the contour from this point onwards. The contour can be traversed clockwise or counterclockwise. If the curve is open, the contour is traversed from one extremity to the other, and if it is closed, the contour is traversed until the initial point is revisited. Let the lower-left corner of the contour of Figura 2.4(a) be the starting point and suppose that the contour is traversed in counterclockwise fashion. Therefore, the parametric representation of the contour begins with the initial point $(2,2)$ (i.e., $x = 2$, $y = 2$), followed by the next point in the counterclockwise direction $(2,3)$, which is followed by $(2,4)$, and so on. The obtained complete ordered set of points is $(2,2)$, $(2,3)$, $(2,4)$, $(2,5)$, $(2,6)$, $(2,7)$, $(3,7)$, $(4,7)$, $(5,7)$, $(6,7)$, $(7,7)$, $(7,6)$, $(7,5)$, $(7,4)$, $(7,3)$, $(7,2)$, $(6,2)$, $(5,2)$, $(4,2)$, $(3,2)$.

It is observed that contour in the above example can also be parametrically represented in terms of two functions $x(t)$ and $y(t)$ where the parameter t corresponds to the order of contour elements along the shape contour (Figura 2.4(b)). Another way to see the parametrization process is as point movement along the contour with parameter t being the time in the described movement functions. This point of view makes the process to unfold in time and models such as Markov based can be used for object classification, (see Subseção 2.2.4).

By using a parametric curve, two usual types of representations are considered, global and local. A global shape descriptor takes into account the position of all other points relative to a reference point (e.g. shape context(BELONGIE; MALIK; PUZICHA, 2002) and inner-distance shape context (LING; JACOBS, 2007)). A local point shape descriptor depends only on its neighborhood and will not change as long as neighbor points do not change (e.g. tangent angle (ZAHN; ROSKIES, 1972), curvature (CUI et al., 2009) and radial vector (HORÁČEK; KAMENICKÝ; FLUSSER, 2008)). For occluded objects, local representations tend to be more



(a) (b)

Figure 2.4: (a) Contour of square shape and (b) parametric contour representation in clockwise manner.

suitable than global ones because changes in some points will not affect all descriptors. In the state of the art review, it was observed that most shape signature are local representations and present resistance to partial missing of contour points with very low computational complexity (YANG; KPALMA; RONSIN, 2008).

A shape signature is a one-dimensional function which is derived from shape boundary coordinates (ZHANG; LU et al., 2002). Usually captures the perceptual shape feature and can describe a shape by itself. Most commonly used shape signatures are complex coordinates, centroid distance function, tangent angle, curvature, area function, triangle-area representation and chord length function (ZHANG; LU, 2004). All these, except complex coordinates signature, are translation, rotation and scale invariant. Of these, only tangent angle and curvature signatures present good resistance to occlusion and non-rigid deformation with low computational complexity. In this section, these two shape signatures are introduced.

2.2.2.1 Tangent angle signature

The tangent angle function (TAS) of parameterized curve at point t is defined by a tangential direction of a contour (ZHANG; LU et al., 2001), see Figura 2.5.

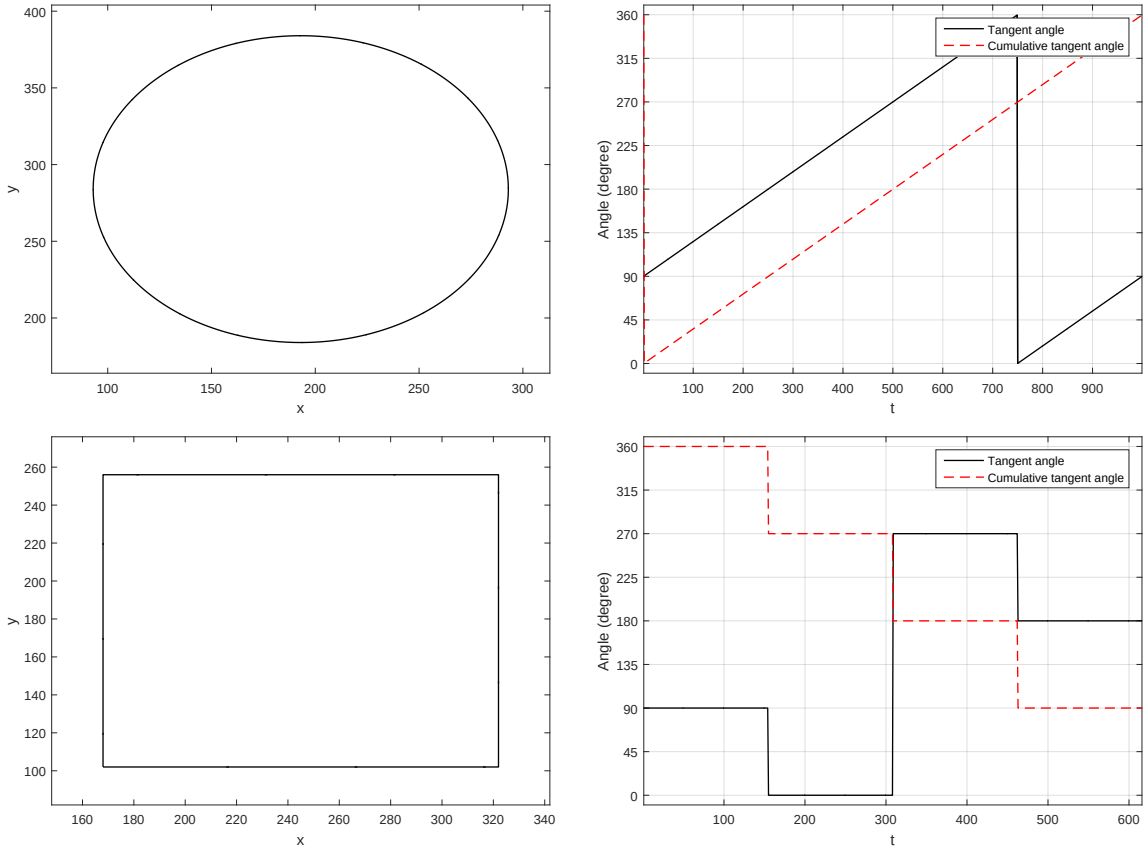
$$\theta(t) = \theta_t = \arctan \frac{y(t) - y(t-k)}{x(t) - x(t-k)} \tag{2.5}$$

since every contour is a digital curve; k is a small window to calculate θ_t more accurately.

One problem of this function is discontinuity, due to the fact that the tangent angle function assumes values in a range of length 2π, usually in the interval of $[\pi, \pi]$ or $[0, 2\pi]$. Therefore θ_t in general contains discontinuities of size 2π. To overcome the discontinuity problem, with an arbitrary starting point, the cumulative angular function φ_t is defined as the angle differences between the tangent at any point t along the curve and the tangent at the starting

point $t = 0$ (YANG; KPALMA; RONSIN, 2008) (see Figura 2.5):

$$\varphi(t) = \varphi_t = |\theta(t) - \theta(0)|$$



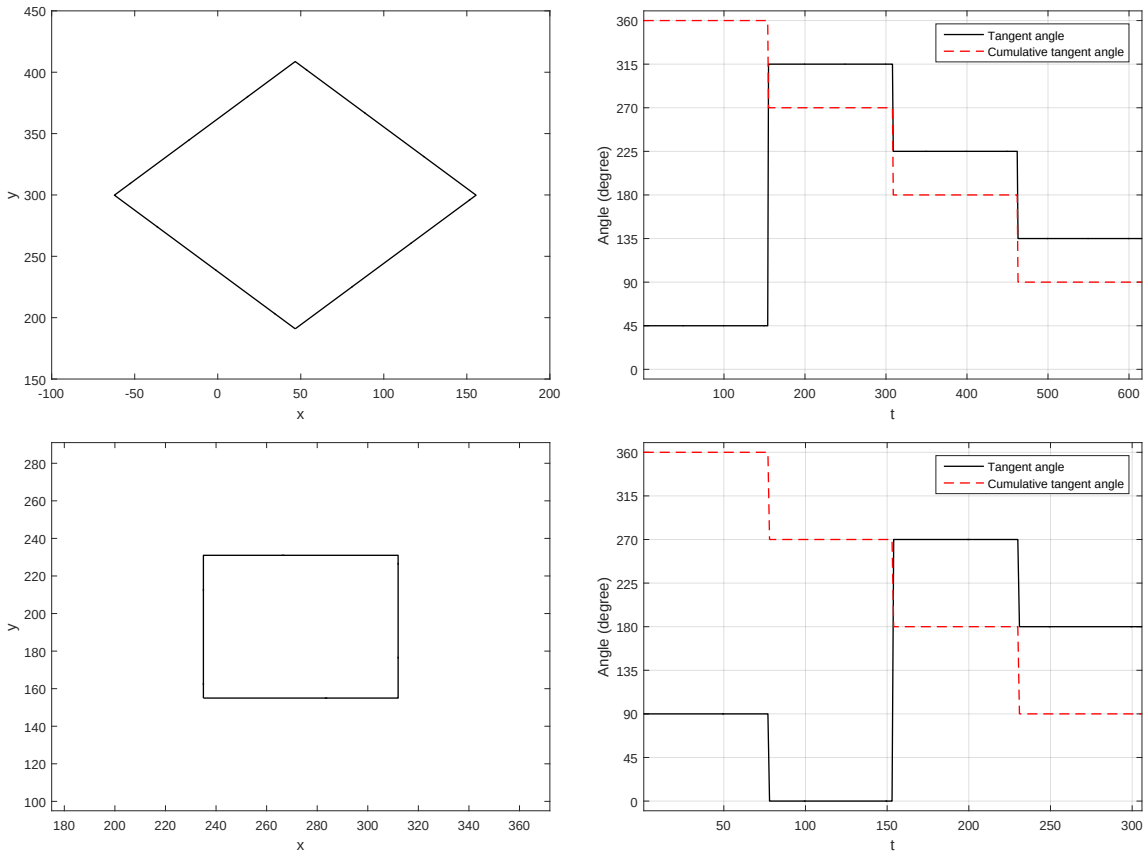
(a) (b)

Figure 2.5: (a)Simples shape and (b) respective tangent angle and cumulative tangent angle signatures.

As it can be seen in Figura 2.6 rotated and scale versions of shapes maintain the same form but different shift where rotated and different sampling in scaled version. In another hand cumulative tangent angle is same for original square shape and rotated version but invariance is restricted for correct initial point selection. This means that for occluded objects representation, multiple versions with different initial points normalization of training objects are needed.

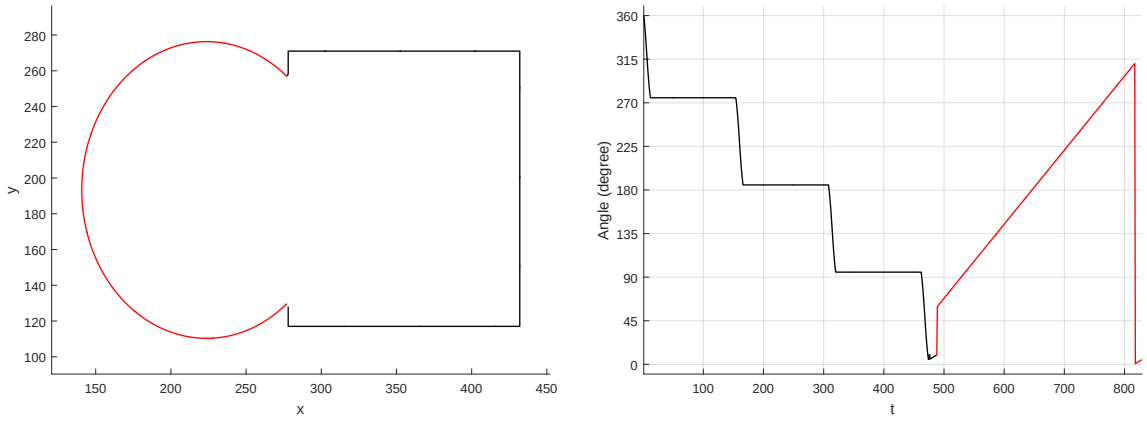
Robustness when shapes presents occlusion can be observed in Figura 2.7 where only the missing contour parts are lost in the signature meanwhile the rest remains invariant. Square shape signature of occlusion is represented in black and maintains step form as observed in the figure. Circle shape signature is also correctly represented and occlusion signature corresponds with a composite of these shapes tangent angle signatures.

The principal drawback of tangent angle signature is noise sensitivity. For solution of this issue is introduced Wavelet Transform (Subseção 2.2.3) in order to make a multi-resolution analysis.



(a) (b)

Figure 2.6: (a) Rotated and scaled shapes contour and (b) respective tangent angle signature.



(a) (b)

Figure 2.7: (a) Occluded shapes contour and (b) respective tangent angle signature.

2.2.2.2 Curvature signature

Curvature is a very important boundary feature for human to judge similarity between shapes. It also has salient perceptual characteristics and has proven to be very useful for shape recognition. In order to use it for shape representation, is quoted the function of curvature, $K(t)$, from (MOKHTARIAN; MACKWORTH, 1992) as:

$$K(t) = \frac{\dot{x}(t)\ddot{y}(t) - \dot{y}(t)\ddot{x}(t)}{(\dot{x}(t)^2 + \dot{y}(t)^2)^{3/2}} \tag{2.6}$$

Could be observed from this equation that estimating the curvature involves the derivatives of contour parametrization $x(t)$ and $y(t)$, which is a problem in the case of computational shape analysis where the contour is represented in digital form. In this notation $\dot{x}(t)$ represents first derivative of $x(t)$ and $\ddot{x}(t)$ second derivative. The simplest approach is to approximate the derivatives of $x(t)$ and of $y(t)$ in terms of finite differences, i.e.,

$$\begin{aligned}
\dot{x}(t) &= x(t) - x(t-1) \\
\dot{y}(t) &= y(t) - y(t-1) \\
\ddot{x}(t) &= \dot{x}(t) - \dot{x}(t-1) \\
\ddot{y}(t) &= \dot{y}(t) - \dot{y}(t-1)
\end{aligned}$$

The curvature can then be estimated by substituting the above-calculated values in the curvature Equação 2.6. Although this simple approach can be efficiently implemented, it is rather sensitive to noise. A more elaborate technique is based on approximating the contour in a piecewise fashion to parametric polynomial. The curve approximating contour point neighborhood could be estimated by least square error technique. Thus, parametric curve definition using second order polynomial is:

$$\begin{aligned}
x(t) &= a_1 t^2 + b_1 t + c_1 \\
y(t) &= a_2 t^2 + b_2 t + c_2
\end{aligned}$$

with first and second order derivative

$$\begin{aligned}
\dot{x}(t) &= 2a_1 t + b_1 \\
\ddot{x}(t) &= 2a_1 \\
\dot{y}(t) &= 2a_2 t + b_2 \\
\ddot{y}(t) &= 2a_2
\end{aligned}$$

Again curvature can be calculated by substituting in Equação 2.6, but this approach consist in several polynomial approximation and is too expensive.

Another more suitable approach is angle-based curvature estimation like the proposed in (GUERRERO-PENA et al., 2015). Curvature approximation is calculated through k-curvature calculation. For this is defined Equação 2.7 and Equação 2.8, which are the k-slope in both directions. Vertical direction slope m_y is considered to eliminates undefined K_x curvatures in

special cases when $x(t) = x(t+k)$. Equação 2.9 and Equação 2.10 define the k-curvatures in each direction as obtained slopes differences.

$$m_x(t,k) = \frac{y(t) - y(t+k)}{x(t) - x(t+k)} \tag{2.7}$$

$$m_y(t,k) = \frac{1}{m_x(t,k)} = \frac{x(t) - x(t+k)}{y(t) - y(t+k)} \tag{2.8}$$

$$K_x(t,k) = m_x(t,k) - m_x(t+k,k) \tag{2.9}$$

$$K_y(t,k) = m_y(t,k) - m_y(t+k,k) \tag{2.10}$$

where k is the analyzed k-neighbor of point t for the k-curvature.

Curvature is calculated as minimum absolute values of k-curvature in horizontal and vertical directions, see Figura 2.8(c) and Figura 2.8(d) respectively.

$$K(t,k) = \min(|K_x(t,k)|, |K_y(t,k)|) \tag{2.11}$$

The resulting signal reaches the highest values in the zones where the curvature is also high in the x and y directions, and the values tend to decrease in the remaining zones. The resulting signal is shown in Figura 2.8(b). Local maximum in resulting $K(t,k)$ are associated to high curvature points.

Two undefined points in curvature of circle part of the shape, can be seen in Figura 2.8(c) and Figura 2.8(d) respectively (show in red). These points appears as local minimum or maximum but are not associated with high curvature points given that circled shape objects has constant curvature. Information of two directional curvature allows to remove these undefined points as can be observed in Figura 2.8(b).

Figura 2.9 shows curvature signatures of objects calculated by polygonal approximation and K-curvature. Both calculations methods has similar results but K-curvature shows significant time improvement respect the first.

As it can be seen in Figura 2.10 rotated and scale versions of shapes maintain the same curvature representation with different sampling in scaled version. Also robustness when shapes presents occlusion can be observed in Figura 2.8(b) as in tangent angle case.

Main difference between tangent angle and curvature signatures could be observed from Figura 2.5 and Figura 2.9 where very different signals were obtained to square and circle shapes for tangent angle meanwhile similar constant signatures are calculated in curvature cases, with only difference in four high curvature points of square shape.

2.2.3 Integral Transform

In shape representation review, it has been found that methods working in spatial domain suffer from two main drawback: noise sensitivity and high dimension. The problems can

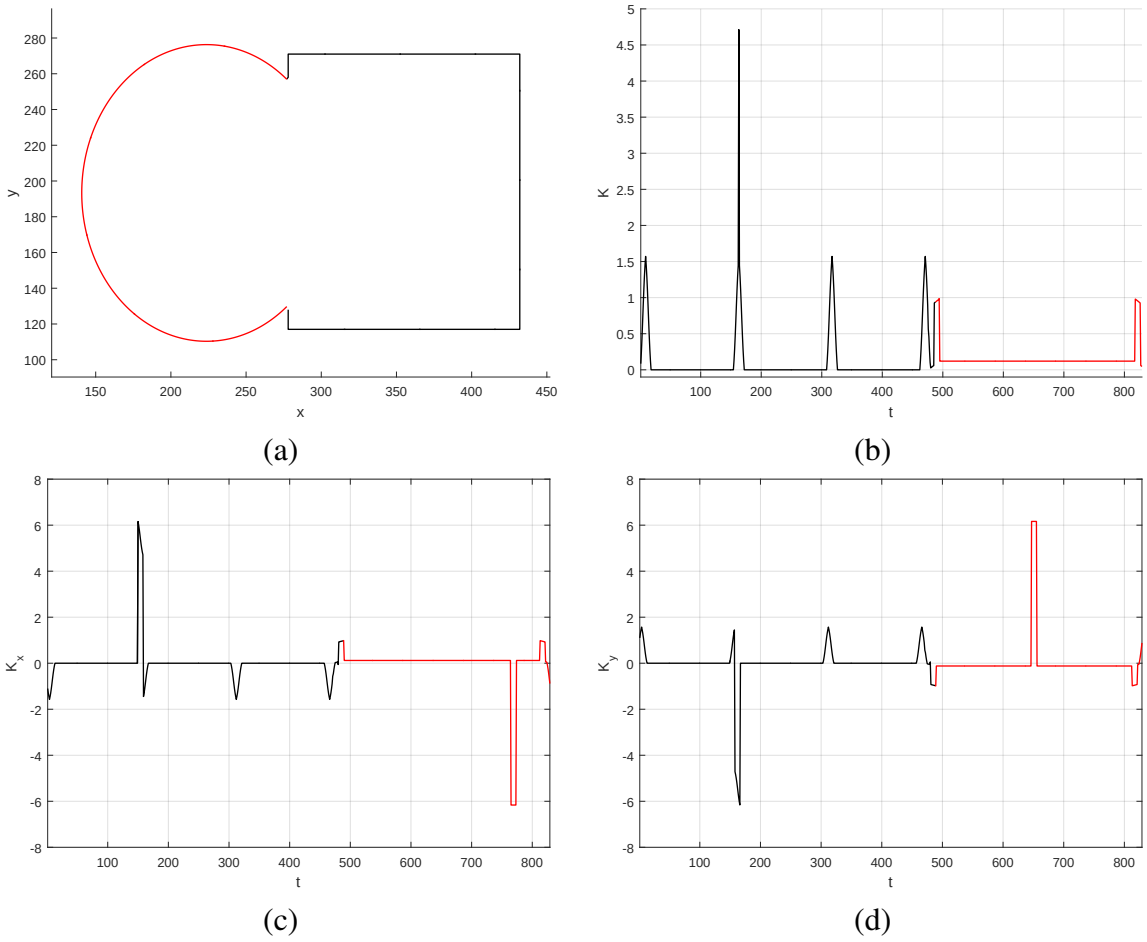


Figure 2.8: (a)Two occluded objects shape and its (b) K-curvature $K(t,k)$ calculated through (c) horizontal K-curvature $K_x(t,k)$ and (d) vertical K-curvature $K_y(t,k)$.

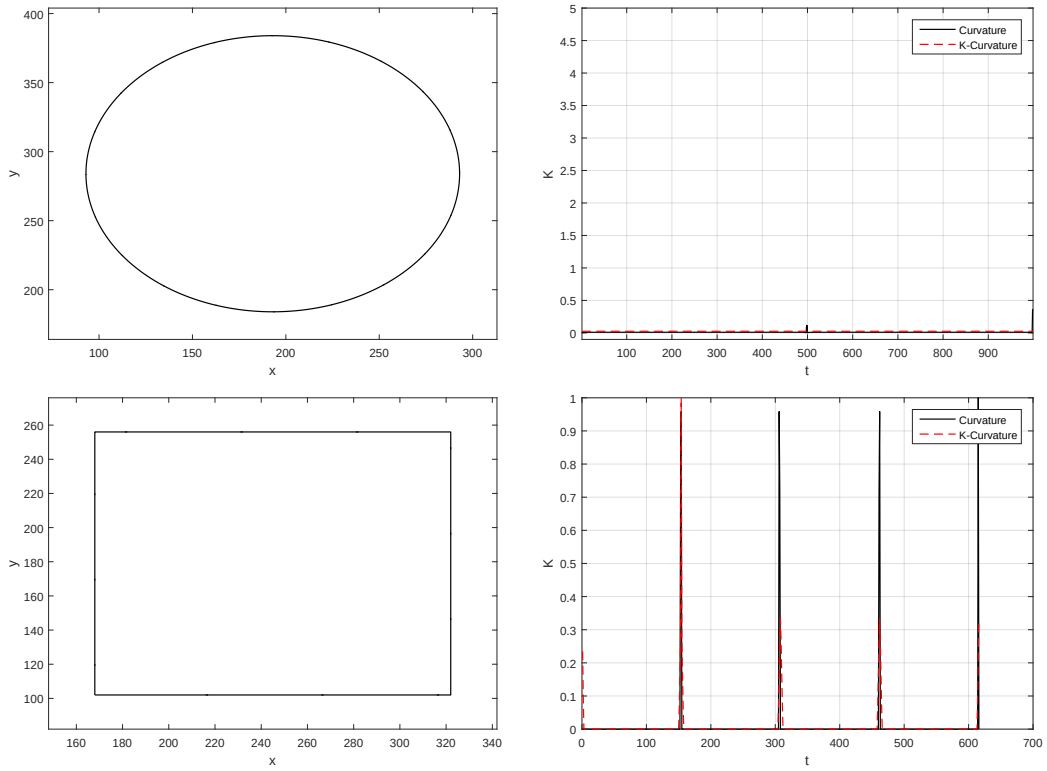
be solved in four ways: histogram, moments, scale space and spectral transforms (ZHANG; LU, 2004). Although histogram and scale space increase robustness to noise and compactness, matching using histogram and scale space can be very expensive. Moments is robust and compact, however, higher order moments are either difficult to obtain or without physical meaning. Among the four solutions, spectral transforms is the most promising.

The existence of problems that are difficult to solve in their original representations or domains is the motivation behind de integral transforms. An integral transform map an equation from its original domain into another more adequate domain, where the solution it is simpler. The solution then is mapped back to original domain with inverse transform like is show in Figura 2.11.

In general an integral transform $F(s)$ of function $f(t)$ is defined as:

$$F(s) = \int_a^b \kappa(s,t) f(t) dt = T[f(t)] \tag{2.12}$$

where $\kappa(s,t)$ is a known function named transformation kernel. If a and b are finite, transformation is called finite otherwise infinite. Depending on kernel selection and integration



(a) (b)

Figure 2.9: (a)Simples shape and (b) respective curvature signatures.

limits are obtained different integral transforms.

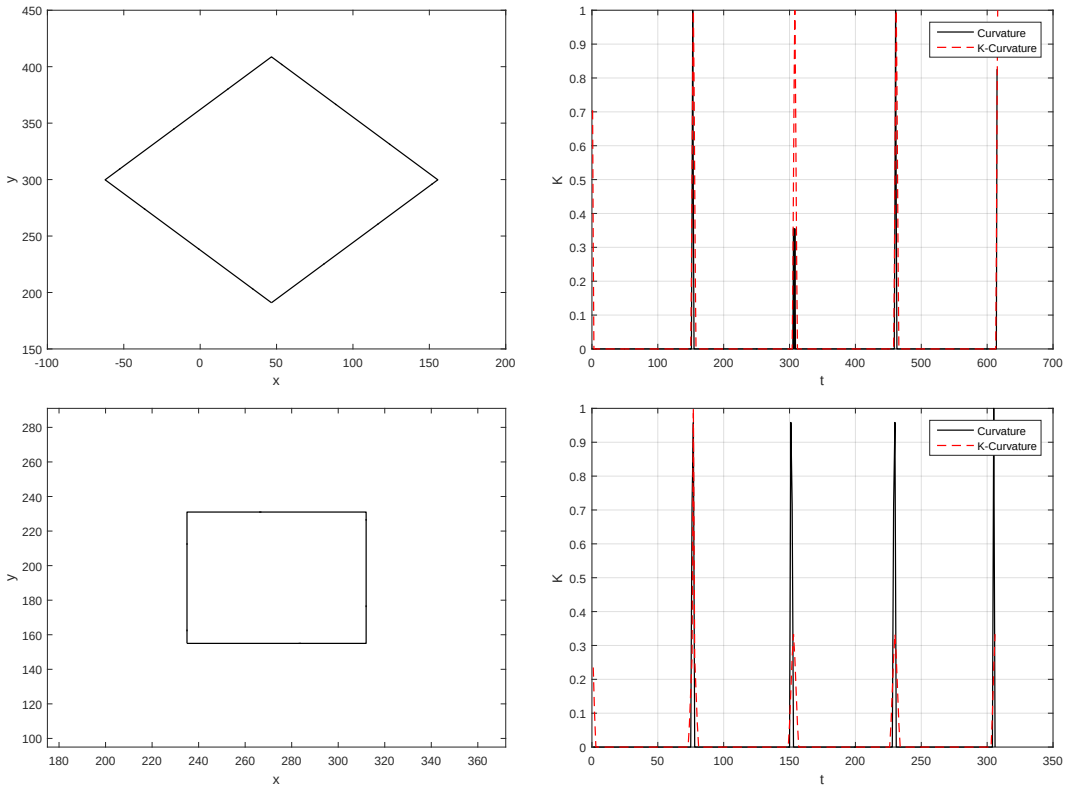
2.2.3.1 Fourier Transform

Fourier Transform (FT) is the most popular and worldwide used integral transform that map temporal/spacial information to frequency domain. Part of the importance of the Fourier approach arises from the fact that it allows a representation of a broad class of functions in terms of a linear combination of sine, cosine or complex exponential basic functions. Moreover, unlike most alternative representations of functions in terms of an orthogonal kernel, the Fourier approach exhibits an inherent and special compatibility with the signals typically found in nature, especially regarding their oscillatory and highly correlated features. In addition, much insight about other important transforms, such as wavelet transform, can be gained by treating them in terms of the Fourier transform.

Signal representation in frequency domain is unique and consist in sinusoids infinite sum been one dimensional Fourier transform defined by equation (COSTA; CESAR JR, 2000):

$$\mathfrak{F}(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-iwt} dt \tag{2.13}$$

where the variables t and w are usually called time and frequency, respectively. Transformed value $\mathfrak{F}(w)$ from Equação 2.13 represents a mean value over all time/space of energy



(a) (b)

Figure 2.10: (a) Rotated and scaled shapes contour and (b) respective curvature signature.

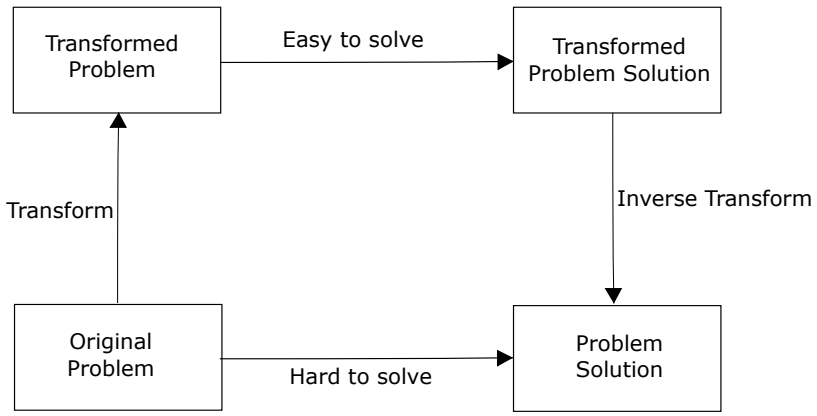


Figure 2.11: Integral transform.

content in certain frequency w named Fourier coefficients. Then, spectral composition of the signal is represented in terms of the power spectrum $P_s f$ of the original signal $f(t)$, which is defined as

$$P_s f = |\mathfrak{F}(w)|^2 = \mathfrak{F}(w) \cdot \mathfrak{F}(w) \tag{2.14}$$

An important property of the power spectrum is that it does not change as the original function is shifted along its domain, which is explored by the so called Fourier descriptors for shape analysis. Although the Fourier transform of complex function is usually a complex

function, it can also be a purely real (or imaginary) function. On the other hand, the power spectrum is always a real function of the frequency. Some example of signals power spectrum are shown in Figura 2.12(b) and Figura 2.13(b).

High frequencies content in signals are associated with very quick variations and anomalous behavior in time domain. These variations generally are noise randomly added to the signal and its identification and removal in frequency space result an easy task. Example of noisy signal and filtered signal by high frequencies elimination (low pass filtering) is shown in Figura 2.12.

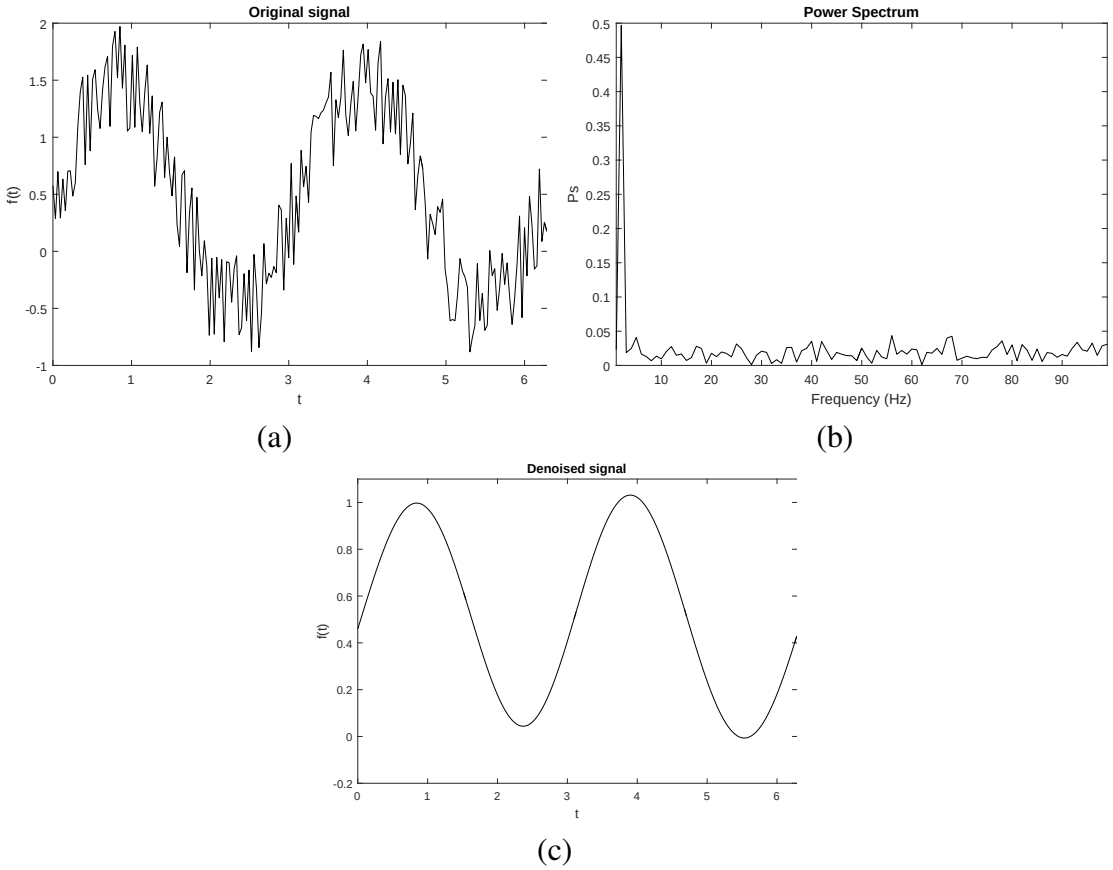


Figure 2.12: (a) Signal $\sin(2x)$ with uniform noise, (b) power spectrum of (a), and (c) low pass filtered signal.

Signals that it frequency content not vary in time are called stationary signals (see Figura 2.12(c) and Fourier analysis result very useful in this cases. On the other hand most of signals, like previously discussed shape signatures, are not stationary and power spectrum in Fourier domain only reveals what is the signal frequency content but time interval where each frequency is contained is not know (Figura 2.13(c)). This means Fourier analysis is not the right tool for non stationary signals analysis and filtering because frequency variations in time are not detected. Additionally, Fourier coefficients are calculated using whole domain because the infinite nature of its kernel (Equação 2.13) and local variations or missing parts of original signal, like in occluded shape signatures, result in modification of all coefficients, including unaltered parts.

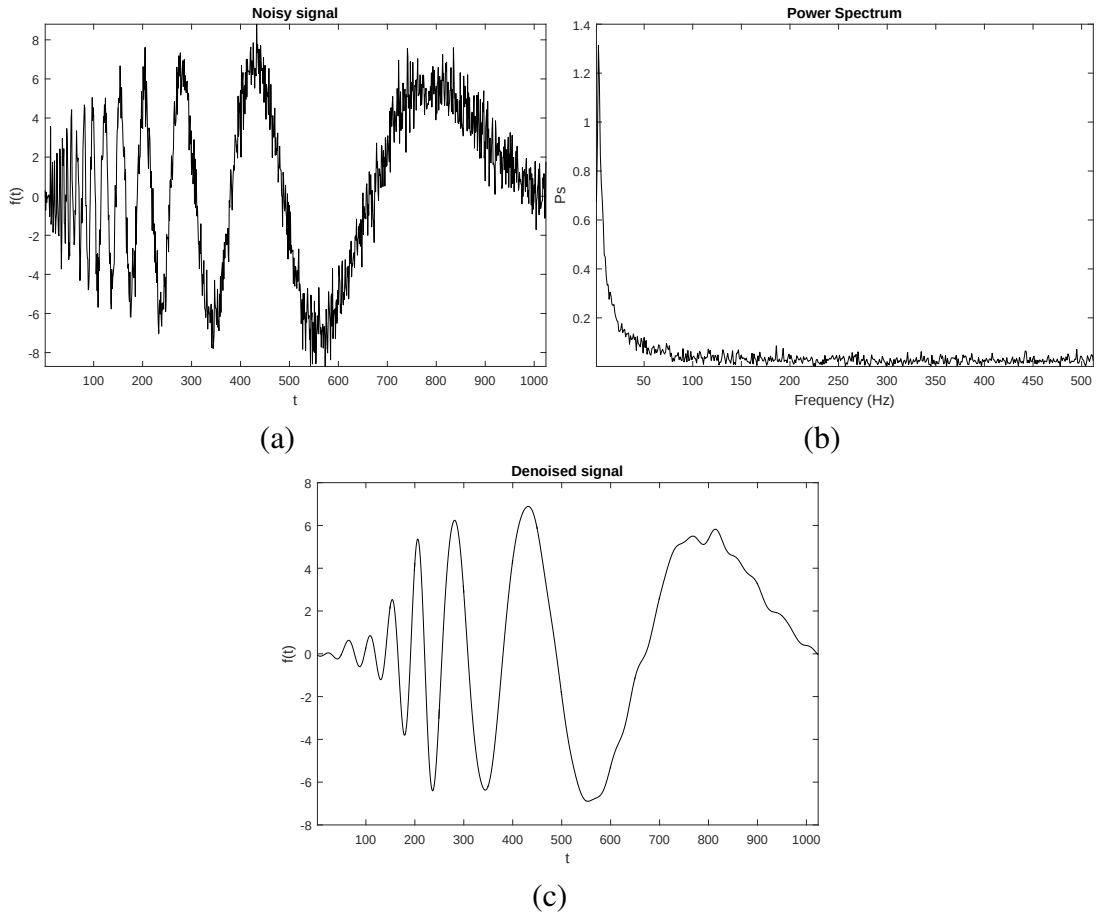


Figure 2.13: (a) Non stationary signal with gaussian noise, (b) power spectrum of (a), and (c) low pass filtered signal.

Further approach of Fourier analysis try to solve this problem dividing $f(t)$ into parts using window function $v(t)$ before transformation. This approach named Short Time Fourier Transform (STFT) use typical Fourier transform over fixed time interval and centering the window in time instant τ. The main disadvantage of Short Time Fourier is that once windows size is fixed is used for whole signal analysis being the method very sensitive to windows size choice.

2.2.3.2 Wavelet Transform

Wavelet Transform (WT) is efficient for analysis of non-stationary local signals and, equals to STFT, maps signals in time-scale representation. The difference is wavelet provides multiresolution analysis with different windows sizes. High frequency analysis is made using narrow window and low frequencies with widest windows. The larger scales in the wavelet space are associated with low frequencies and its used to eliminates the influence of the noise.

To obtain WT its required a real predefined function $\psi(t)$ named mother wavelet (Figura 2.14(a)) that satisfies some mathematics criteria: finite energy ($E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty$), admissibility ($\mathfrak{F}_{\psi}(0) = 0$), progressive ($\psi(f) = 0, f \leq 0$) and a number of vanishing moments

($\int_{-\infty}^{\infty} t^r \psi(t)dt = 0, r = 0,1,...K$) (COSTA; CESAR JR, 2000). This function is manipulated through a process of translation (i.e. movement along time axis) and dilation (i.e. spreading out of the wavelet) to transform original signal into another form which unfolds it in time (Figura 2.14(b)) and scale (Figura 2.14(c)). There are a large number of mother wavelets to choose and the best one depends on both the nature of the signal and what its required for the analysis.

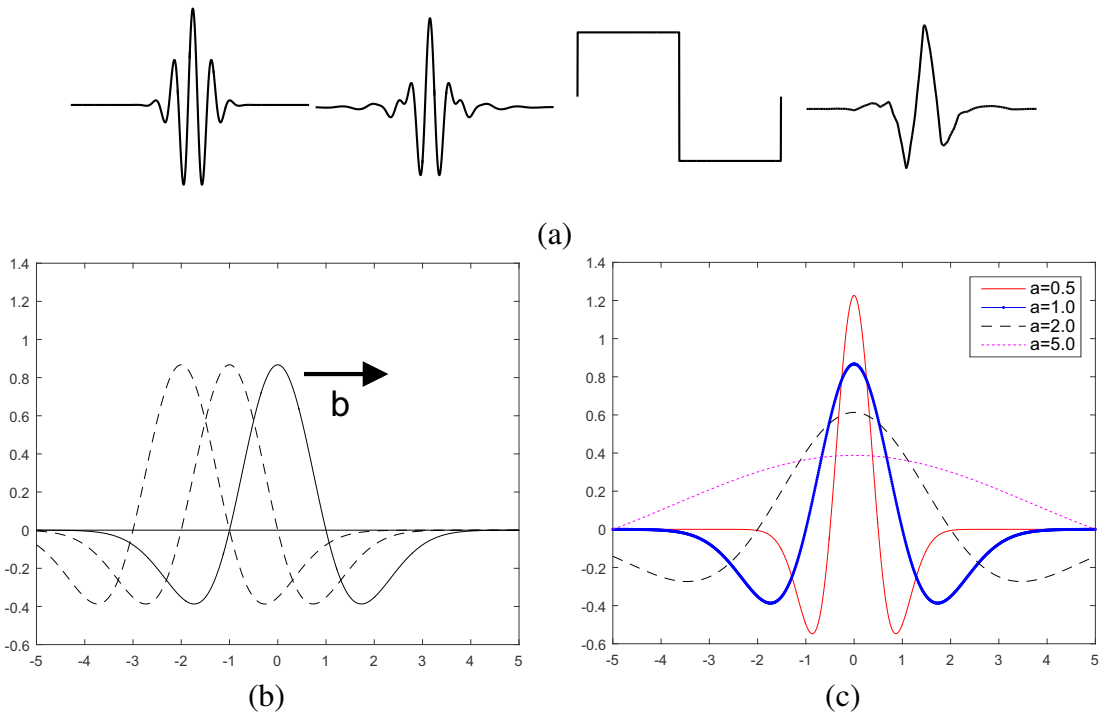


Figure 2.14: Examples of (a) mother wavelets; (b) shifted versions and (c) scaled versions.

Second derivative of gaussian function satisfy mathematics criteria of wavelets and it is the basic form of wavelet know as Mexican hat.

$$\psi(t) = \frac{2}{\pi^{\frac{1}{4}}\sqrt{3\sigma}}\left(1 - \frac{t^2}{\sigma^2}\right)e^{-\frac{t^2}{2\sigma^2}} \tag{2.15}$$

This wavelet is very useful for detection of local maximum and has a poor frequency resolution, leading to high frequencies discriminations which usually are associated with the contour noise. The parameter σ were set to 1 for balance between noise reduction and significant feature extraction.

2.2.3.3 Wavelet Coefficient

In order to generate dilated and translated versions of original mother wavelet $\psi(t)$ dilation parameter a and location parameter b are included in wavelet definition Equação 3.1. The shifted and dilated versions of the mother wavelet are denoted $\psi(\frac{t-b}{a})$ leading to definition of wavelet family:

$$\psi\left(\frac{t-b}{a}\right) = \left[1-\left(\frac{t-b}{a}\right)^2\right] e^{-\frac{1}{2}[(t-b)/a]^2} \tag{2.16}$$

A signal $f(t)$ could be transformed to wavelet spaces using integral transformation equation (Equação 2.12) and substituting kernel function by Equação 2.16.

$$T_\psi(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right) dt \tag{2.17}$$

This is one dimensional Continuous Wavelet Transform (CWT) and can be though as the cross-correlation of a signal with a wavelet family. Different to Fourier, wavelet kernels has finite energy and transformation of partially modified signals only results in associated wavelet coefficient modification. Signal spectral composition in wavelet space are calculated in the same way of Fourier (see Equação 2.14) obtaining a two dimensional signal, named scalogram, for one dimensional CWT (see Figura 2.15(b)). Remembering that lower scales are associated with signal noise, a low pass filtering can be done but selecting noisy frequencies for each time location b. Example of noisy signal with is corresponding scalogram and low pass filtering is show in Figura 2.15. Can be observed from Figura 2.15(c) that low pass filtering result in wavelet space has better signal reconstruction than Fourier filtering (Figura 2.13(c)).

Figura 2.16(a) and (b) shows original signal corresponding to function $\sin(2x)$ and locally modified signal from this function. Wavelet and Fourier coefficient were calculated using Equação 2.17 and Equação 2.13. Varied signal Fourier coefficients differs to original signal FT and, as observed in Figura 2.16(c), invariant part can not be recognized. Contrary to this, Wavelet coefficients for the two signals remains equals for signal unmodified part Figura 2.16(d).

2.2.4 Hidden Markov Model

In problems that have an inherent temporality, that is, consist of a process that unfolds in time, it may have states at time t that are influenced directly by a state at $t-1$. Hidden Markov models (HMM) have found greatest use in such problems, for instance speech recognition or gesture recognition. Hidden Markov models have a number of parameters, whose values are set to best explain training patterns for the known category. Later, a test pattern is classified by the model that has the highest posterior probability, i.e., that best explains the test pattern (DUDA; HART; STORK, 2012).

Considering a sequence of states at successive times; the state at any time t is denoted $\omega(t)$. A particular sequence of length T is denoted by $\omega = \{\omega(1), \omega(2), ..., \omega(T)\}$ as for instance we might have $\omega = \{\omega 1, \omega 4, \omega 2, \omega 2, \omega 1, \omega 4\}$. The system can revisit a state at different steps, and not every state need be visited.

Model for production of any sequence is described by transition probabilities $P(\omega_j(t)|\omega_i(t-1)) = a_{ij}$, the time-independent probability of having state ω_j at step t given that the state at time $t-1$ was ω_i. There is no requirement that the transition probabilities be symmetric ($a_{ij} \neq a_{ji}$, in

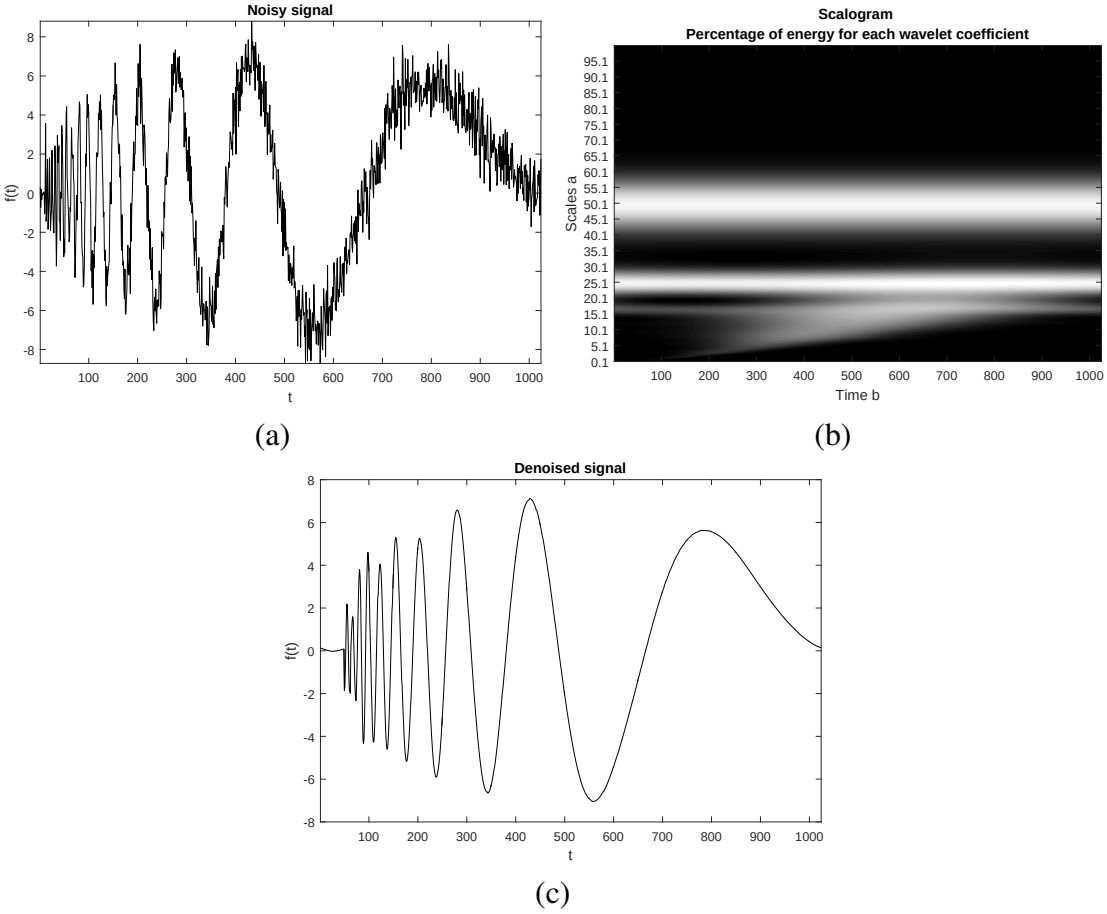


Figure 2.15: (a) Non stationary signal with gaussian noise, (b) scalogram of (a), and (c) low pass filtered signal.

general) and a particular state may be visited in succession ($a_{ii} \neq 0$, in general), as illustrated in Figura 2.17.

Given a particular model λ, that is, the full set of a_{ij}, as well as a particular sequence ω^T. In order to calculate the probability that the model generated the particular sequence its simply multiply the successive probabilities. For instance, to find the probability that a particular model generated the sequence described above, we would have $P(\omega|\lambda) = a_{14}a_{42}a_{22}a_{21}a_{14}$. Up to here its been discussed a first-order discrete time Markov model since the probability at t depends only on the states at $t-1$.

However in some applications the perceiver does not have access to the states $\omega(t)$. Instead, are measured some properties named observations. Thus it will have to allow the introduction in Markov model of visible states, which are directly accessible to external measurement, as separate from the ω states, which are not.

Assuming that at every time step t the system is in a state $\omega(t)$, now is also assumed that it emits some discrete visible symbol $o(t)$. As with the states, is defined a particular sequence of such visible states as O={o(1),o(2),...,o(T)} and thus it could be obtained $O = \{o_5, o_1, o_1, o_5, o_2, o_3\}$.

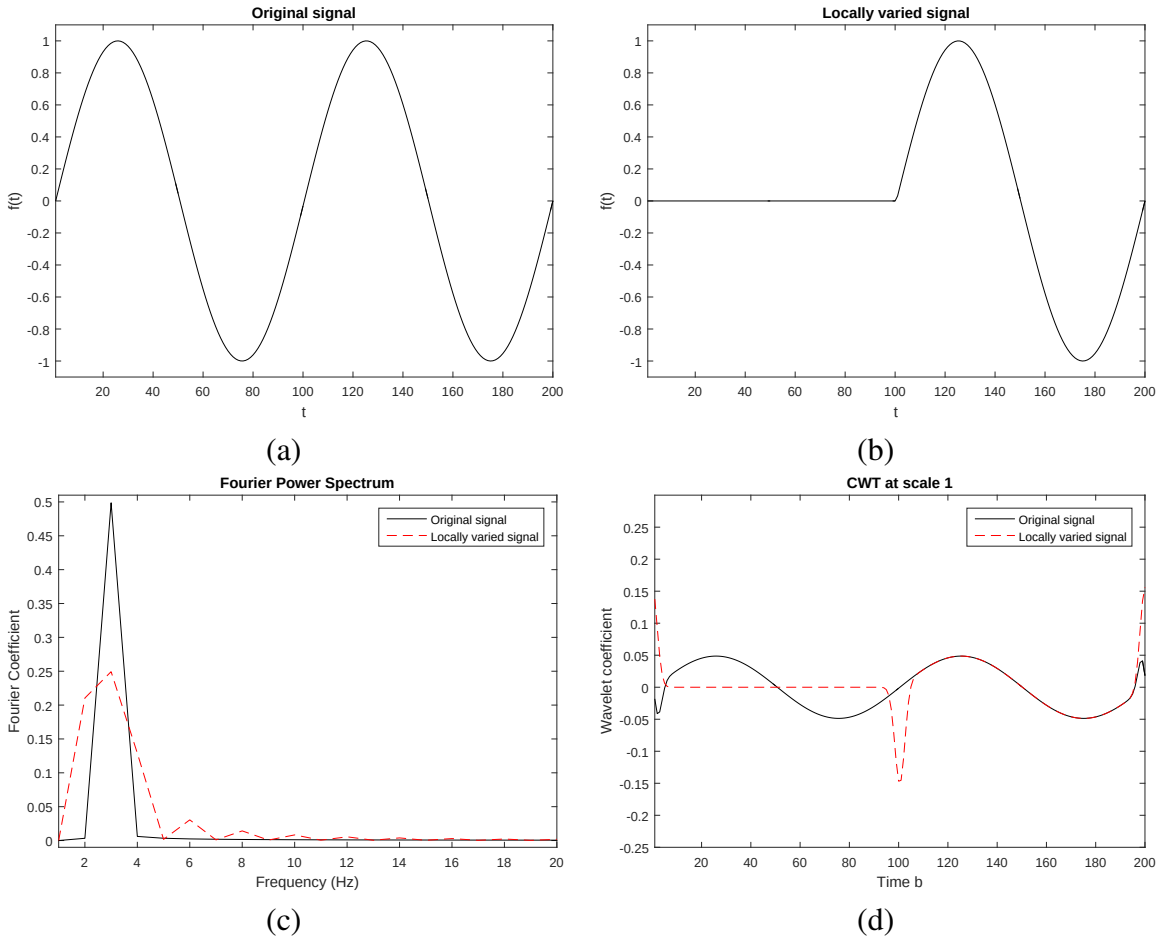


Figure 2.16: (a) $\sin(2x)$ signal, (b) locally varied $\sin(2x)$ signal, (c) Fourier coefficients for (a) and (b), and (d) Wavelet coefficients for (a) and (b).

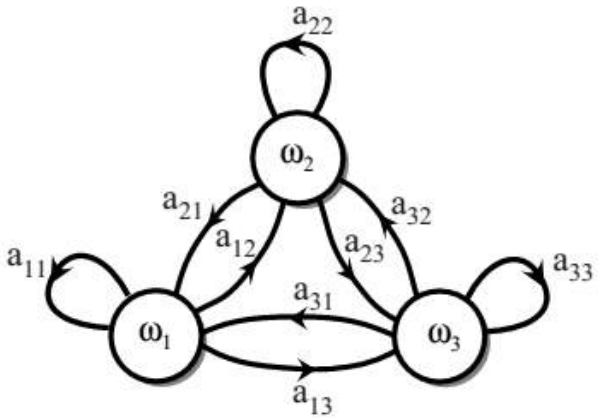


Figure 2.17: Transition probabilities a_{ij} for a basic Markov model with three states (ω_1,ω_2 and ω_3). Image obtained from DUDA; HART; STORK (2012).

The model is then that in any state $\omega(t)$ has a probability of emitting a particular visible state or observation $o_k(t)$. This probability is denoted as $P(o_k(t)|\omega_j(t)) = b_{jk}$. This full model is called hidden Markov model because it has access only to the visible states, while the ω_i are unobservable (see Figura 2.18).

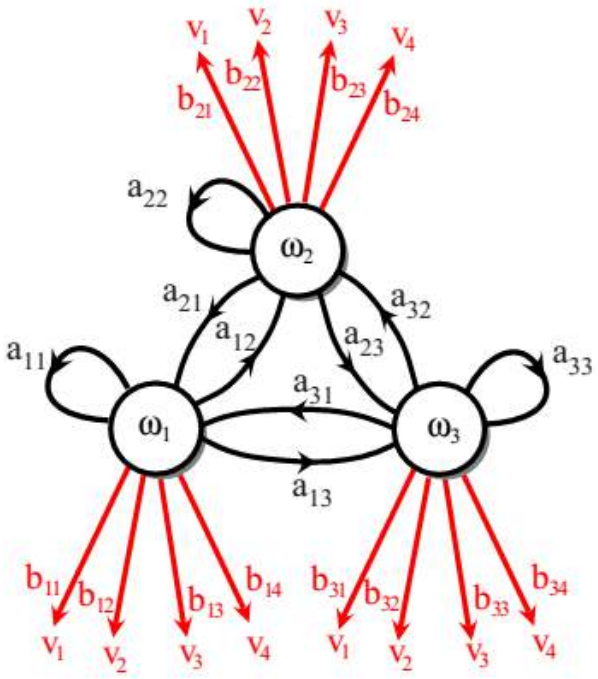


Figure 2.18: Transition a_{ij} and emission b_{ij} probabilities for a hidden Markov model with three states (ω_1,ω_2 and ω_3). Image obtained from DUDA; HART; STORK (2012).

A hidden Markov model denoted as λ is, therefore, completely described by

- a finite set of states $\{\omega_i | 1 \leq i \leq T\}$ that are usually only referred to by their indices in the literature,
- a matrix A of state-transition probabilities $A = \{a_{ij} | a_{ij} = P(\omega_j(t) | \omega_i(t-1))\}$
- and state specific probability distributions $B = \{b_{jk} | b_{jk} = P(o_k(t) | \omega_j(t))\}$ for the outputs of the model.
- a vector π of start probabilities $\pi = \{\pi_i | \pi_i = P(\omega_i(1))\}$. In this work initial probability distribution is assumed to be uniform and therefor could be ignored from further formulation (DUDA; HART; STORK, 2012).

Its demanded that some transition occur from step $t \rightarrow t+1$ and that some visible symbol be emitted after every step. Thus we have the normalization conditions:

$$\sum_j a_{ij} = 1 \text{ for all i}$$

and

$$\sum_k b_{jk} = 1 \text{ for all k}$$

With these preliminaries, it can be defined the three central issues in hidden Markov models:

- **Learning problem**.Given the coarse structure of a model (the number of states and the number of visible states) but not the probabilities a_{ij} and b_{jk}. Given a set of training observations of visible symbols O, determine these parameters.
- **Evaluation problem**. Given an HMM λ with transition probabilities a_{ij} and b_{jk}. Determine the probability that a particular sequence of visible states O^T was generated by that model $P(O|\lambda)$.
- **Decoding problem**. Given an HMM as well as a set of observations O. Determine the most likely sequence of hidden states ω that led to those observations.

In this work only the first two issues are needed to be solved and are related next.

2.2.4.1 Evaluation

The probability that the model produces a sequence O of observations is:

$$P(O|\lambda) = \sum_{r=1}^{r_{max}} P(O, \omega_r|\lambda) = \sum_{r=1}^{r_{max}} P(O|\omega_r, \lambda)P(\omega_r|\lambda) \tag{2.18}$$

where each r indexes is a particular sequence $\omega_r = \{\omega_{r1}(1), \omega_{r2}(2), ..., \omega_{rT}(T)\}$ of T hidden states. In the general case of c hidden states, there will be $r_{max} = c^T$ possible terms in the sum of Equação 2.18, corresponding to all possible sequences of length T. Thus, in order to compute the probability that the model generated the particular sequence of T visible states O, it should take each conceivable sequence of hidden states, calculate the probability they produce O, and then add up these probabilities. The probability of a particular visible sequence is merely the product of the corresponding (hidden) transition probabilities a_{ij} and the (visible) output probabilities b_{jk} of each step.

In first order HMM likelihood and prior terms of Equação 2.18 are calculated as:

$$P(O|\omega_r, \lambda) = \prod_{t=1}^{T} P(o_k(t)|\omega_{r_j}(t)) = \prod_{t=1}^{T} b_{jk}^t \tag{2.19}$$

$$P(\omega_r|\lambda) = \prod_{t=1}^{T} P(\omega_{r_j}(t)|\omega_{r_i}(t-1)) = \prod_{t=1}^{T} a_{ij}^t \tag{2.20}$$

Substituting Equação 2.19 and Equação 2.20 in Equação 2.18 is obtained

$$P(O|\lambda) = \sum_{r=1}^{r_{max}} \prod_{t=1}^{T} a_{ij}^t b_{jk}^t \tag{2.21}$$

which has complexity O($c^T T$) and is prohibitive in practice.

Since each term $P(O|\omega_r)P(\omega_r)$ involves only $o_k(t), \omega_j(t)$ and $\omega_i(t-1)$ a recursively calculation can be done defining

$$\alpha_j(t) = \begin{cases} \pi_j b_{jo_0} & t = 0 \\ \sum_i \alpha_i(t-1) a_{ij} b_{jk} & \text{otherwise} \end{cases} \tag{2.22}$$

Thus $\alpha_i(t)$ represents the probability that our HMM is in hidden state ω_i at step t having generated the first t elements of O. This calculation is implemented in the Forward algorithm in the following way:

Algorithm 2.1: Forward algorithm

input : O

output : $P(O|\lambda)$

1 **Initialize** $t = 0, A, B, \alpha_j(0) = \pi_j b_{jo_0}$;

2 **for** $t = t+1$ **to** $t = T$ **do**

3 $\quad \alpha_j(t) = \sum_{i=1}^{c} \{\alpha_i(t-1) a_{ij}\} b_{jk}$;

return $\sum_{i=1}^{c} \alpha_T(i)$

An example of calculation with Forward algorithm is shown in Figura 2.19.

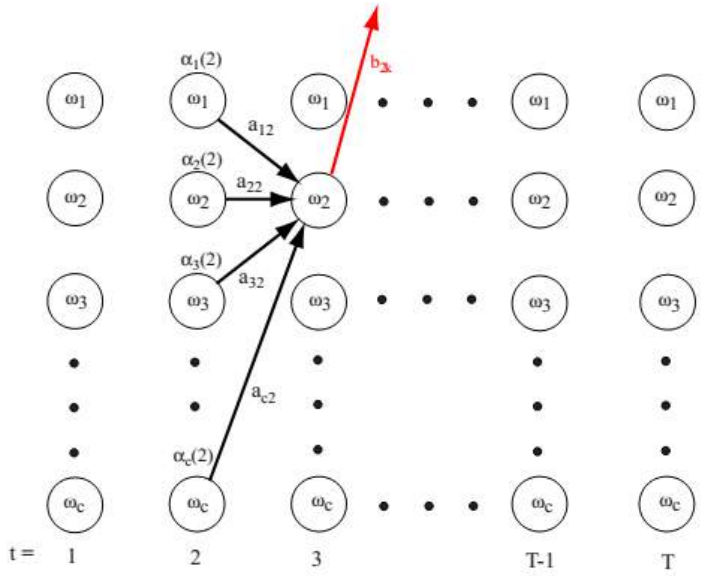


Figure 2.19: Probabilities computation of Forward algorithm. Image obtained from DUDA; HART; STORK (2012).

For the mathematical exact determination of the total output probability it is necessary to include all possible paths through the respective model into the computations. For the evaluation of the modeling quality of an HMM this summarizing consideration is, however, not necessarily the only procedure that makes sense. A model which is satisfactory on the average can be discriminated from another one which works especially well in certain cases if only the respective optimal possibility is considered to generate a certain observation sequence with a given model.

When disregarding efficiency considerations, the optimal output probability $P^*(O|\lambda)$, i.e. the optimal probability $P(O, \omega^*|\lambda)$ for generating the observation sequence along a specific path,

can be determined by maximization over all individual output probabilities given in Equação 2.21.

$$P(O,\omega^*|\lambda) = \max_{\omega_i} P(O,\omega_i|\lambda) = \max_{\omega_i} \prod_{t=1}^{T} a_{ij}^t b_{jk}^t \tag{2.23}$$

A much more efficient method for the computation of this quantity is obtained as a slight variation of the forward algorithm by again applying the considerations about the finite memory of HMMs. The resulting method is named Viterbi algorithm for which the computation procedure is summarized in algorithm 2.2.

Algorithm 2.2: Viterbi algorithm

input : O

output : $P(O|\lambda)$

1 **Initialize** $t = 0, A, B, \delta_j(0) = \pi_j b_{jo_0}$;

2 **for** $t = t+1$ **to** $t = T$ **do**

3 $\quad \delta_j(t) = \max_i\{\delta_i(t-1)a_{ij}\}b_{jk}$;

return $\sum_{i=1}^{c} \delta_T(i)$

An example of calculation with Viterbi algorithm is shown in Figura 2.20.

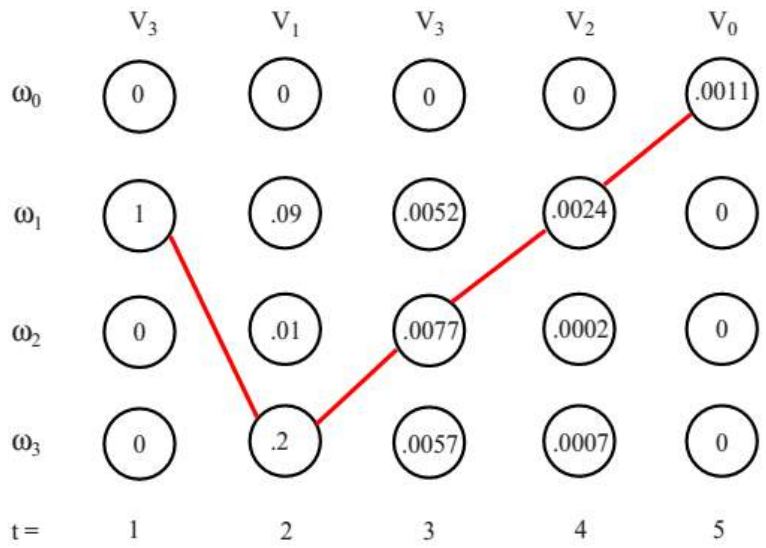


Figure 2.20: Probabilities computation by Viterbi algorithm. Image obtained from DUDA; HART; STORK (2012).

2.2.4.2 Learning

The goal in HMM learning is to determine model parameters, the transition probabilities a_{ij} and b_{jk}, from an ensemble of training samples O. The way to do this is through the Forward-backward algorithm, also known as Baum-Welch algorithm. This approach will be iteratively update the weights in order to better explain the observed training sequences.

The algorithm uses $\alpha_i(t)$ defined in Equação 2.22 and $\beta_i(t)$ which is probability that

model is in state $\omega_i(t)$ and will generate the remainder of the given target sequence. The definition of $\beta_i(t)$ is:

$$\beta_i(t) = \begin{cases} 1 & t = T \\ \sum_j \beta_j(t+1) a_{ij} b_{jk} & \text{otherwise} \end{cases} \tag{2.24}$$

But the $\alpha_i(t)$ and $\beta_i(t)$ determined only estimates of their true values, since are not know the actual value of the transition probabilities a_{ij} and b_{ij}. It can be calculated an improved value by first defining $\gamma_{ij}(t)$, the probability of transition between $\omega_i(t-1)$ and $\omega_j(t)$, given the model generated the entire training sequence O by any path. Definition of $\gamma_{ij}(t)$ is as follows:

$$\gamma_{ij}(t) = \frac{\alpha_i(t-1) a_{ij} b_{ij} \beta_j(t)}{P(O|\lambda)} \tag{2.25}$$

where $P(O|\lambda)$ is calculated as in Equação 2.21.

Thus, state transition probability estimation $\hat{a}_{ij}$ can be found by taking the ratio between the expected number of transitions from ω_i to ω_j and the total expected number of any transitions from ω_i. That is:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T} \gamma_{ij}(t)}{\sum_{t=1}^{T} \sum_{k=1}^{c} \gamma_{ik}(t)} \tag{2.26}$$

Similarly, improved estimate b_{jk} can by calculated as the ratio between the frequency that any particular symbol o_k is emitted and that for any symbol:

$$\hat{b}_{jk} = \frac{\sum_{t:o(t)=o_k} \gamma_{jk}(t)}{\sum_{t=1}^{T} \sum_{k=1}^{c} \gamma_{jk}(t)} \tag{2.27}$$

The Baum-Welch algorithm is then defined as an instance of a generalized Expectation-Maximization algorithm and start with rough or arbitrary estimates of a_{ij} and b_{jk}, calculate improved estimation by Equação 2.26 and Equação 2.27, and repeat until some convergence

criterion is met. The full algorithm is shown in algorithm 2.3.

Algorithm 2.3: Baum-Welch algorithm

input : O, convergence criterion ε

output : A, B

Initialize $a_{ij}^0, b_{jk}^0, z = 0$;

while $\max_{i,j,k}[a_{ij}^z - a_{ij}^{z-1}, b_{jk}^z - b_{jk}^{z-1}] > \varepsilon$ **do**

$z = z + 1$;

$\hat{a}_{ij}^z = \frac{\sum_{t=1}^{T} \gamma_{ij}^{z-1}(t)}{\sum_{t=1}^{T} \sum_{k=1}^{c} \gamma_{ik}^{z-1}(t)}$;

$\hat{b}_{jk}^z = \frac{\sum_{t:O_t=o_k} \gamma_{jk}^{z-1}(t)}{\sum_{t=1}^{T} \sum_{k=1}^{c} \gamma_{jk}^{z-1}(t)}$;

$a_{ij}^z = \hat{a}_{ij}^{z-1}$;

$b_{jk}^z = \hat{b}_{jk}^{z-1}$;

return $a_{ij} = a_{ij}^z, b_{jk} = b_{jk}^z$

In HMM pattern recognition we would have a number of trained HMMs, one for each category and classify a test sequence according to the model with the highest probability.

3

Proposed Method

It has been seen that the object recognition problem can be defined as a labeling problem based on models of known objects. The underlying issue particularly here investigated comes when objects are highly occluded and no information are given about the quantity of objects in the occlusion. In addition, it is unknown which parts of the contour belong to the same object and which object categories are present in analysed occlusion. Previous literature works lacks of ability to deal with these problems, aggravated when severe occluded objects are analysed. Proposed method solves multiples severe occluded objects recognition task without any of related issues using following steps: (1) separating the occlusion into parts, where each point of the part meets those other belonging to the same object; (2) estimating the best hypotheses for each part, as an image retrieval problem; and (3) validating the hypotheses for elimination of duplicated or wrong estimated objects.

The method here first separates parts by defining High Curvature Points (HCP) (GUERRERO-PENA et al., 2015), which have the property of representing object intersections in an occlusion with high probability, and provide invariance to traslation, rotation and scale. Then, Tangent Angle Signature (TAS) (ZAHN; ROSKIES, 1972) is calculated for each part to obtain local shape descriptors that can handle occlusion. Continuous wavelet transform (CWT) (MALLAT, 2008) is computed for each part to represent its main signal and filter the desired amount of noise. Further, the best matching object is calculated for each part using a bayesian approach with Pearson's correlation coefficient between representation of the part and most probable objects in the database as likelihood. For selection of most probable objects, an ensemble of hidden Markov models (HMM) is created with a model for each object category trained with the one-class approach (KOCH et al., 1995). Models are scale invariant and are used to recognize the most probable class for each object without the need for part length normalization. For hypotheses validation, two area relation constraints are set proposed to determine the prior leaving only the hypotheses with visual consistency with the occlusion. Figura 3.1 illustrates above steps.

The bayesian method for recognition of non-deformable occluded object is proposed without any information about quantity of objects present in the occlusion, the parts of the occluded contour that belong to each object and the classes of the occluded objects. Also, an

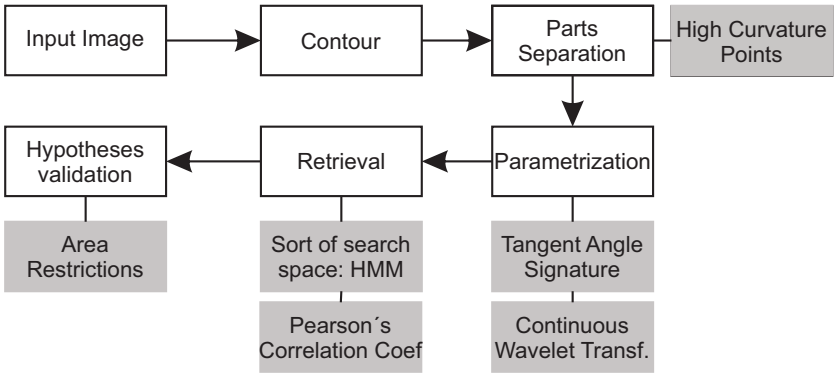


Figure 3.1: Proposed method scheme.

optimization is proposed for the retrieval of objects based on an ensemble of Hidden Markov Models to a reduction of search space. The input to the method is an object shape represented by an external closed contour, given by a segmentation process or background subtraction technique, as seen in Figura 3.2. In the next sections, details for each step of the method are provided.

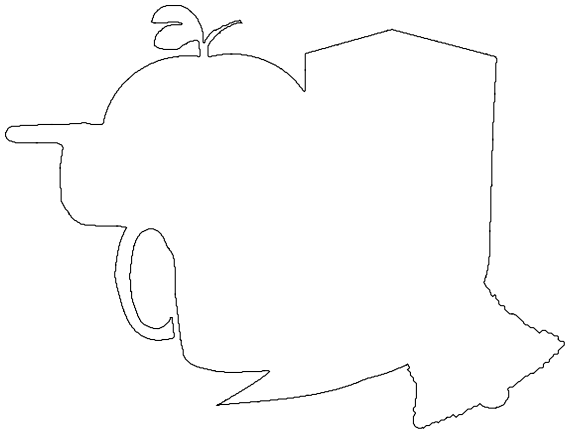


Figure 3.2: Example of input contour.

3.1 Parts Separation

In order to recognize occluded parts in an image, the first step should consider isolating each part from the rest. Then, a representation for every part is needed. As discussed in previous sections, for occluded object recognition, local representations tend to be more suitable than global ones because changes in some points will not affect all descriptors.

To accomplish that, geometric attributes are employed to generate local descriptors for each separate part, thus maintaining representation of main object elements even if contours are partially missing or modified. High curvature points are used to define the parts, located through the k-curvature method (GUERRERO-PENA et al., 2015), which calculates HCPs by local maximum point of k-curvature using Equação 2.11. A threshold (sT) is established for filtering the new signal, which conveniently softens the curve and eliminates false concave points

that correspond to contour noise concavities. All of the values that exceed the threshold are fixed to 1, which generates small point sequences that correspond to the contour and that have k-curvature values over the threshold (see Figura 3.3(d)), where the last point of these small points sequences (marked with * in Figura 3.3(d)) is taken as a point of interest. Figura 3.3 shows obtained k-curvatures according to Equação 2.9, Equação 2.10 and Equação 2.11 for input contour of Figura 3.2.

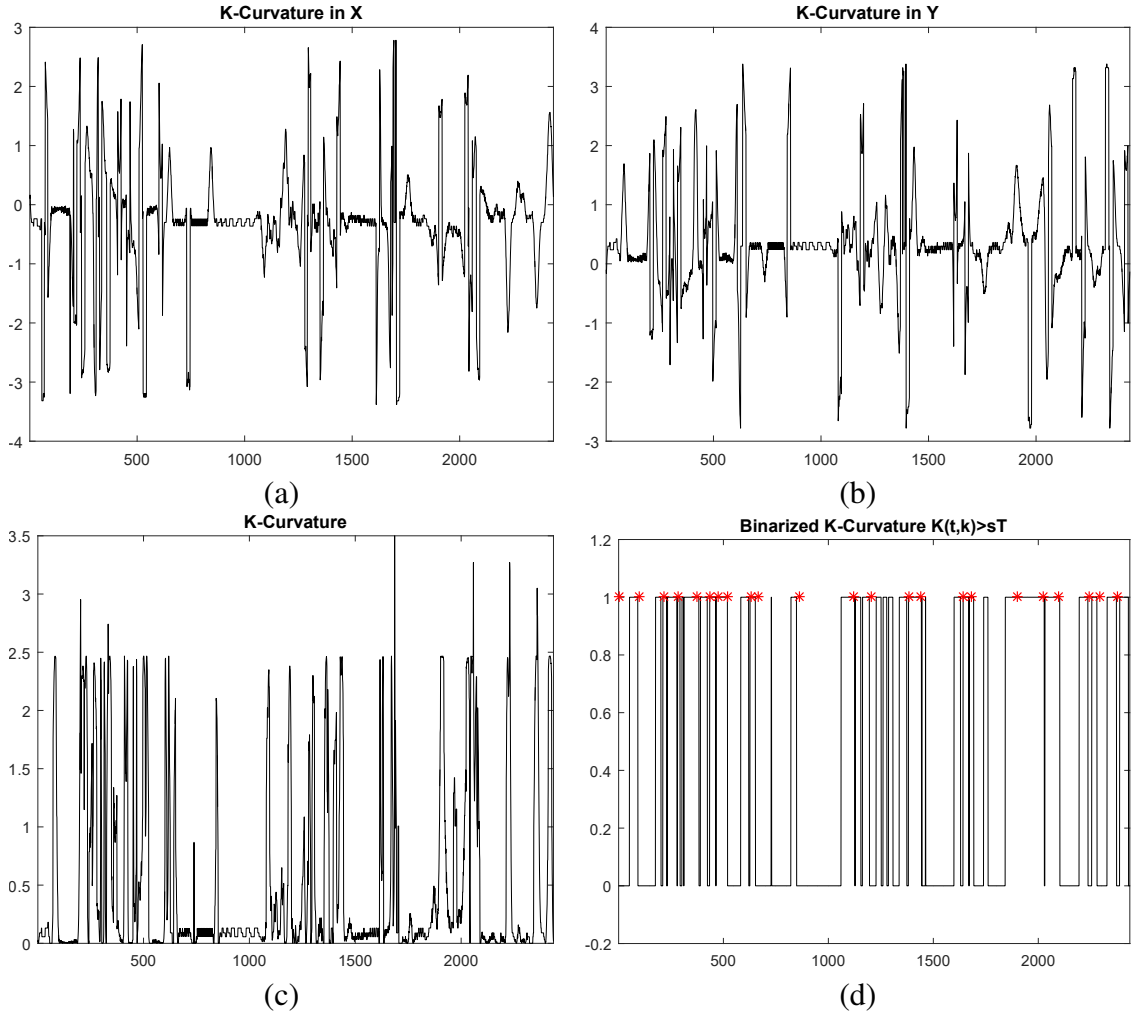


Figure 3.3: Examples of k-curvature graphs (a) x k-curvature, (b) y k-curvature, (c) k-curvature, and (d) k-curvature binarization for $sT = 0.15$.

To determine whether the point is concave or convex, their neighbors in the k positions to the right and left are taken and the middle of the line formed between the neighbors is considered. If this midpoint is interior to the region, the point of interest represents a convexity, but otherwise it represents a concavity (see Figura 3.4). This method facilitates the efficient detection of concave points in the contour and it almost completely avoids the points that correspond to contour noise concavities or details of the contour. Concave and convex points detected by the method are show in Figura 3.4 where blue circles represents convex points and red squares concave points. Only concave points are selected for part separation.

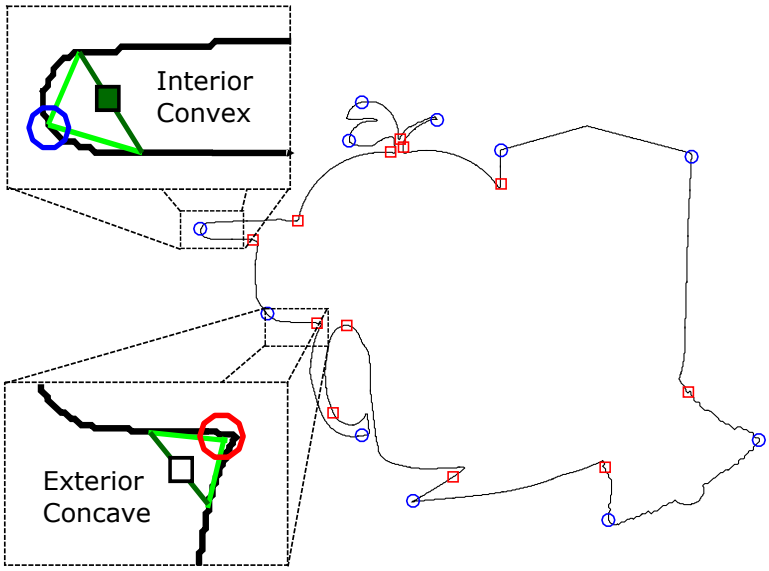


Figure 3.4: High curvature points detection and part separation.

Points selected are indicative of the joint between two objects along the occlusion, and are then considered to determine the parts. Also, such points maintain the same high curvature under different Euclidian shape transformations. Method operation depends on parameter k, regarding the desire smoothness of the curvature, whose value is set empirically. In this work, k was set to 5 (pixels), after several trials, providing the best location of high curvature points along shape contours. The result of part separation is shown in Figura 3.5, been detected 12 parts for the example.

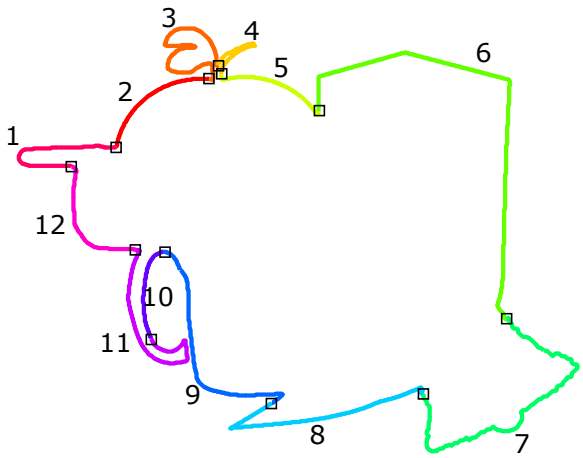
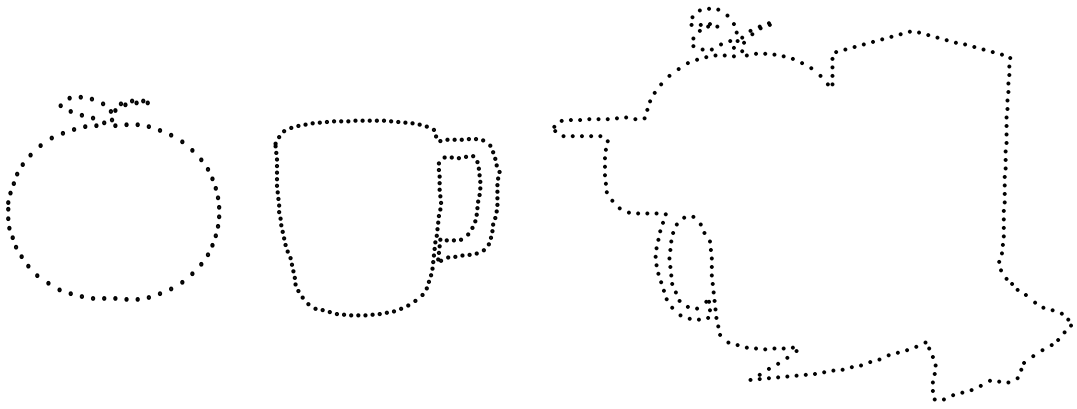


Figure 3.5: High curvature points detection and part separation (in different colors).

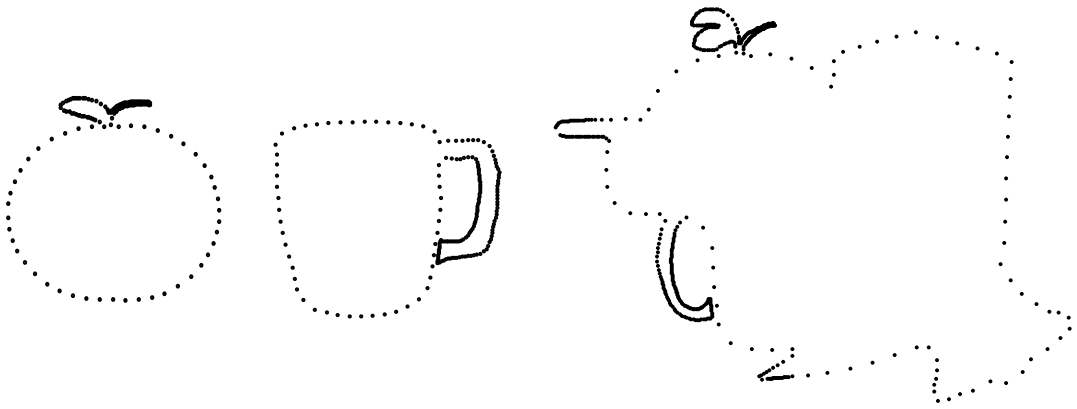
The *k-th* part of the occlusion O is defined as $P_k = \{O_j \mid Hc_k \leq j \leq Hc_{k+1}\}$, being O_j the *j-th* point of the contour O and Hc_k the *k-th* HCP of the contour. In general, the part are formed by the points of the contour between every pair of adjacent HCP (see Figura 3.5). The Tangent Angle Signature of the elements of the set of all founded parts $\{P_k\}$ is then obtained.

3.2 Wavelet Filtering

To reduce noise, some works carry out a sampling of the contour points (BELONGIE; MALIK; PUZICHA, 2002; CHEN; FERIS; TURK, 2008; MICHEL; OIKONOMIDIS; ARGYROS, 2011). The uniform sampling approach (BELONGIE; MALIK; PUZICHA, 2002; CHEN; FERIS; TURK, 2008) (Figura 3.6 (a)) selects a subset of contour points with fixed length between them, achieving an approximation of the shapes with thinner representations. Non-uniform sampling (MICHEL; OIKONOMIDIS; ARGYROS, 2011) (Figura 3.6 (b)), on the other hand, selects a contour points subset but with variable steps where small concave parts are obtained with a smaller step size and wide open parts have greater step sizes. Example of isolated and occluded objects sampling using uniform and non-uniform variants are shown in Figura 3.6.



(a)



(b)

Figure 3.6: Examples of isolated and occluded objects with (a) uniform and (b) non-uniform sampling.

In addition to spacial domain sampling approaches for noise reduction, the problem could be addressed through histograms (BELONGIE; MALIK; PUZICHA, 2002), moments (ZHANG; LU, 2004), scale space (MOKHTARIAN et al., 1997) and spectral transform (YANG; KPALMA; RONSIN, 2008) methods. Although histogram and scale space can increase robustness to noise

and compactness, matching using histogram and scale space can be very expensive. Moments is robust and compact, however, higher order moments are either difficult to obtain or without physical meaning. Among the four techniques, spectral transforms is the most used for this task.

One-dimensional continuous wavelet transform of TAS is employed here as an alternative to sampling approaches because it is an efficient strategy for analysing non-stationary local signals such as the obtained signatures. Advantage of using wavelet among other spectral approaches like Fourier and Short Time Fourier is described in Subseção 2.2.3. Two wavelet based approaches are considered here. The first one is calculate 1-Dimensional Continuous Wavelet Transform of obtained TAS and use it in retrieval step. In this work the two first scales were sufficient to reduce the contour noise.

Occlusion signature in wavelet space is a composite of multiples parts of signatures of the isolated objects that compound the occlusion. In Figura 3.7, is shown the CWT of the occlusion of the Figura 3.5; and the CWT of the isolated apple and cellular phone objects that are occluded in this example. The signalized matched part of the CWT corresponds with the visible parts of the objects in the occlusion.

One of the advantages of the combination TAS+CWT is that maintain local features of each part (see Figura 3.7) and its very efficient in perform calculation. This aspect leads to facilitates the shape retrieval process. The resulting CWT of a part has the same length of the respective part in terms of contour points that belongs to the part.

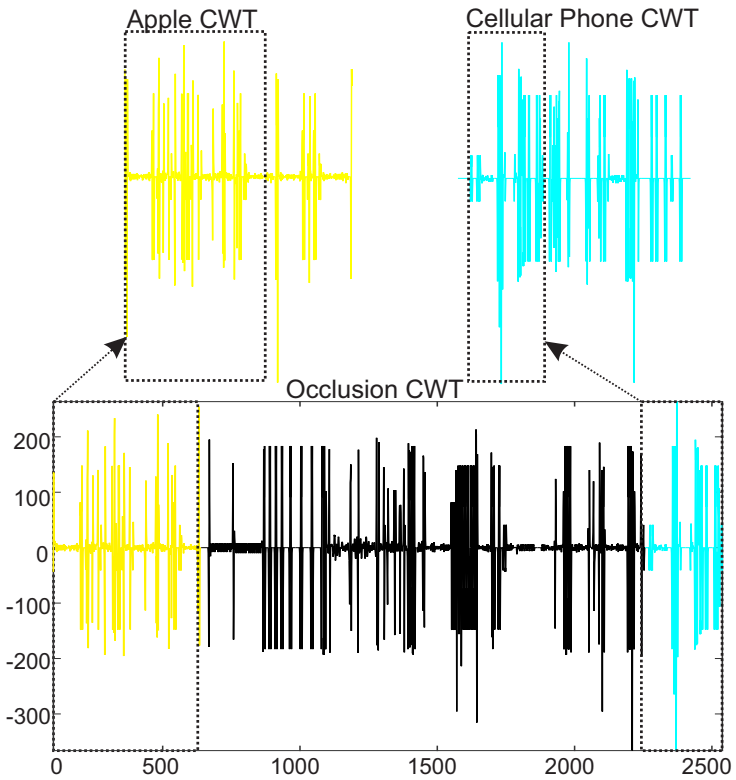


Figure 3.7: Occlusion signature in wavelet space with matched part of isolated objects.

The second approach is wavelet based filtering, where a low pass filter is applied for

the 30 lower scales as proposed in KON'YA; KUSHIMA (1998), reducing significantly noise influence while keeping TAS relevant features. For this, all coefficients of lower scales that are minor than selected τ_0 are set to 0. A reconstructed signal x from the filtered wavelet coefficients is then obtained and corresponds to a denoised TAS representation, used then in retrieval step. This approach performs better noise reduction but involve more computation because filtering and a signal reconstruction step are added when compared to previous approach.

In both cases, the mother wavelet used was the Mexican Hat (Equação 3.1), very useful for detection of local maximum since it has poor frequency resolution (MALLAT, 2008).

$$\psi(t) = \frac{2}{\pi^{\frac{1}{4}}\sqrt{3\sigma}}\left(\frac{t^2}{\sigma^2} - 1\right)e^{-\frac{t^2}{2\sigma^2}} \tag{3.1}$$

with σ set to 1 for balance between noise reduction and relevant feature extraction.

The output of this step is a set of one dimensional signals $\{x_k\}, k = 1..N$ that in first approach corresponds to CWT of TAS for a given scale and in second approach corresponds to denoised TAS signal.

3.3 Retrieval

Once parts of the occlusion are obtained and denoised, hypotheses for each part in the occlusion are formulated. Hypotheses correspond to isolated objects from the database and object that best matches the analyzed part is taken as selected hypothesis. This leads to a shape retrieval problem where an object that contains a given open curve has to be found. Supposing every part x comes from a certain complete shape y from the dataset, the retrieval process is formulated as a Maximum a Posteriori (MAP) problem to obtain the most probable unknown object y given the open curve x,

$$y^* = \arg\max_y p(y|x) = \arg\max_y \frac{p(x|y)p(y)}{\int p(x|y)p(y)dy} \tag{3.2}$$

The recovered object not only has to contain query x but also must satisfy the constraint it is a valid hypothesis for a given occlusion O. So, the MAP problem of Equação 3.2 is defined as,

$$y^* = \arg\max_y p(y|O,x) \tag{3.3}$$

For a retrieved object y as an object recognition problem there is a hidden variable indicating the class of the object, so the posterior in Equação 3.3 is computed by,

$$p(y|O,x) = \sum_{i=1}^{C} p(y|O,x,\lambda_i)p(\lambda_i|x) \tag{3.4}$$

The $p(y|O,x,\lambda_i)$ term is the probability that the recovered object of the class i satisfies

occlusion constrain and $p(\lambda_i|x)$ is the probability that the part x is present in the objects of the class i. In practice however each object only belongs to one of the classes leading to the approximation of the Equação 3.4,

$$p(y|O,x) \approx p(y|O,x,\lambda_i^*)p(\lambda_i^*|x) \tag{3.5}$$

removing constant marginal term from Equação 3.2 and assuming object retrieval's a priori probability $p(y)$ follows a uniform distribution, where λ_i^* is the class model of the object with maximum posterior probability.

By Bayes theorem its obtained the relation

$$p(\lambda_i|x) = \frac{p(x|\lambda_i)p(\lambda_i)}{p(x)} \tag{3.6}$$

and because probability of classes λ_i and curves x occurs is uniform, could be approximated $p(\lambda_i|x) \approx p(x|\lambda_i)$.

Given a reference class i, the best object that satisfy the constrains is selected maximizing the $p(y|O,x,\lambda_i)$ probability define by,

$$p(y|O,x,\lambda_i) \approx p(O|y,x,\lambda_i)p(x|y) \tag{3.7}$$

The $p(O|y,x,\lambda_i)$ is the likelihood and represents the partial consistency between the occlusion O and the object y. The probability $p(x|y)$ is the prior and measures the similarity between the query x and the best match part of the object in other words it calculates how probable part x belongs to y.

Replacing Equação 3.7 in Equação 3.5 a retrieval method is obtained which incorporates the knowledge of the coherence of the retrieved object with the occlusion, the similarity of the query part to the best match part in the object and the occurrence frequency of the part into the objects of the class.

$$p(y|O,x) = p(O|y,x,\lambda_i^*)p(x|y)p(x|\lambda_i^*) \tag{3.8}$$

Despite many objects could have high priori probability given a query x (classical object retrieval), the likelihood will ensure a hypothesis that has visual consistency with the occlusion. In the following will be introduced how to compute each probability term of the method.

3.3.1 Class posterior probability

The retrieval process has to be repeated to a query part for each object in the database and for problems with many objects, this could be a time-consuming task. Because the estimation of y is based on a dataset with multiple classes, the knowledge of how likely part x is for a class i is incorporated to the posteriori calculation in Equação 3.5 in order to sort out and reduce the search space.

For computation of the $p(x|\lambda_i)$ term an ensemble of classifiers is defined as $E = \{\lambda_i \mid 1 \leq i \leq C\}$ where λ_i is a trained HMM and C is the number of object classes. HMMs capabilities for shape classification were studied in BICEGO; MURINO (2004); MANDAL; Mahadeva Prasanna; SUNDARAM (2015) and demonstrated robustness for occluded object recognition.

Every λ_i is trained with the CWTs of objects from class i (one-class problem KOCH et al. (1995)), with CWTs corresponding to the set of observations received by the Baum-Welch algorithm (DURBIN et al., 1998).

Each λ_i is trained with the representation of objects from class i (one-class problem (KOCH et al., 1995)), being the $\{x_k\}, k = 1..N$ the set of observations received from Baum-Welch algorithm (Subsubseção 2.2.4.2). Training is made although that the objects of the class have different length. For determining HMMs parameters (number of states, transition and emission matrices) two frequent approaches are employed: automatic and experimental (MANDAL; Mahadeva Prasanna; SUNDARAM, 2015). In this work, the experimental approach was used, observing different parameter values just changed the search space ordering and larger number of states only resulted in slower training. For this reason, the number of states in HMMs was set to 2, as initial choice, with transition and emission matrices randomly initialized. For every class were generated various models varying randomly the initialization and the model with lowest training error were selected to represent the class. Note that the training set of the λ_i corresponds to non-occluded examples of objects of class i.

For the selection of most probable objects in the retrieval step they are sorted in a manner that the objects of most likely classes go first. For this, the probability $p(x|\lambda_i)$ that the set of query observations x could be generated by the model λ_i is calculated through the Viterbi algorithm (Subsubseção 2.2.4.1). The sort of objects is done in descending order of probabilities, in such a way that the objects probability corresponds to their respective class posterior probability $p(x|\lambda_i)$. In the search space of retrievals(Search Space Retrieval (SSR)) unsorted version, conversely, the default order of SSR for all parts is usually considered the same and predefined.

If the query part is very frequently in most classes of objects, for example, a straight line, the hypotheses obtained will have a minor retrieved posteriori probability $p(y|O,x)$ than the others, and the validation step ensure in most of the cases the elimination of wrong retrieved hypotheses using this information. An example of contour parts classification after the class posterior probability calculation is illustrated in Figura 3.8.

3.3.2 Query prior

The best match between an open curve and a closed contour is a partial shape matching problem. The similarity of the curve and the matched part of the target is used as prior probability $p(x|y)$ on Equação 3.7 and gives how likely is that the part x is contained in the object y. For this purpose, cross-correlation match of signals in wavelet space between query part x and target object y in the database is used.

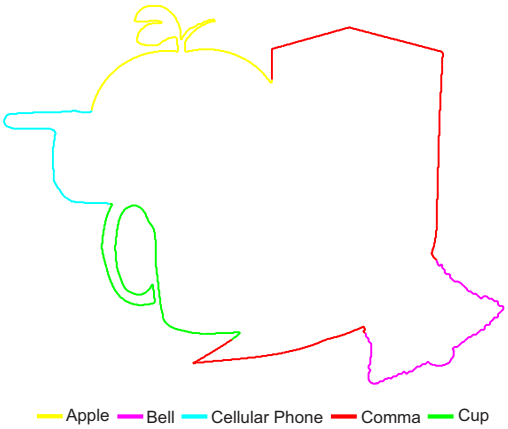


Figure 3.8: HMM classification for contour parts.

The most significant value of the cross-correlation signal indicates the best correlation between the query and the part of the target y at location l^*. Pearson's correlation coefficient between query and matched target part is used as prior $p(x|y)$, Equação 3.9. This equation calculates how similar two given signals are and was used for the partial shape matching with success in CUI et al. (2009).

$$p(x|y) = corr(x, y_{l^*}) \tag{3.9}$$

with

$$corr(X,Y) = \frac{\sum_{i=1}^{n}(X(i)-\overline{X})(Y(i)-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X(i)-\overline{X})^2 \sum_{i=1}^{n}(Y(i)-\overline{Y})^2}} \tag{3.10}$$

been n the length of the signals X and Y.

The object with major priori probability is the best match for the reference class i, therefore, the hypothesis for the *k-th* query in that class:

$$y_k = \arg\max_{y}(p(x|y)) \tag{3.11}$$

The set of hypotheses is defined as $H = \{y_k\}$, being y_k defined in Equação 3.11. For each hypothesis, Euclidean transformation that best fits the hypothesis with the part is calculated using the RANSAC algorithm (FISCHLER; BOLLES, 1981).

As search stopping rule, the search space of object recognition is explored (in any order) until a threshold τ_1 is reached, indicating the minimum similarity required between the query part and the target for accepting the object identification, as given by Equação 3.12.

$$p(x|y) > \tau_1 \tag{3.12}$$

with y been selected in any given established order.

3.3.3 Occlusion likelihood

For occlusion likelihood computation the visual consistency of the retrieved object y with the occlusion O has to be measure. Visual consistency could be defined as the probability that a point on the object y corresponds to a point in the occlusion. This involve computing the area of hypothesis inside the occlusion and the area of the hypothesis y. Then, the occlusion likelihood in Equação 3.7 is just a ratio of these two areas.

$$p(O|y,x,\lambda_i) \approx p(O|y) = \frac{area(y \bigcap O)}{area(y)} \tag{3.13}$$

Objects with low $p(O|y)$ are caused by wrong transformation estimations or wrong object recognition. If the hypothesis is not consistent with the occlusion the likelihood will be much less than 1.0 but the retrieval posterior probability could be higher than other valid hypotheses depending on the prior and class posterior probability. In practice a minimum fitness τ_2 between hypothesis y and occlusion O could be established in order to reduce processing time and discard wrong estimated hypotheses where $p(O|y) = 0$. For that, likelihood probability (Equação 3.13) is redefined as

$$p(O|y) = \max\left[\frac{area(y \bigcap O)}{area(y)} - \tau_2, 0\right] \tag{3.14}$$

Six examples of estimated hypotheses with their associated occlusion likelihood probabilities calculated as Equação 3.13 are shown in Figura 3.9. Blue contours correspond to occlusion O and red contours to estimated hypothesis y. Dark red areas correspond to $p(O|y)$ likelihood and query x is delimited by two marked contour points in each case.

Using Equação 3.14, $p(O|y)$ probability of case (e) from Figura 3.9 becomes equal to zero for $\tau_2 = 0.85$. These means, these hypotheses will have zero $p(y|O,x)$ posterior probability and could be discarded for posterior processing.

3.4 Hypotheses Validation

Although only hypotheses with $P(y|O,x) > 0$ are considered as valid, objects with concavities in their geometry will be divided into smaller parts and multiple hypotheses will tend be associated with the same object. For validation of hypotheses, an area constraint is used. This restriction is used to eliminate duplicate y hypotheses. The area ratio calculated for hypothesis y_k is presented in Equação 3.15 and measures the rate of duplicated hypotheses respectively. Objects with high A_i are redundant because either the same hypothesis was already considered in another part or a wrong retrieved hypothesis was formulated. The valid set of objects recognized for occlusion O is defined as $V = \{y_k | A_k \leq \tau_3\}$.

$$A_k = \frac{area(y_k \bigcap O_k^{'})}{area(y_k)} \tag{3.15}$$

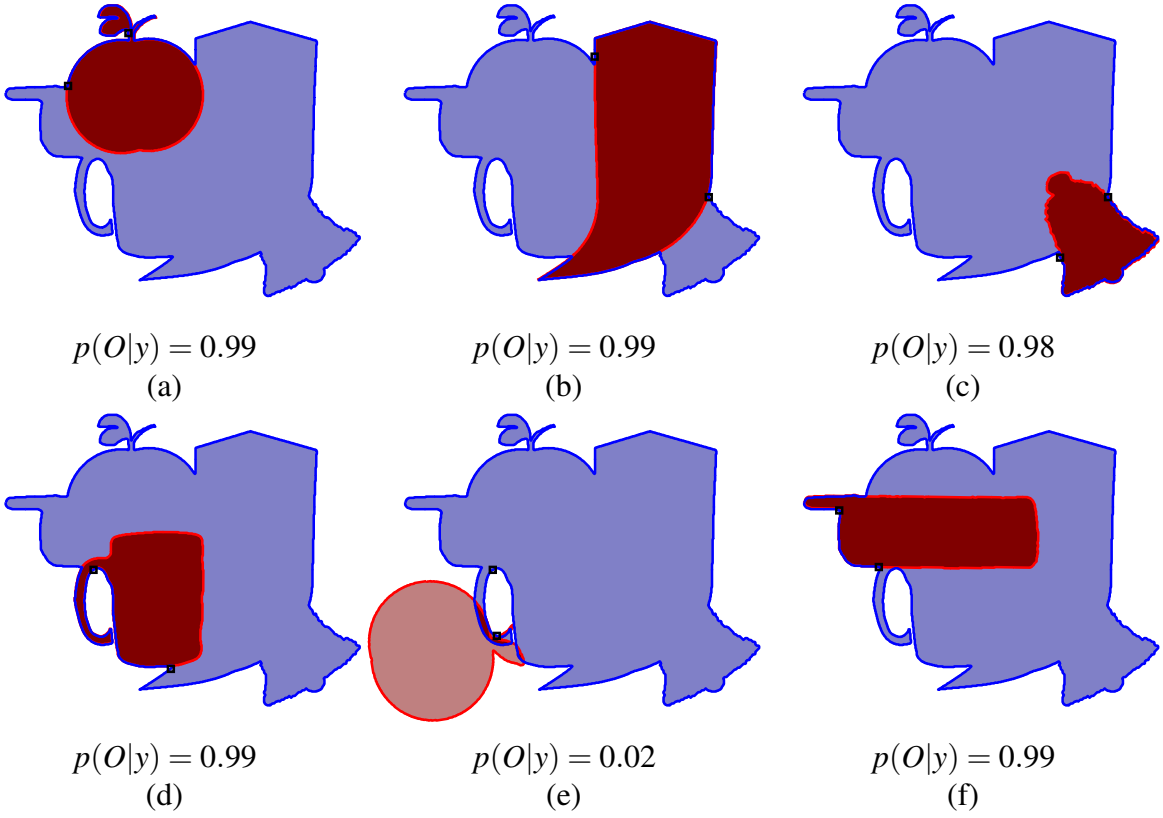


Figure 3.9: Examples of estimated hypotheses with corresponding occlusion likelihood (best viewed in color).

$$O'_k = O \bigcap \left(\bigcup_{\forall j \neq k} y^*_j \right) \tag{3.16}$$

The $\bigcup_{\forall j \neq k} y^*_j$ term in Equação 3.16 refers to the conjunction of all valid hypotheses except y_k and returns the generated occlusion of all valid hypotheses except y_k. The term $area(y_k \bigcap O'_k)$ in Equation 3.15 is the area of intersection between object y_k and generated occlusion O'_k. In each step, if calculated A_k do not meet restriction ($A_k \leq \tau_3$) hypotheses y_k is eliminated from V and this makes the process of validation iterative, and order of hypotheses evaluation affects the results. Then, for efficiency, hypotheses are sorted out in increasing order, given their corresponding posterior probability $p(y|O,x)$, to discard less probable hypotheses first.

Result of the validation step for each hypotheses is shown in Figura 3.10(a)-(h). Hypotheses are shown in the described order, where red contours represent analysed hypothesis y_k, blue contours represent generated occlusion O'_k, and dark red areas correspond to calculated numerator of A_k expression. In this example, hypotheses 2,3 and 6 are considered invalid ($A_2 > \tau_3$, $A_3 > \tau_3$ and $A_6 > \tau_3$ with $\tau_3 = 0.85$), and the final recognition result can be observed in Figura 3.10(i).

The intersection of two objects may not be a high curvature point, leading to wrong object retrieval. However, if objects are composed of multiple parts, correct hypotheses can still be retrieved, being duplicate hypotheses eliminated through validation constraint (Equação 3.15).

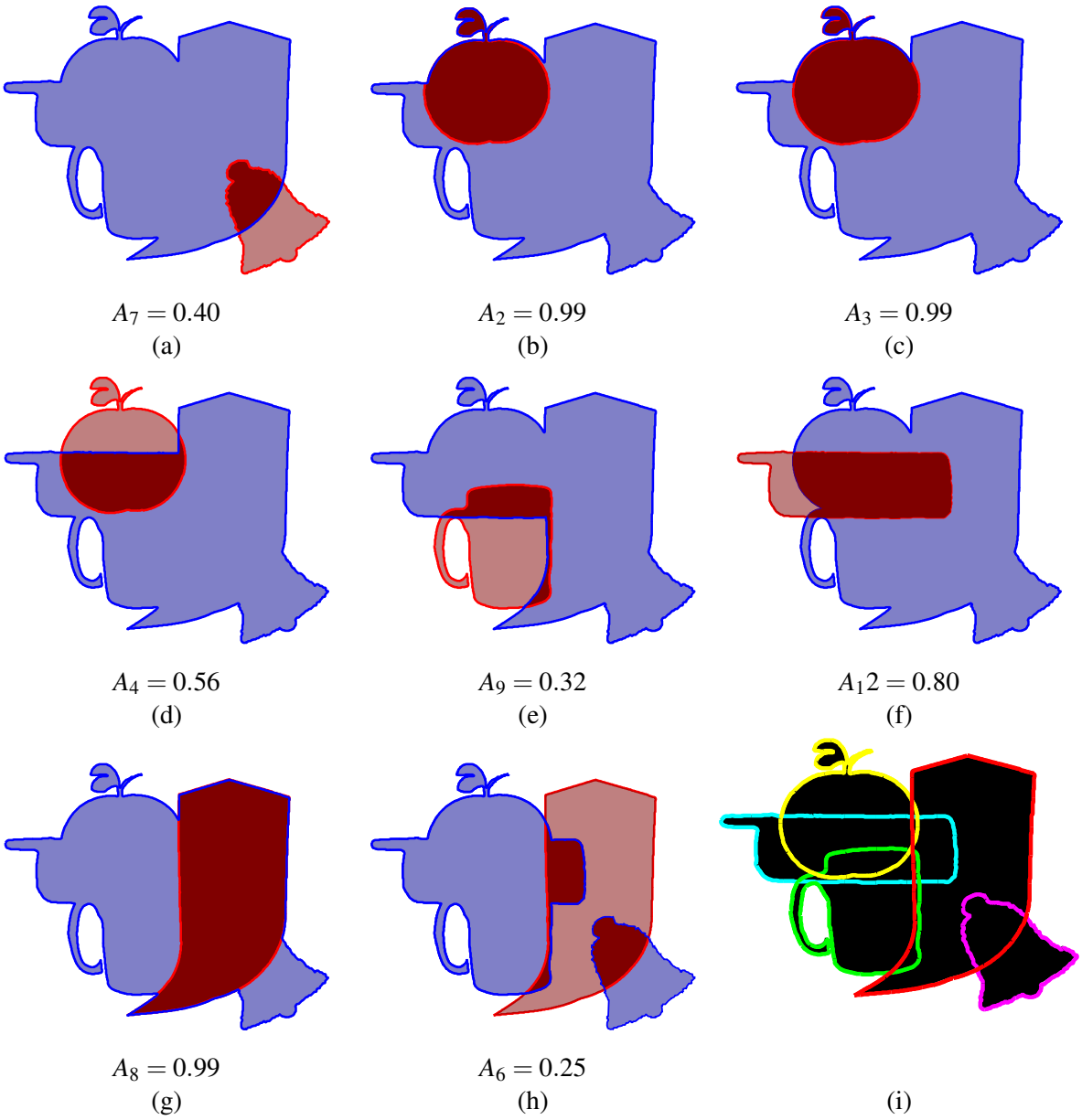


Figure 3.10: (a)-(h) Hypotheses validation step and (i) final recognition result.

Some examples of this situation are analysed step by step in Appendix A.

4

Experimentation and Comparison

To test the method, four experiments were carried out to evaluate performance, according to (1) level of object occlusion, (2) number of object classes in the problem, (3) number of objects in the occlusion and (4) inner occlusion. Also influence of proposed sorting search space retrievals and wavelet filtering is analyzed in several tests. For carrying out experiments, real life objects CMU_KO dataset and synthetical occluded objects from MPEG-7 dataset were used.

For carrying out the experiments, the widely used dataset MPEG-7 for shape analysis problems ((BELONGIE; MALIK; PUZICHA, 2002; LATECKI et al., 2007; CAO et al., 2011; MICHEL; OIKONOMIDIS; ARGYROS, 2011; MARVANIYA; GUPTA; MITTAL, 2015)) was employed. This dataset contains 1400 binary shape objects of 70 different classes and, in the experiments, a set of non-deformed severe-occluded shape contours was constructed in a synthetic way. For occlusion creation, n objects of any class were randomly selected, with $n = 2$ in experiment of subsections 4.1.2, 4.1.3, 4.2.2 and 4.3.2 and $n = (3, 4 \text{ or } 5)$ in experiment of subsection 4.1.4. Selected objects were randomly positioned always maintaining contour intersection. Some examples of shapes in dataset are shown in Figura 4.1 (a). For the experiments, a set of non-rigidly deformed and severe-occluded shape contours was constructed in a synthetic way (Figura 4.1 (b)). Details for synthetic datasets construction are given in every case.

(a) (b)

Figure 4.1: Example of (a) shapes from MPEG-7 dataset and (b) generated synthetic occluded objects.

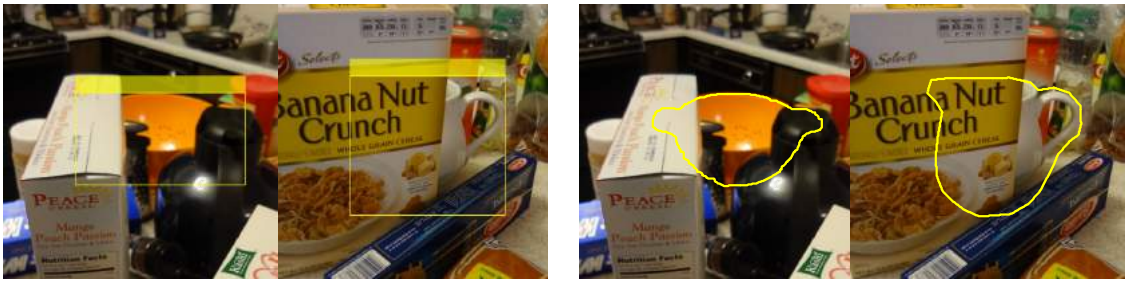
For real images experimentation was used the challenge set of daily life objects CMU_KO occluded objects dataset (HSIAO; HEBERT, 2014). Single view dataset version contains 8 classes with 100 occluded objects by class for testing and 1 non occluded object per class for training. Multiples view dataset version also contains 8 classes with 25 non occluded samples by class for training and 100 occluded objects for each class for test from different views. Objects classes examples in training and test set are shown in Figura 4.2.

(a) (b)

Figure 4.2: Example of images from CMU_KO dataset for (a) training and (b) test.

The most popular form of representing a detection hypothesis is a bounding box, which is a rectangular region enclosing the area in the image in which the detector thinks an object is present. While a rectangular is the simplest way to indicate a region of the image, it is not optimal for many applications. If the object itself is not rectangular, or if the object is partially occluded, not all the pixels in the bounding box belong to the object (BRAHMBHATT, 2014). For detection representation purpose in this work is used recognized object contour with estimated transformation. Figura 4.3 (a) shows detection representation with bounding box which is less informative than contour representation Figura 4.3 (b).

For method evaluation, individual precision and recall metrics were employed as well as F-Measure for combined observation of precision and recall. Also receiver operating characteristic (Receiver Operating Characteristic (ROC)) curves and area under curve (Area Under Curve (AUC)) were used for results over CMU_KO. True positives were considered all objects in the occlusion correctly classified. Invalid objects that met restriction measured with Equação 3.15 and were accepted as valid hypotheses were also counted as false positives (OZDEMIR et al., 2010).



(a) (b)

Figure 4.3: Examples of detection represented with (a) bounding box and (b) recognized object contour.

4.1 Occluded Object Recognition

To test proposed method, four situations were tested to evaluate performance, according to: (1) Level of object occlusion: varied amount of objects visible area in occlusion, ranging from 1% to 99%. (2) Number of object classes in the problem: varied quantity of objects classes and cardinality of training set. (3) Number of objects in the occlusion: varied quantity of objects in occlusion resulting in complex situations to analyze. (4) Inner occlusion: situation where objects appear completely in front of others.

In this experiment, CMU_KO and synthetical MPEG-7 based datasets were used.

4.1.1 Parameter Selection

For best parameter (τ_1, τ_2, τ_3) estimation, values were varied from 0.0 to 0.9 with steps of 0.15 (all possible combinations). Each combination was analyzed in precision-recall space. For experimentation, a synthetic dataset of 100 isolated objects from 5 different classes was used for training and 300 occluded objects were employed for testing. Figure 4.4 shows obtained results, with every point referring to certain parameter values combination. Influence of τ_1, τ_2 and τ_3 values could be observed in Figure 4.4 a), b), and c). F-Measure of each combination is shown in Figure 4.4 d). It is seen higher F-Measure values associated with higher values of τ_1 and τ_2, and τ_3 close to 0.85. From this experiment, the best values for the synthetic dataset were set to $\tau_1 = 0.85$, $\tau_2 = 0.9$ and $\tau_3 = 0.85$, also repeated in further experiments.

4.1.2 Different amount of occlusion

The first experiment was conducted to measure the influence of the level of occlusion of the objects in the proposed method. For this experiments were selected 10 pairs of objects of the dataset. The objects belong to one of five selected classes and the HMM were trained with all objects of the 5 classes. For each pair, one object was put in a fixed position and the other was moved from right to left in steps of 10 pixels varying the amount of occluded area. As results, multiples occlusions were generated for each pair of objects. Illustration of generated

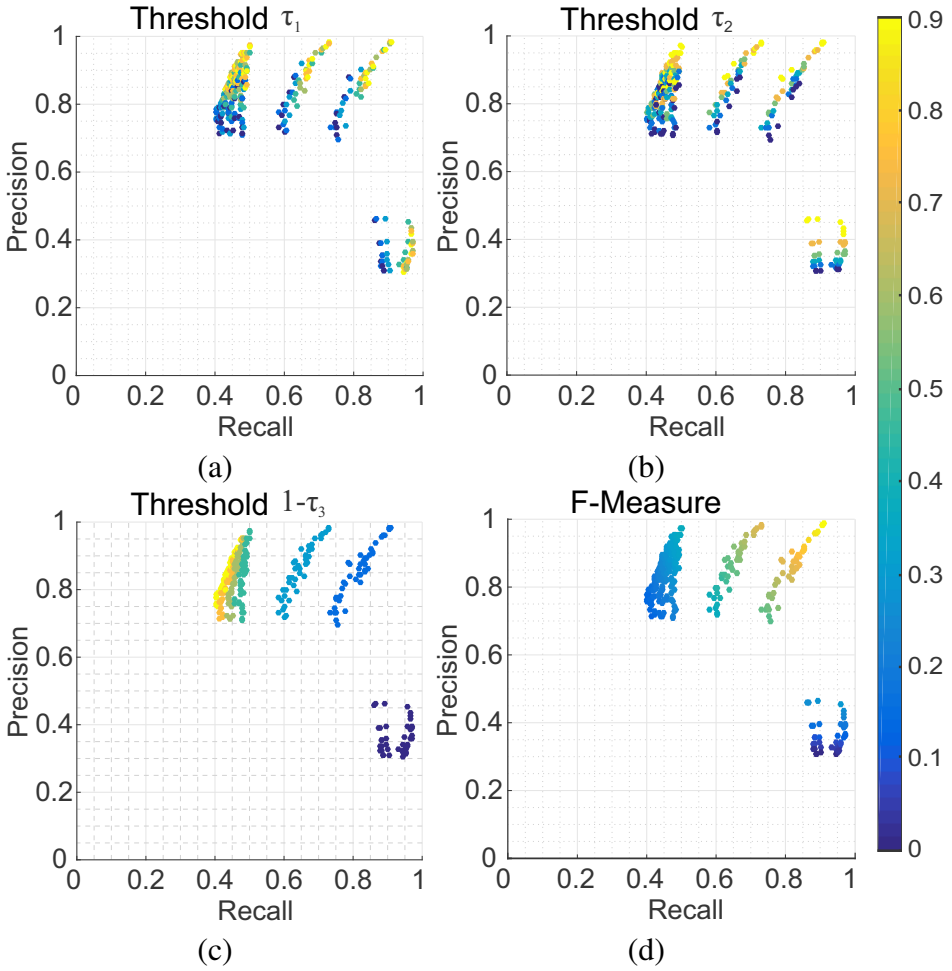


Figure 4.4: Results in precision-recall space with different values of τ_1 a), τ_2 b), $1-\tau_3$ c) and F-Measures d).

occlusions for two pair of objects is shown in Figura 4.5. A total of 300 occlusions of 2 objects were analyzed in this experiment.

The obtained F-Measure for every amount of occlusion is shown in Figura 4.6. The increment of the amount of occlusion leads to a diminution of the F-Measure mostly because the breach of the second restriction that dictates the valid amount of occluded area. The relaxation of the thresholds associated with this restriction leads to the increment of False Positive. No False Positive were recognized in this experiment due the simplicity of the generated occlusions. It can be observed the stability of the method that keeps high F-Measure, even with high level of occlusions where at least one of the two objects were recognized. For the amount of occlusion between 90% and 100% the mean F-Measure was 0.67. The results show that even when the level of occlusion is up to 80% the F-Measure is greater than 0.95. This experiment proves the viability of the proposed method.

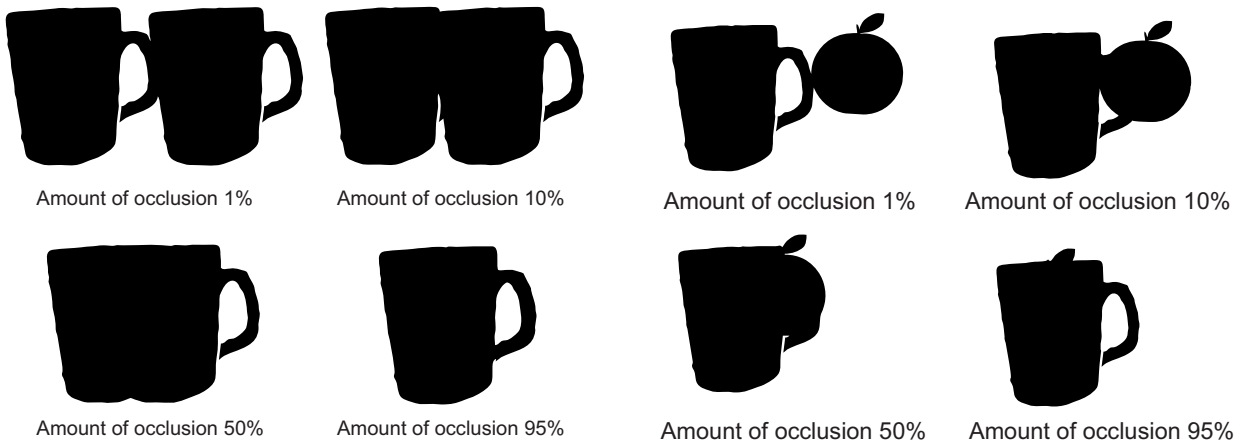


Figure 4.5: Examples of occlusions generated for two pair of objects.

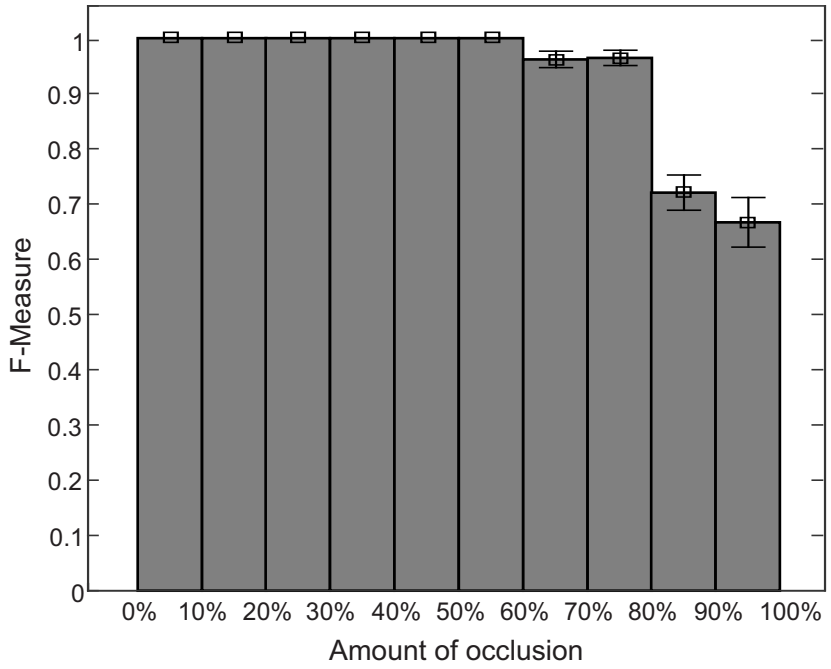


Figure 4.6: F-Measure of the proposed method under different amount of occlusion.

Table 4.1: Results of the experiment with different number of classes

Classes	Total Objects	True Positive	False Positive	Precision	Recall	F-Measure
5	600	552	1	0.99	0.92	0.96
16	600	542	4	0.99	0.90	0.95
30	600	534	10	0.98	0.89	0.93

4.1.3 Different number of classes

In this experiment was tested the behavior of the proposed method when different classes of objects are available. For this experiment were selected 3 amount of classes (5, 16, 30) with 20 objects every of each. For each amount of classes, 300 occlusions were generated. Each occlusion was composed by two randomly chosen objects. The HMMs were trained with the objects of the 5, 16 and 30 classes respectively.

The performance of the proposed method under this conditions can be observed in Tabela 4.1. While more classes of objects, more confusion will exist between parts of objects of different classes and consequently a decrease in the recall of the shape retrieval process. The False Positive increments with the number of classes because the amount of wrong classified parts it also increases. It can see that the increment of the number of classes has a low decrease in the True Positive objects. The F-Measure of the 30 class problem it is 0.93 that is very promising result in the occlusion problem.

An illustration of the results obtained is shown in the Figura 4.7. The examples of a), b) and c) corresponds to results obtained with 5 classes. Similarly, the examples d), e) and f) corresponds to 16 classes and g), h) and i) to 30 classes. The examples a), b), d), e), g) and h) present well-recognized objects. Each of this images has the binary occlusion presented to the method and the result of the recognition. Could be observed in the example e) that the glass shape is divided into two different parts because the head shape. This leads to found 2 hypotheses for the glass. Besides the exact object was not matched, in terms of object recognition the correct class of object was recognized. The Figura 4.7 c), f) and i) shows an example of wrong recognized objects. This has the 3 main problems that occur in this experiment; the non-recognized of HCP that leads to wrong part separation; the False Positive recognition due to the non-discriminative parts wrong classification; and the non-object recognition because the wrong classification of the parts.

4.1.4 Different number of objects

In order to measure the performance of the proposed method with more than 2 objects were generated a total of 300 occlusions with 3, 4 and 5 objects. The number of classes was set to 5 and 20 objects every of each. For the occlusion generations, the objects were selected from the five classes in a randomly way. The ensemble of HMMs was trained with all objects of the selected classes. The generated occlusions have no restrictions on the amount of occlusion

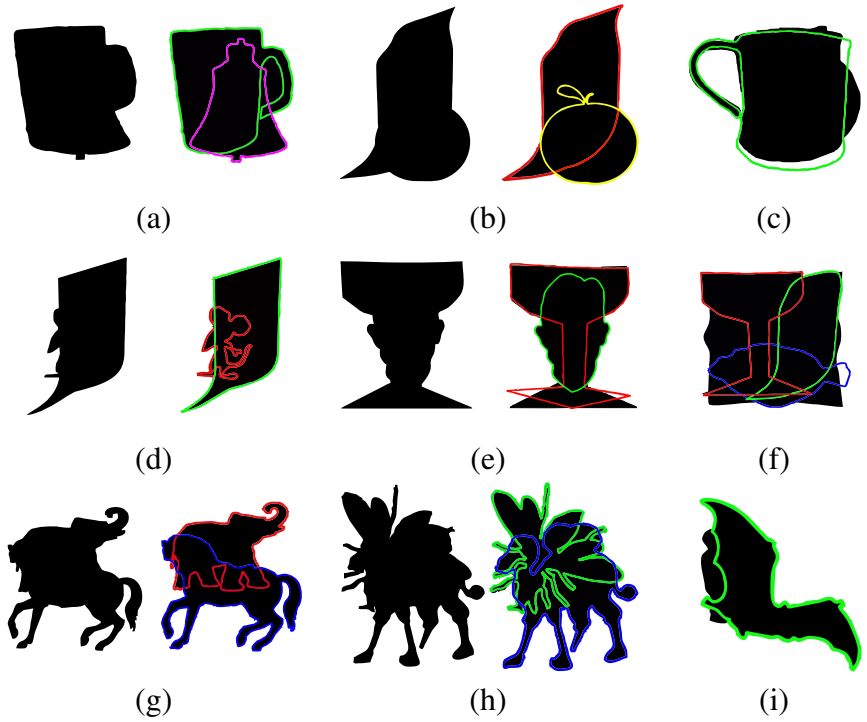


Figure 4.7: Results of the proposed method with different amount of classes. Examples of correct recognition a), b), d), e), g) and h) and examples of incorrect recognition c), f) and i)

Table 4.2: Results of the experiment with different number of objects

Objects	Total Objects	True Positive	False Positive	Precision	Recall	F-Measure
3	300	276	4	0.99	0.92	0.95
4	400	347	11	0.97	0.87	0.92
5	500	438	12	0.97	0.87	0.92

between the objects, this way some objects of the occlusions have a non-discriminative part.

The results of the third experiment are shown in Tabela 4.2. The increase in the number of objects leads to more complex occlusions. This means that the occlusion could have objects only represented by one non-discriminative part. In the previous experiments, the most of the objects were represented for more than one part and the failure in the retrieval of less discriminative parts was not so significant because other parts of the same object not fail the retrieval. In the case of the occlusion with five objects the False Positive has a significant increment because the larger area occlusions allow the acceptance of wrong hypotheses as valid. In addition, other valid objects were rejected because the generated dataset in this experiment do not restrict the amount of occlusion and do not meet the second restriction.

An illustration of the results obtained is shown in the Figura 4.8. The examples of a), b) and c) corresponds to results obtained with 3 objects, examples d), e) and f) corresponds to 4 objects and g), h) and i) to 5 objects. Like in the Figura 4.7 the two first column has the occlusion presented to the method and the recognition result. The Figura 4.8 c), f) and i) present example of wrong recognized objects. The most common problems were the non-HCP recognition; very

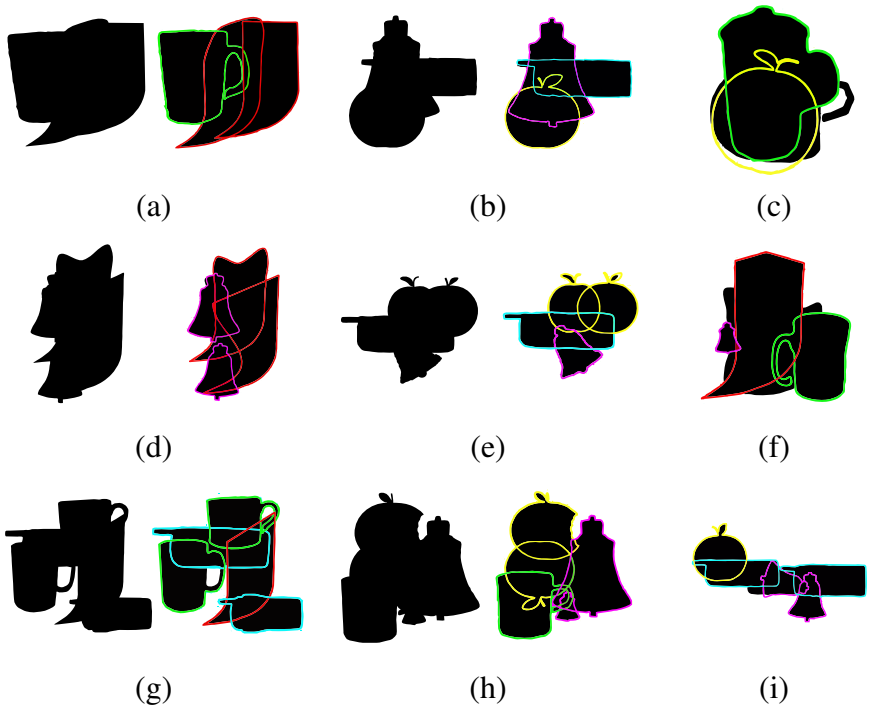


Figure 4.8: Results of the proposed method with different amount of objects. Examples of correct recognition a), b), d), e), g) and h) and examples of incorrect recognition c), f) and i)

occluded objects; and wrong hypotheses estimation due to the non-discriminative part wrong classification. The F-Measure values obtained are considered as a promising result because the complexity of the generated occlusions.

4.1.5 Inner occluded objects

Other experiments were performed to evaluate situations where objects appear completely in front of others (Figure 4.9). A dataset was built with 100 inner occluded objects with 5 classes from MPEG-7 dataset and 20 objects per class. For occlusion generation, 2 objects were randomly drawn out of the 5 classes and an ensemble of HMMs was trained with all objects from selected classes. The 2 selected objects were positioned, rotated and scaled randomly, making sure image of one object was "inside" the other (with no contour intersection). Created images were compounded of 2 objects with different intensities (Figure 4.9). Since input for the method is a closed contour, a previous step for selecting regions of interest was carried out. For this, an intensity threshold for contour selection was applied, obtaining all closed shapes for classification. The method was applied and, as a result, all objects in the dataset were correctly recognized with no false positives and F-Measure of 1.0.

4.1.6 CMU_KO Real Image Dataset

For further testing, images from the challenge 2012 CMU_KO database were considered. Recognition performances were measured with parameter values $\tau_1 = 0.85$, $\tau_2 = 0.98$ and

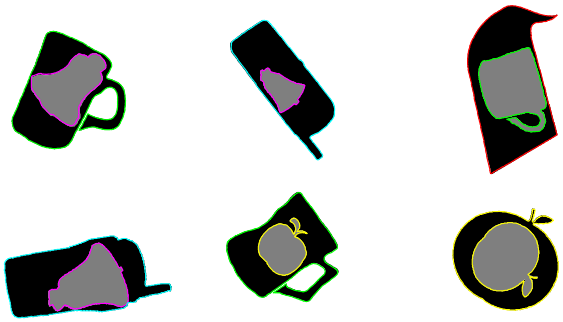


Figure 4.9: Examples of object recognition with inner occluded objects.

$\tau_3 = 0.95$. Each isolated object for all 8 classes of CMU_KO's single view database were employed for training. For testing, a total of 800 objects from the test set were selected with 100 objects per class. Since input to the method is a closed contour of occluded shapes, segmented images are required. For this, the existing segmentation ground truth images in the dataset were employed. Receiver operating characteristic curves (ROC) computed for each class are shown in Figure 4.10. High ROC rates were obtained with areas under the curve (Area Under ROC (AUROC)) close to or above 0.9, for most cases, with FP rates up to 0.2. The shaker class presented the worst performance since has less discriminant parts and recognition fails when more discriminant parts are occluded. Some examples of occluded object recognition with CMU_KO dataset are shown in Figure 4.11.

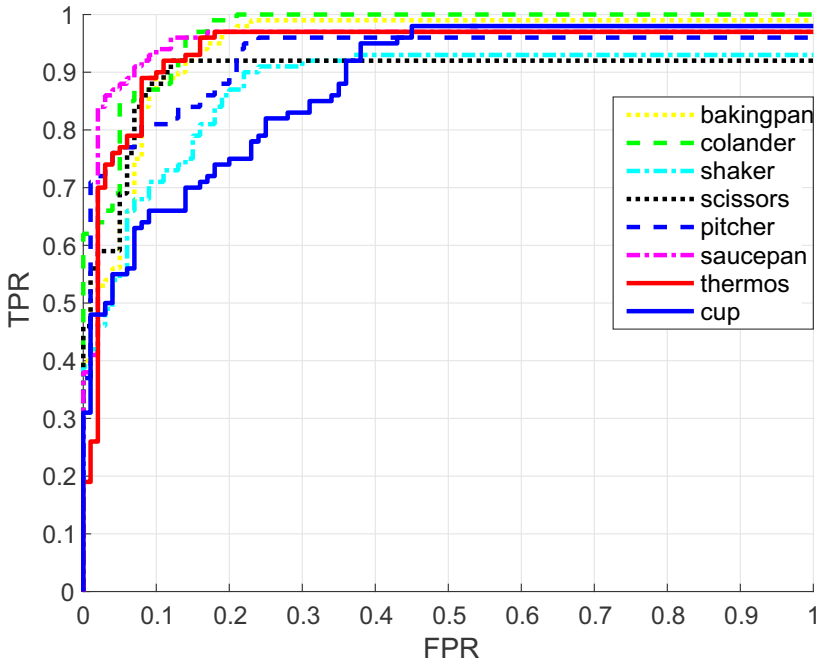


Figure 4.10: ROC curves for every class of CMU_KO dataset.

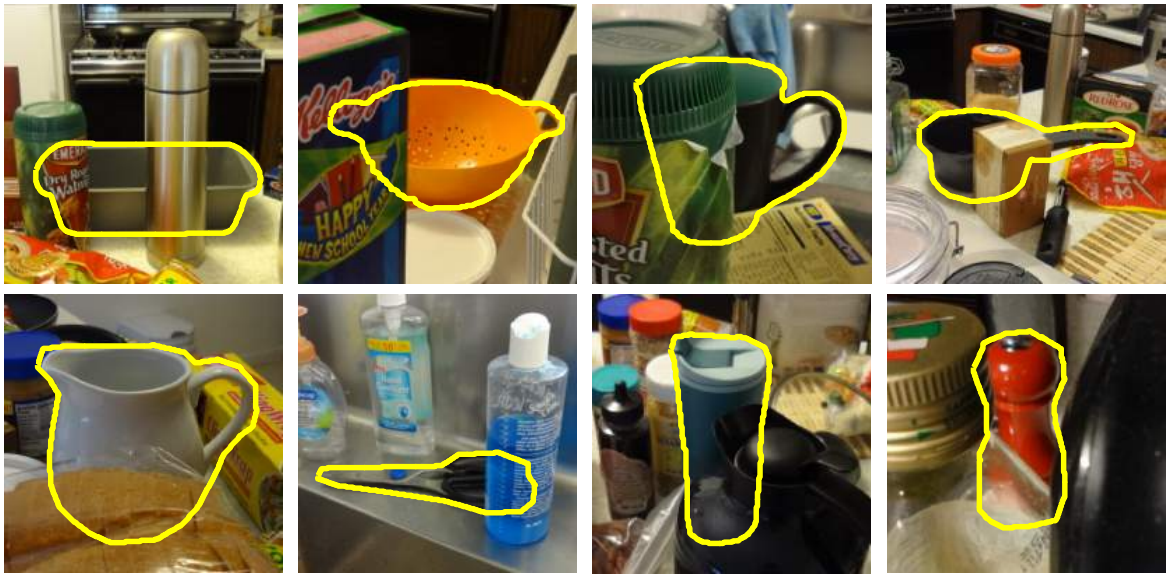


Figure 4.11: Severe occluded object recognition with CMU_KO dataset.

4.1.7 Comparison

The methods proposed in (LATECKI et al., 2007) and (MERHY et al., 2014) only solves the shape retrieval problem based in a query part of the contour. This task is robust in the presence of occlusions because only a part of the contour is necessary for the most similar object recognition but for this type of methods all the points of the query part has to ensure that belongs to the same object. This means that a previous step for the parts identification has to be taken. To separate the objects in parts some works divide by the Zero Curvature Points (HORÁČEK; KAMENICKÝ; FLUSSER, 2008), High Curvature Points (MARVANIYA; GUPTA; MITTAL, 2015), (ZHANG; XU; LIU, 2015) and the shape retrieval process taking the recognized parts as queries is made. This approach can manage the partially occluded object recognition task where part of the objects are occluded and the contour points that belongs to each object meets that are consecutive. In this type of problems, no hypotheses validation step is needed because the most similar object of the retrieval is the only hypothesis for the object (SABER; XU; Murat Tekalp, 2005), (CHEN; FERIS; TURK, 2008), (CUI et al., 2009) ,(MICHEL; OIKONOMIDIS; ARGYROS, 2011). In order to make a severely occluded object recognition, an exhaustive point matching method is developed in (CAO et al., 2011) and no retrieval process is made, implying that the object to be recognized has to be giving to the method. The method proposed in this work differs from all above methods that the severely occluded object recognition task is made without information of the quantity of objects and the class of the object. Also has the ability to recognize severely occluded objects where the contour points of the occluded object are not consecutive, this means that the objects could be described by more that one non-adjacent part, Figura 4.7 e) in a non-exhaustive way.

The most important step to obtain good results in the proposed method is the retrieval process. In order to quantitatively compare the retrieval with other methods of the literature were

Table 4.3: Comparison of the proposed retrieval with SDIM and MVM

	Queries of (LATECKI et al., 2007)
MVM	24.0%
SDIM	52.5%
Proposed method	50.5%

measured the Bull's eye score (MICHEL; OIKONOMIDIS; ARGYROS, 2011). This metric measures the retrieval accuracy of the system counting the number of correct retrievals among the top $2R$ retrievals, where R is the number of shapes which are relevant to the test image in the database, in this case, 20. The metric is normalize by the maximum number of correct shapes $(10x20)$. In this experiment was used the 10 queries open contours proposed in (LATECKI et al., 2007) and were selected and matched with each shape in the MPEG7 dataset. This experiment setup is proposed in MICHEL; OIKONOMIDIS; ARGYROS (2011).

The Tabela 4.3 shows the obtained results for the Bull's eye score for the Minimum Variance Matching (MVM) (LATECKI et al., 2007), Scale Invariant and Deformation tolerant partial shape Matching (SDIM) (MICHEL; OIKONOMIDIS; ARGYROS, 2011) and the proposed method. The results respect to MVM is highly improved by SDIM and the proposed method as can be observed. Also, the retrieval method in this work has comparable performance with the SDIM, that was designed for deformation tolerant partial shape matching. The problem of using this type of retrieval in a severely occluded object recognition task is that the deformations could lead to a hard-to-prove hypotheses set.

4.2 Search Space Sorting with Hidden Markov Models

To measure influence of sorting the search space (use of HMM) in recognition performance, a first experiment was conducted on a synthetic occluded object dataset using MPEG-7 database. The dataset contained 300 binary objects distributed in 5 different classes, each class with 20 objects. Individual objects were used for training and test images with 60% to 90% occlusions were employed for testing. Then, for testing with real-image objects, a second experiment was carried out in CMU_KO database multiples views with 3 first classes, 25 samples of objects per class for training and 100 images of varied occluded objects for testing.

4.2.1 Sorting search space retrievals in synthetic dataset

The first experiment observed how sorting the search space affected recognition performance in terms of precision, recall, F-measure and rate of objects visited before reaching stopping criterion. The generated synthetic occluded dataset was used with length of search space set to 100 training objects and 300 diverse occluded test objects. The three method parameters τ_1, τ_2 and τ_3, described in Seção 3.3, were varied from 0.0 to 0.9 with steps of 0.15 (all possible combinations), with the stopping rule threshold τ_1 seen as the particular parameter

of interest, in this case. For that, parameters τ_2 and τ_3 were set to 0.85, and precision and recall rates for different values of stopping thresholds τ_1 were computed. Results are seen in Tabela 4.4 and Tabela 4.5. The stopping rule was applied in the retrieval process in two different configurations, with and without sorting. For object representation, the two object features TAS and Curvature (CV), described in Subseção 2.2.2, were considered for analysis.

Table 4.4: Precision for different values of τ_1 with $\tau_2 = 0.15$ and $\tau_3 = 0.15$

τ_1	TAS+HMM	TAS	CV+HMM	CV
1.00	0.9855	0.9855	0.9929	0.9947
0.85	0.9837	0.9910	0.9929	0.9911
0.70	0.9819	0.9749	0.9876	0.9856
0.55	0.9603	0.9107	0.9802	0.9481
0.40	0.9044	0.6144	0.9519	0.8027
0.25	0.8764	0.3635	0.8967	0.3632
0.00	0.8792	0.3162	0.8911	0.3234

Table 4.5: Recall for different values of τ_1 with $\tau_2 = 0.15$ and $\tau_3 = 0.15$

τ_1	TAS+HMM	TAS	CV+HMM	CV
1.00	0.9083	0.9083	0.9300	0.9350
0.85	0.9050	0.9133	0.9267	0.9300
0.70	0.9067	0.9050	0.9267	0.9150
0.55	0.8867	0.8333	0.9067	0.8517
0.40	0.8200	0.5417	0.8583	0.6917
0.25	0.7917	0.3417	0.7667	0.3450
0.00	0.7883	0.3083	0.7500	0.3083

Tabela 4.4 and Tabela 4.5 shows sorting the SSR produced higher or equal precision and recall rates than those of the unsorted version for all values of τ_1, except the two first (1.00 and 0.85). The most noticeable difference was with $\tau_1 = 0.0$, the case where no minimum similarity was required for object recognition. The difference in precision between the sorted and unsorted versions were 0.56 with TAS and 0.57 with CV (last row in Tabela 4.4), and in recall were 0.48 with TAS and 0.44 with CV (last row in Tabela 4.5), respectively. Complete results in the precision-recall space for all different configurations are shown in Figura 4.12. From Figura 4.12, it can be observed sorting the SSR produced more condensed regions with high performances than those with the unsorted version. The best performance for all configurations in Figura 4.12 was obtained with $\tau_1 = 1.0$, as expected, because the non-stopping rule guarantees searching all solution space.

Figure 4.13(a) shows the relation between τ_1 and the percentage of SSR visited with both representations (TAS and CV). As mentioned before, with $\tau_1 = 1.0$ all SSR is tracked, with precision around 0.98 with TAS and 0.99 with CV. However, with $\tau_1 = 0.0$ (only 20% of all objects visited), the HMM approach reached precision and recall performances around 0.88 (TAS) and 0.89 (CV), respectivelly, while with typical no sorting perfomances dropped

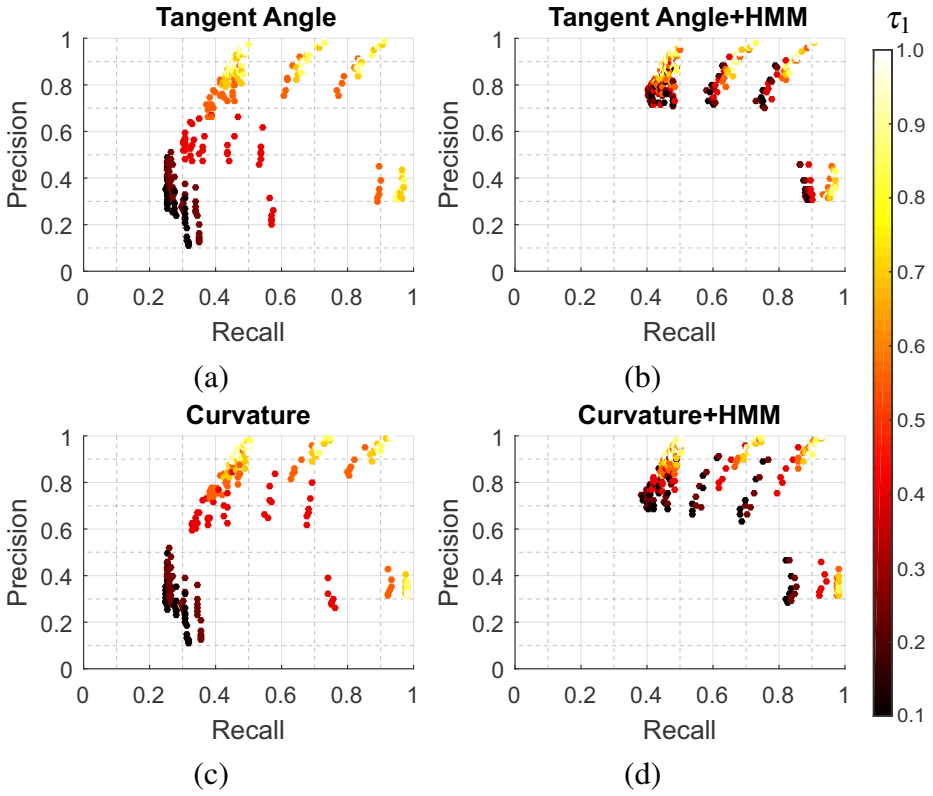


Figure 4.12: Results in precision-recall space with different values $of\,\tau_1$, τ_2 and τ_3 for TAS without sorting SSR (a), TAS with sorting SSR (b), CV without sorting SSR (c) and CV with sorting SSR (d).

to 0.32 (both with TAS and CV). Mean F-Measure values for every percentage of SSR visited with TAS and CV are shown in Figure 4.13 (b) and (c), respectively. With HMMs, F-Measure values of 0.83 and 0.81, were observed respectively, for TAS and CV with "quick stopping" thresholds ($\tau_1 = 0.0$). Without sorting, however, the highest F-Measure in each case was 0.33 and 0.34, respectively. For further analysis, significative statistical difference between the sorted and unsorted versions was tested with Friedman test with significance of 0.05. For values of τ_1 between 0.0 and 0.6 there is significative difference with p-values of 0.0143. P-values greater than 0.1025 were obtained with $\tau_1 > 0.6$.

According to Friedman test, significative difference in results ocurred with TAS+HMM F-Measures of Figura 4.13(b) depending on values of τ_1, obtaining p-value of $5.05 * 10^{-6}$.

The Nemenyi pos test revealed there was no-significative difference in F-Measure values from 100% to 40% percents of SSR visited. The same analysis was made with CV representation rejecting Friedman's null hypotheses with p-value $3.56 * 10^{-6}$ and no-significative difference in F-Measures was found for percentages above 40%. This result shows even with 60% search space reduction, the performance was not meaningfully different than that of visiting the hole lot of objects in the database. Also, SSR sorting with HMMs produced consistent high F-Measure performance even when visiting only 20% of candidate objects, in a highly occluded object recognition scenario and with a relevant number of retrieved objects being processed. No

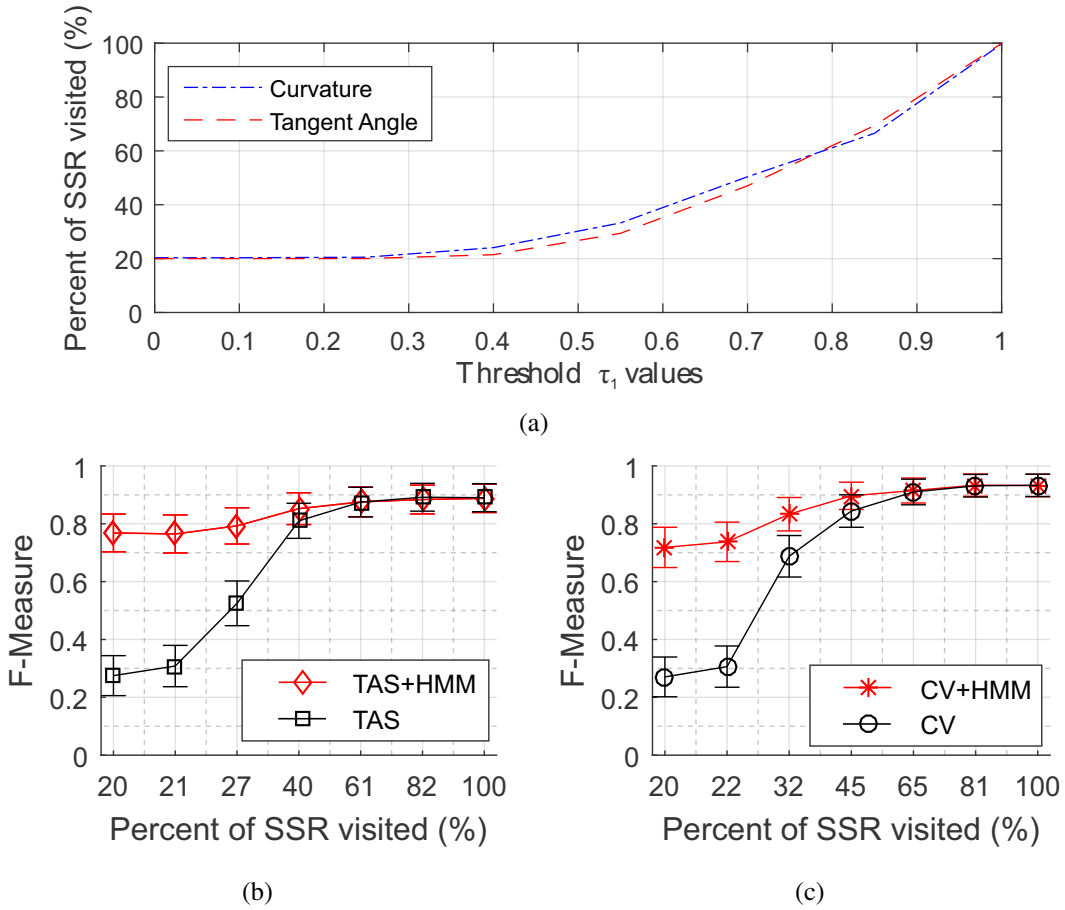


Figure 4.13: Percentage of SSR visited for each value of τ_1 (a) and F-Measure with distinct percentage of SSR visited with and without sorting. (b) for TAS and (c) for CV.

significative difference was observed between TAS and CV. Therefore, lesser time consuming representation was used in experimentation from here, corresponding with Tangent Angle Signature.

4.2.2 Sorting search space retrievals in CMU_KO dataset

For testing real images CMU_KO dataset was experimented. Recognition performances were measured with and without sorting and diffetent values of stopping parameter (τ_1). 75 isolated objects from first three object classes of CMU_KO database, with multiples views, were employed for training. For testing, approximately the same number of objects from the test set of 3 classes were randomly selected with a total of 100 objects. First, segmentation was carried out using the approach described in Subseção 2.2.1 and TAS representation was used for experimentation. Occlusion likelihood and constraint thresholds values were fixed to $\tau_2 = 0.99$ and $\tau_3 = 0.97$ after several trials. Then, four different stopping thresholds were tested, $\tau_1 = \{0.1, 0.5, 0.9, 1\}$, both for sorting and unsorting SSR configurations. Obtained receiver operating characteristic (ROC) curves for each configuration are shown in Figure 4.14.

From Figure 4.14, it can be seen the smaller the value of τ_1 the lesser accurate is the

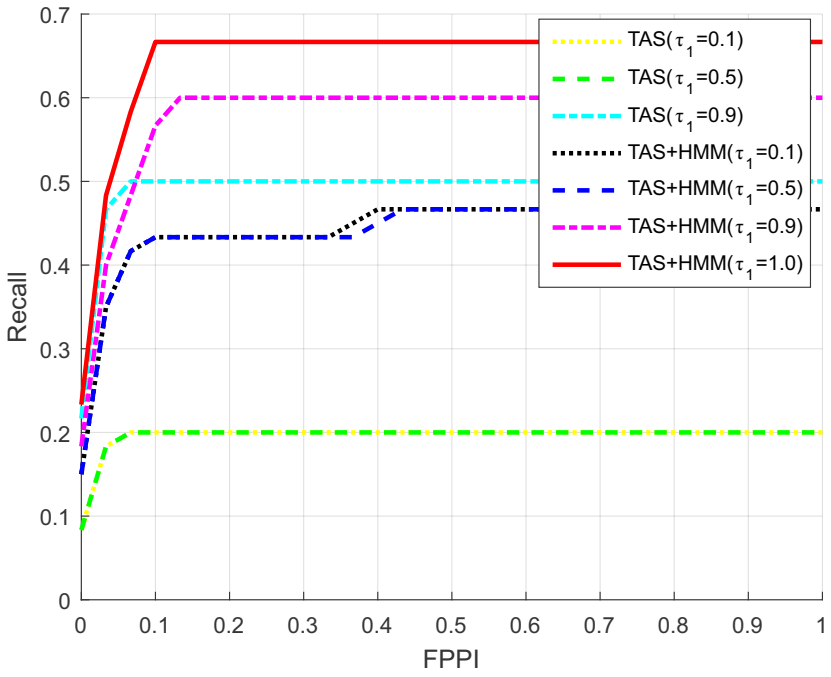


Figure 4.14: ROC curves with four different values of τ_1 using TAS and TAS+HMM.

recognition task as expected. However, the major difference between the sorted and unsorted SSRs is with the quick stopping threshold ($\tau_1 = 0.1$) with AUROC measures of 0.50 and 0.22, respectively. The overall attained AUROC with the non-stopping threshold was 0.70 and the AUROC with $\tau_1 = 0.90$ was 0.63 for TAS+HMM and 0.54 for TAS (SSR without sorting).

Some examples of object recognition with unsorted and sorted SSRs are shown in Figure 4.15. The HMM maximum a posteriori (Maximum A Posteriori (MAP)) classification for each part is showed in Figure 4.15(b) for the contour parts corresponding to learned objects (bakingpan, colander and cup). Classification was correct in all cases, indicating the search began from the right class. With low value of τ_1 (0.1), searching with no sorting was unable to recognize the objects (f, column 2, 3 and 4), whereas TAS+HMM correctly recognized the objects even with τ_1=0.1 (e, column 1, 2, 3 and 4). With a high value of τ_1 (0.9), TAS+HMM also correctly recognized all objects but the unsorted version could not recognized the cup (d, column 4) even in this case $\tau_1 = 0.9$.

4.3 Wavelet Filtering

To measure influence of wavelet filtering (second approach of Seção 3.2) in recognition performance, an experiment was conducted on a synthetic occluded object dataset using the MPEG-7 database. The generated dataset contained 300 occluded objects distributed in 5 different classes, each class with 20 objects. Individual objects were used for training and test images with 60% to 90% occlusions were employed for testing. Then, for testing with real-image objects, another experiment was carried out with the CMU_KO database with 8 classes and 800

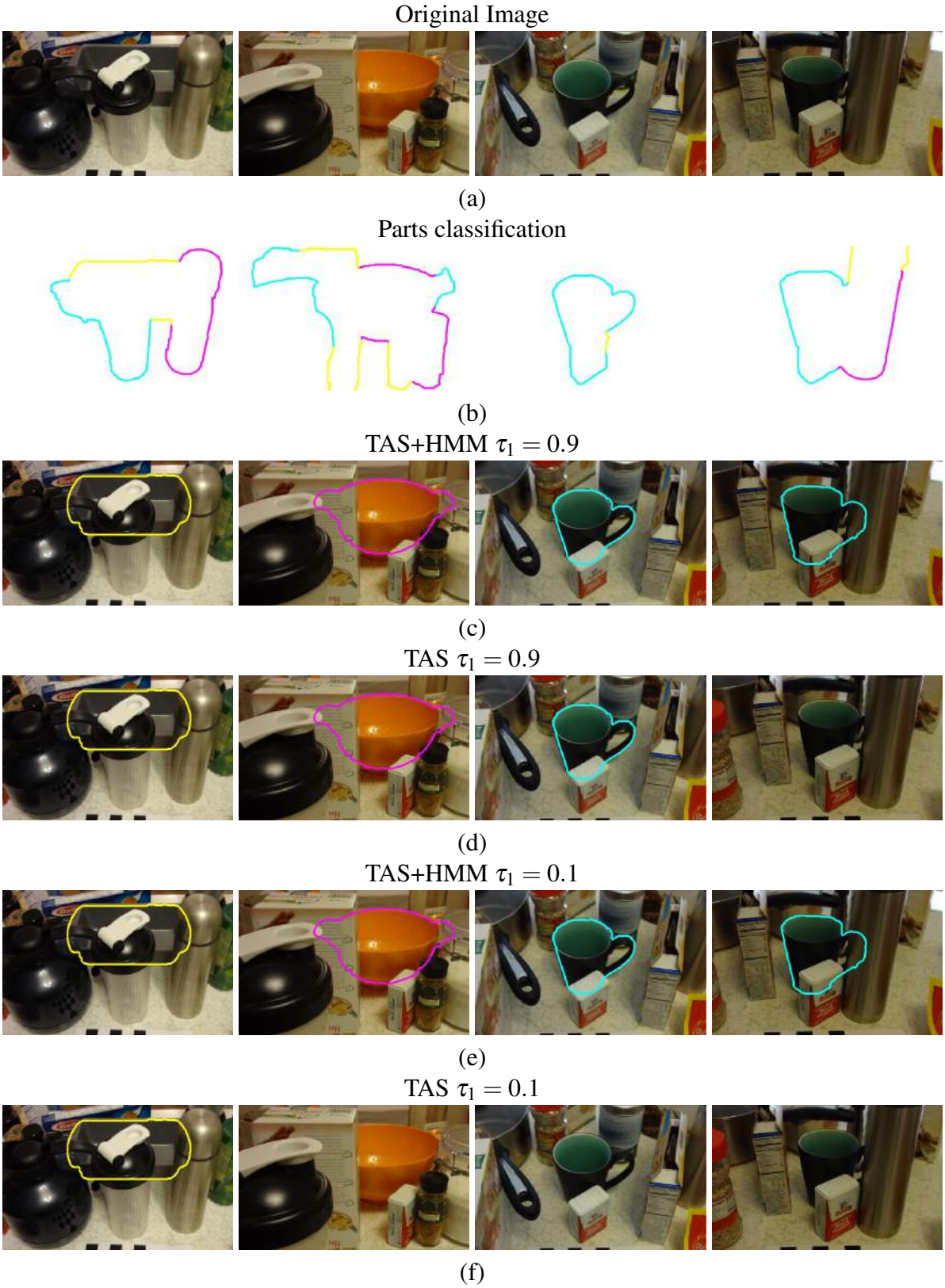
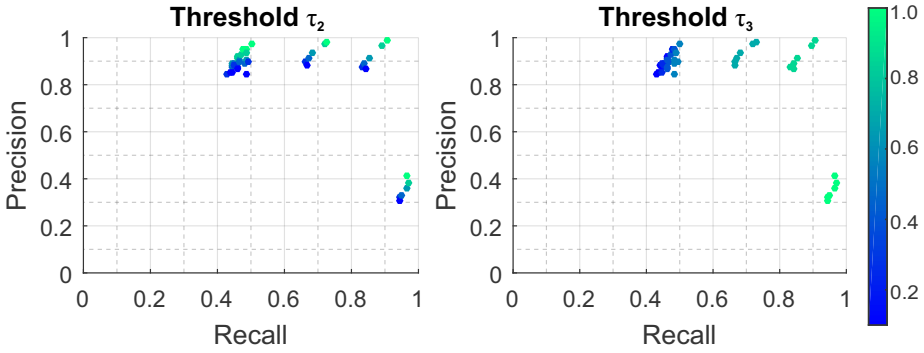


Figure 4.15: Results with (a) CMU_KO images; (b) HMM classification of baking pan (1st column), colander (2nd column) and cup (3rd and 4th column); (c) different stopping thresholds when sorting (c) and (e), and unsorting (d) and (f), the SSR.

images of varied occluded objects for testing. Unsorted version of SSR, this means no HMM search space sorting, was used in experimentation in order to remove sorting influence for noise removal.

4.3.1 Parameters selection

For best parameter selection, values of τ_2 and τ_3 were varied from 0.0 to 0.9 with steps of 0.15 (all possible combinations). Each combination was analyzed in precision-recall space. For experimentation, a synthetic dataset of 100 isolated objects from 5 different classes was used for training and 300 occluded objects without noise were employed for testing. Figura 4.16 shows obtained results, with every point referring to certain parameter values combination. Influence of τ_2 and τ_3 values could be observed in Figura 4.16 (a) and (b). It is seen higher F-Measure values (about 0.95) associated with higher values of τ_2 and values of τ_3 close to 0.85. From this experiment, the best values for the synthetic dataset were set to $\tau_2 = 0.9$ and $\tau_3 = 0.85$, also repeated in further experiments.



(a) (b)

Figure 4.16: Results in precision-recall space with different values of (a) τ_2 and (b) τ_3.

Once hypotheses estimation parameters τ_2 and τ_3 were set, influence of τ_0 threshold was analysed. Same testing set of 300 occluded object was used, this time with addition of Gaussian noise varying from $\sigma = 0.5$ to $\sigma = 1.0$ and $\sigma = 1.5$. Value of scale parameter τ_0 was varied from 1 to 300 with steps of 10. Figura 4.17 shows obtained F-Measure with every value. As expected, while τ_0 grows more relevant features are removed from TAS signature. Higher F-Measure values was 0.85 for $\tau_0 = 1$ and $\tau_0 = 10$ in added $\sigma = 0.5$ noise. With values of τ_0 up to 50, F-Measure values were above 0.8 for lesser quantity of noise and above 0.5 for higher gaussian noise amount.

4.3.2 Wavelet filtering in synthetic dataset

In this experiment was observed how wavelet filtering affected recognition performance in terms of F-measure. The generated synthetic occluded dataset was used with length of search space set to 100 training objects and 300 diverse occluded test objects. Each occlusion was composed by two random selected objects.

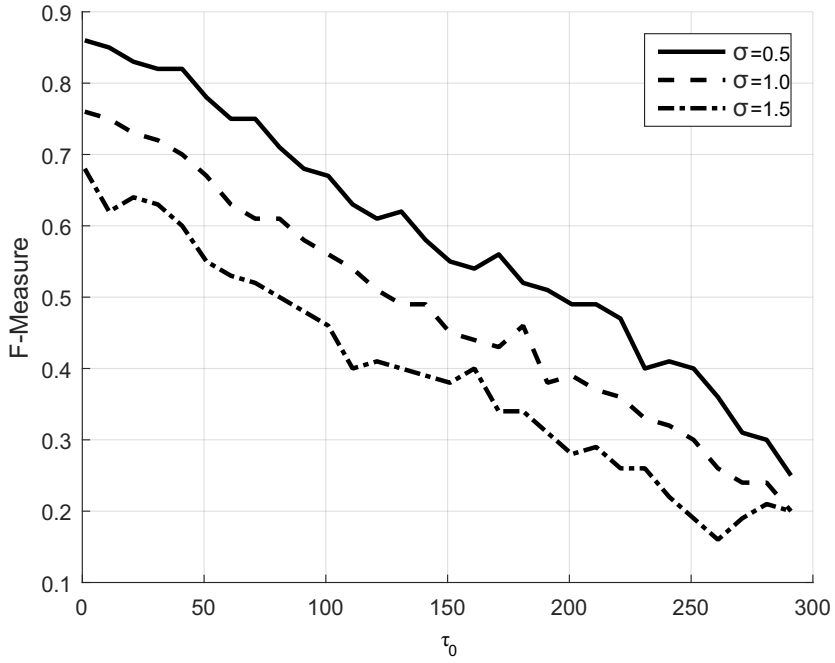


Figure 4.17: F-Measure results with different scale parameter τ_0 variation.

Tabela 4.6 shows F-Measure results with the generated synthetic dataset where UNFILT refers to unfiltered TAS representation. USMP5 and USMP2 correspond to uniform sampling methods (BELONGIE; MALIK; PUZICHA, 2002) with steps 5 and 2, respectively, and NUSMP corresponds to non-uniform sampling (MICHEL; OIKONOMIDIS; ARGYROS, 2011). TAS+WAV refers to using wavelet coefficients of TAS for retrieval (first approach of Seção 3.2). WAV10 method is the investigated wavelet filtering where $\tau_0 = 10$.

Method	No-noise	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 1.5$
NUSMP	0.08	0.06	0.06	0.05
USMP5	0.19	0.14	0.13	0.12
USMP2	0.63	0.46	0.41	0.38
UNFILT	0.90	0.74	0.64	0.50
TAS+WAV	**0.96**	0.64	0.49	0.30
WAV10	0.94	**0.84**	**0.76**	**0.65**

Table 4.6: F-Measure results for occluded object recognition without noise and three different amount of Gaussian noise.

As result of experimentation, it was observed small concave parts to be the most discriminant parts for correct object retrieval. It explains, therefore, why NUSMP method has the worse performance since this type of sampling keeps noise in discriminant parts and remove it from less important parts. The bigger the step in uniform sampling the more relevant contour features were removed from TAS. USMP2 sampling achieved better results than USMP5 but a worse performance than the unfiltered version, because small concave parts are described with too many general descriptors. Also, small scaled objects have larger sampling step size and

less accuracy. The wavelet filtering approach, on the other hand, achieved the best attainable performance, with a consistent performance increase of 10% or more, even in the presence of noise. In no-noise case there is small difference between TAS+WAV and WAV10 filtering and in noise case this difference is higher in favor to WAV10 method. In real images dataset the two best method in case of noise are compared corresponding with WAV10 and UNFILT.

4.3.3 Wavelet filtering in CMU_KO dataset

For testing with real images CMU_KO was also experimented. Recognition performances were measured with and without filtering. Segmentation result involve contours with certain noise. Additionally, different amount of Gaussian noise ($\sigma = 0.5$, and $\sigma = 1.0$) were added to occlusion contours. One isolated object of each class from the CMU_KO database, with single views, were employed for training. For testing, 800 objects from the test set were selected. Results of receiver operating characteristic (ROC) curves are shown in Figura 4.18.

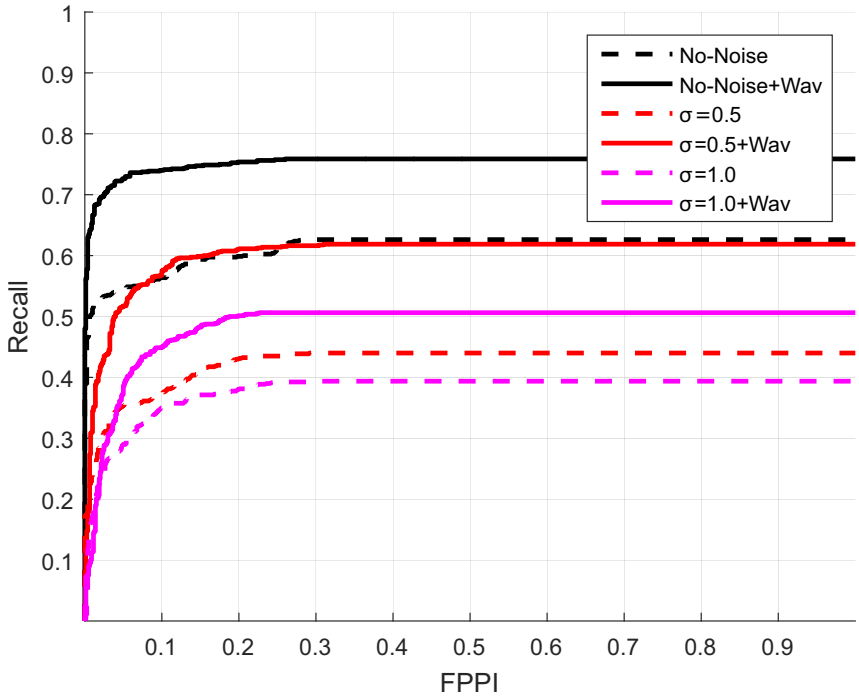
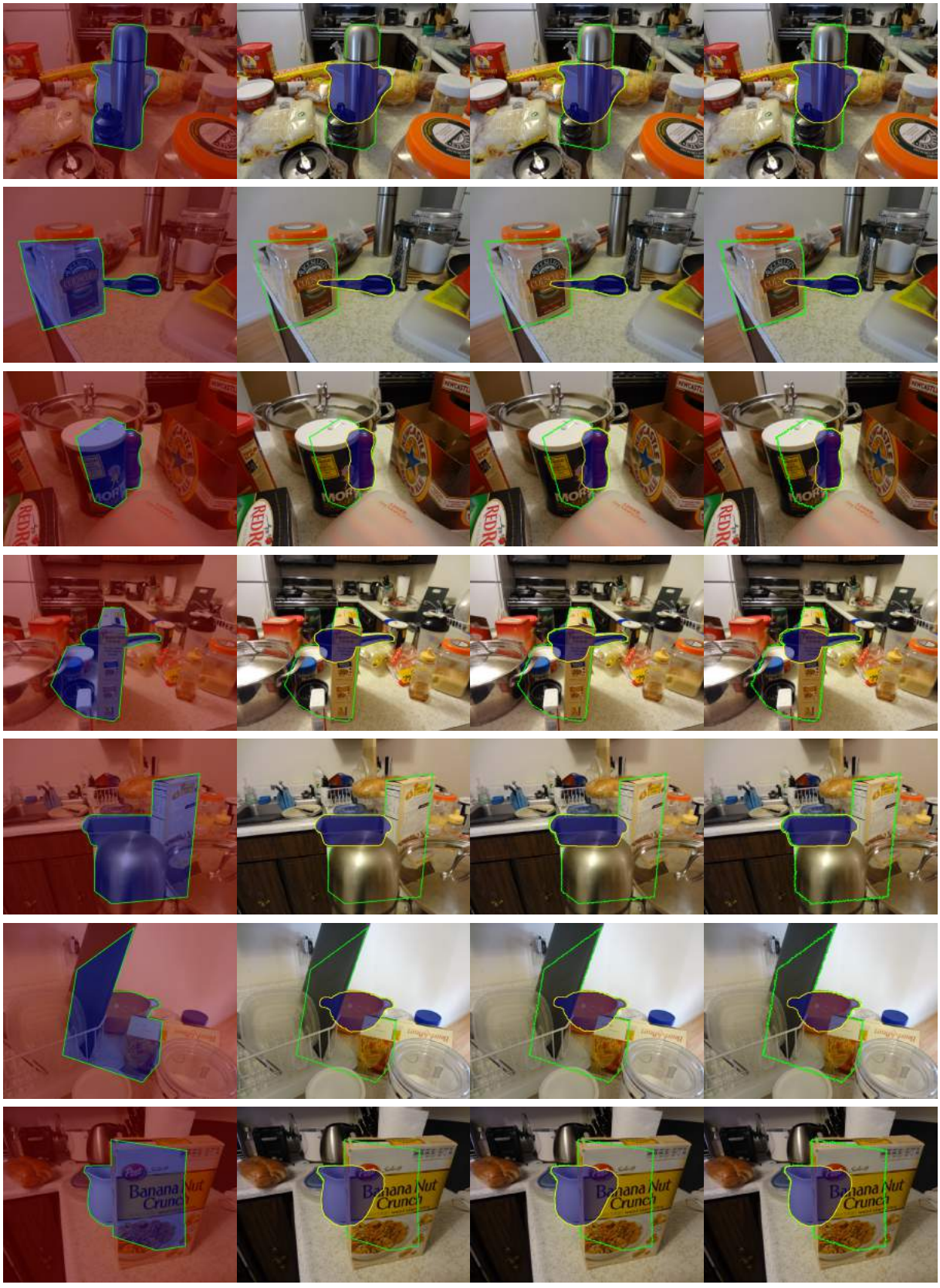


Figure 4.18: Performance of ground truth with different amount of Gaussian noise.

From Figura 4.18, it can be seen the greater the value of σ the lesser accurate was classification, as expected. However, a major difference between the filtered and unfiltered signal models could be observed with AUROC measures of 0.61 and 0.44, for added Gaussian noise with $\sigma = 0.5$, respectively. The overall attained AUROC rates without noise and non-filtering was 0.61 and the AUROC rate with wavelet noise reduction was 0.74. It could be observed that wavelet filtering for $\sigma = 0.5$ had similar performance to segmented contour without added noise and filtering, demonstrating high improvement with wavelet uses. Some examples of object recognition with different amount of noise are shown in Figura 4.19.



(a) (b) (c) (d)

Figure 4.19: Results over CMU_KO images with wavelet filtering in (a) obtained contour; added gaussian noise with (b)$\sigma = 0.5$; (c) $\sigma = 1.0$; and (d) $\sigma = 1.5$.

5

Conclusions

In this chapter conclusions of this occluded object recognition research are presented. Also a list of contributions and limitations of this work are presented, as well as future work directions.

5.1 Work summary

This work presented a method for recognition of severe occluded objects. An state-of-the-art method on occluded object recognition resulted as an identification of more suitable approach for feature extraction, grouping and hypotheses estimation. Object shape was used as high-level feature for recognition of occluded objects and part division scheme was used as features grouping. The method uses High Curvature Points for the division of the occlusion in parts with some invariance to noise. Shapes representations state-of-the-art revealed that local shape representations are more suitable than global for occluded object recognition task. For shape-based hypotheses estimation, partial shape matching state-of-the-art was made leading to low-complexity accurate method for retrieval based on Pearson's correlation coefficient.

The proposed method has three main steps: representation, retrieval and hypotheses validation. For parts representation, Tangent Angle and Curvature signatures were used. Due to robustness when representing occlusion and non-rigid deformation and its low computational complexity. From experimentation both TAS and CV did not show significant difference in recognition results, and TAS was further used in the experiments because it is less computational expensive than CV.

Tangent Angle signature showed high sensitive to noise and Continuous Wavelet Transform coefficients were used in two ways for noise reduction: for representation and for low-pass filtering. The second approach showed better results with high improvements over synthetic and real image datasets.

The shape retrieval process for the parts was guided by a new proposed sort of the search space through an ensemble of one class Hidden Markov Models and similarity in the retrieval was measured by Pearson's correlation coefficient with very high performance. The new bayesian retrieval method incorporates the knowledge of the coherence of the retrieved object with the

occlusion, the similarity of the query part to the best match part in the object and the occurrence frequency of the part into the objects of the class.

Results showed the feasibility of the proposed method. Hidden Markov models based experimentation demonstrated the use of sorting in SSR improved efficiency over typical unsorted version. With only 40% of SSR visited, a difference of more than 0.28 was observed in precision and recall between sorted and unsorted version. For the bayesian method test, several scenarios were tested where the amount of occlusion, number of classes and quantity of objects in the occlusion were varied giving an insight on the capabilities and limitations of the method. The minimum F-Measure obtained in each experiment was 0.67, 0.93 and 0.92 respectively. An experiment over segmentation ground truth of CMU_KO dataset was conducted obtaining AUROC close to 0.9 for most classes. The proposed method achieved recognition of non-deformable severe occluded objects without any information about(1) quantity of objects present in the occlusion, (2) which parts of the occluded contour belong to each object, and (3) the categories of occluded objects with high performance.

5.2 Contributions

In this dissertation, several techniques have been explored for recognition of objects in scenes with severe occlusions. Our work has focused on the three key components of this problem: feature extraction and grouping, hypotheses retrieval, and hypotheses validation. Our findings in those areas are summarized as follow:

- *A new bayesian method for recognition of non-deformable severe occluded objects*: the problem of occluded object recognition was addressed with a bayesian method that incorporates the knowledge of the retrieved object coherence with occlusion, the similarity of the query part to the best match part in the object and the occurrence frequency of the part into the objects of the class.

- *A new local shape descriptor from the wavelet coefficients of the Tangent Angle signature*: feature extraction was made by means of object shape. Division of object into parts was accomplished by High Curvature Points, grouping all shape descriptors for the same part into the part signature. Tangent Angle was used for shape representation and Continuous Wavelet Transform was used for TAS noise reduction.

- *Proposal of search space retrieval sorting based on an ensemble of Hidden Markov Models*: an optimization for search reduction in shape retrieval space based on one class ensemble of Hidden Markov Models was proposed. Hypotheses retrieval using Pearson's correlation coefficients was done in the most probable class order according to a HMM ensemble. An stopping rule was set to stop the search of hypothesis from a query part.

- *Two new area based constraints for consistency checking between retrieved hypotheses and occlusion*: Were introduced as occlusion likelihood and hypotheses validation. Occlusion likelihood guarantee minimum fitness of hypothesis with occlusion, including transformation estimation validation. Hypotheses validation area constraint was used to eliminate duplicate and wrong estimated hypotheses after the retrieval step.

5.3 Limitations

The most common problem in the proposed method was non-High Curvature Point recognition in smooth or non-concave intersections. Those points are not recognized by definition and object parts division fails to group different objects descriptors in different parts. From erroneous part division is derived wrong hypotheses retrievals and, therefore, wrong recognition results.

Wrong hypothesis estimation also may occur because of non-discriminative part identification, such as for example straights lines. Error in retrieval is because very common parts are present in most objects and query prior probability are maximum for all these objects, hardening a correct hypothesis identification.

Objects with large area occluded and wrong parts classification are two issues directly related to the established thresholds for hypotheses validation and search stopping rule. The first one could lead to miss valid hypotheses because inner area constraint breaches depending on the τ_3 value. The second one can lead to wrong hypothesis retrieval and consequent erroneous recognition because of earlier stop of the search depending on the τ_1 value. Variations of thresholds must be done according to best parameter values in Subseção 4.1.1 for correct elimination of such problems.

In case of severe noise, i.e. $\sigma > 1.0$, wrong transformation estimation is obtained by RANSAC algorithm.

5.4 Future work

Despite contributions, this work is only a small step towards recognizing occluded object robustly. Several areas may be useful for further investigation as future works to this research. An scale invariant shape retrieval method based on Dynamic Time Warping could be examined for further improvement of retrieval step. Also a color based prior could be created for problems where region information is also in hand. Variational methods can be investigated for deformable shapes recognition aiming for articulated severe occluded object recognition approach. Further investigation is required for the establishment of an automatic method for determining the number of state and initialization of emission and transition matrixes for every class of objects in the HMM, for the optimal sort of the search space.

References

ADAMEK, T.; O'CONNOR, N. E. A multiscale representation method for nonrigid shapes with a single closed contour. **IEEE Transactions on Circuits and Systems for Video Technology**, v.14, n.5, p.742–753, 2004.

ALAJLAN, N.; KAMEL, M. S.; FREEMAN, G. H. Geometry-based image retrieval in binary image databases. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.30, n.6, p.1003–1013, 2008.

ANDREOPOULOS, A.; TSOTSOS, J. K. 50 Years of object recognition: directions forward. **Computer Vision and Image Understanding**, v.117, p.827–891, 2013.

BASRI, R. et al. Determining the similarity of deformable shapes. **Vision Research**, v.38, n.15, p.2365–2385, 1998.

BAY, H. et al. Speeded-Up Robust Features (SURF). **Computer Vision and Image Understanding**, v.110, p.346–359, 2008.

BELONGIE, S.; MALIK, J.; PUZICHA, J. Shape matching and object recognition using shape contexts. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.24, n.4, p.509–522, 2002.

BICEGO, M.; MURINO, V. Investigating hidden markov models capabilities in 2d shape classification. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.26, n.2, p.281–286, 2004.

BOYKOV, Y.; FUNKA-LEA, G. Graph cuts and efficient ND image segmentation. **International journal of computer vision**, v.70, n.2, p.109–131, 2006.

BOYKOV, Y.; JOLLY, M.-P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. **Eighth IEEE International Conference on Computer Vision (ICCV)**, v.1, p.105–112, 2001.

BRAHMBHATT, S. M. Detecting partially occluded objects in images. **Dissertação (Mestrado em Ciência da Computação)**, University of Pennsylvania, 2014.

BRONSTEIN, A. M. et al. Analysis of two-dimensional non-rigid shapes. **International Journal of Computer Vision**, v.78, n.1, p.67–88, 2008.

BRYNER, D. et al. 2D affine and projective shape analysis. **IEEE transactions on pattern analysis and machine intelligence**, v.36, n.5, p.998–1011, 2014.

CALONDER, M. et al. BRIEF: binary robust independent elementary features. **Springer Berlin Heidelberg**, p.778–792, 2010.

CAO, Y. et al. 2D nonrigid partial shape matching using MCMC and contour subdivision. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, v.1, p.2345–2352, 2011.

CHELLAPPA, R.; BAGDAZIAN, R. Fourier coding of image boundaries. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, n.1, p.102–105, 1984.

CHEN, L.; FERIS, R.; TURK, M. Efficient partial shape matching using Smith-Waterman algorithm. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops**, 2008.

COSTA, L. d. F. D.; CESAR JR, R. M. Shape analysis and classification: theory and practice. **CRC Press, Inc.**, 2000.

CUI, M. et al. Curve matching for open 2D curves. **Pattern Recognition Letters**, v.30, n.1, p.1–10, 2009.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)**, v.1, p.886–893, 2005.

DALIRI, M. R.; TORRE, V. Robust symbolic representation for shape recognition and retrieval. **Pattern Recognition**, v.41, n.5, p.1782–1798, 2008.

DAS, M.; PAULIK, M. J.; LOH, N. A bivariate autoregressive technique for analysis and classification of planar shapes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.12, n.1, p.97–103, 1990.

DAVIES, E. R. Machine vision: theory, algorithms, practicalities. **Elsevier**, 2004.

DICKINSON, S.; PIZLO, Z. Shape perception in human and computer vision. **Springer**, 2015.

DOLLAR, P. et al. Pedestrian detection: an evaluation of the state of the art. **IEEE transactions on pattern analysis and machine intelligence**, v.34, n.4, p.743–761, 2012.

DUBOIS, S. R.; GLANZ, F. H. An autoregressive model approach to two-dimensional shape classification. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, n.1, p.55–66, 1986.

DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern classification. **John Wiley & Sons**, 2012.

DURBIN, R. et al. Biological sequence analysis: probabilistic models of proteins and nucleic acids. **Cambridge university press**, 1998.

FELZENSZWALB, P. F. Representation and detection of deformable shapes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.27, n.2, p.208–220, 2005.

FELZENSZWALB, P. F. et al. Object detection with discriminatively trained part-based models. **IEEE transactions on pattern analysis and machine intelligence**, v.32, n.9, p.1627–1645, 2010.

FELZENSZWALB, P. F.; SCHWARTZ, J. D. Hierarchical matching of deformable shapes. **IEEE Conference on Computer Vision and Pattern Recognition**, p.1–8, 2007.

FISCHLER, M. A.; BOLLES, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. **Communications of the ACM**, v.24, n.6, p.381–395, 1981.

FRANSENS, R.; STRECHA, C.; VAN GOOL, L. A mean field em-algorithm for coherent occlusion handling in map-estimation prob. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)**, v.1, p.300–307, 2006.

GAO, T.; PACKER, B.; KOLLER, D. A segmentation-aware object detection model with occlusion handling. **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, p.1361–1368, 2011.

GDALYAHU, Y.; WEINSHALL, D. Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.21, n.12, p.1312–1328, 1999.

GIRSHICK, R. B.; FELZENSZWALB, P. F.; MCALLESTER, D. A. Object detection with grammar models. **Advances in Neural Information Processing Systems**, p.442–450, 2011.

GONZALEZ, R. C.; WOODS, R. E. Digital image processing. **Prentice hall Upper Saddle River, NJ, USA**, 2008.

GOPALAN, R.; TURAGA, P.; CHELLAPPA, R. Articulation-invariant representation of non-planar shapes. **European Conference on Computer Vision**, p.286–299, 2010.

GUERRERO-PENA, F. A. et al. Red Blood Cell Cluster Separation From Digital Images for Use in Sickle Cell Disease. **IEEE Journal of Biomedical and Health Informatics**, v.19, n.4, p.1514–1525, 2015.

GUO, G. et al. A shape reconstructability measure of object part importance with applications to object detection and localization. **International Journal of Computer Vision**, v.108, n.3, p.241–258, 2014.

HONG, B.-W. et al. Shape representation based on integral kernels: application to image matching and segmentation. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)**, v.1, p.833–840, 2006.

HORÁČEK, O.; KAMENICKÝ, J.; FLUSSER, J. Recognition of partially occluded and deformed binary objects. **Pattern Recognition Letters**, v.29, n.3, p.360–369, 2008.

HSIAO, E.; HEBERT, M. Occlusion reasoning for object detectionunder arbitrary viewpoint. **IEEE transactions on pattern analysis and machine intelligence**, v.36, n.9, p.1803–1815, 2014.

HU, M.-K. Visual pattern recognition by moment invariants. **IRE transactions on information theory**, v.8, n.2, p.179–187, 1962.

HUANG, X.; PARAGIOS, N.; METAXAS, D. N. Shape registration in implicit spaces using information theory and free form deformations. **IEEE transactions on pattern analysis and machine intelligence**, v.28, n.8, p.1303–1318, 2006.

HUTTENLOCHER, D. P.; KLANDERMAN, G. A.; RUCKLIDGE, W. J. Comparing images using the Hausdorff distance. **IEEE Transactions on pattern analysis and machine intelligence**, v.15, n.9, p.850–863, 1993.

KOCH, M. W. et al. Cueing, feature discovery, and one-class learning for synthetic aperture radar automatic target recognition. **Neural Networks**, v.8, n.7, p.1081–1102, 1995.

KON'YA, S.; KUSHIMA, K. A rotation invariant shape representation based on wavelet transform. **Workshop on Image Retrieval**, p.1–9, 1998.

KUMAR, N. et al. Leafsnap: a computer vision system for automatic plant species identification. **Computer Vision–ECCV**. p.502–516, 2012.

KWAK, S. et al. Learning occlusion with likelihoods for visual tracking. **International Conference on Computer Vision**, p.1551–1558, 2011.

LATECKI, L. J. et al. An elastic partial shape matching technique. **Pattern Recognition**, v.40, n.11, p.3069–3080, 2007.

LATECKI, L. J.; LAKÄMPER, R. Shape similarity measure based on correspondence of visual parts. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.22, n.10, p.1185–1190, 2000.

LATECKI, L. J.; LAKAMPER, R.; ECKHARDT, T. Shape descriptors for non-rigid shapes with a single closed contour. **IEEE Conference on Computer Vision and Pattern Recognition**, v.1, p.424–429, 2000.

LECUMBERRY, F.; PARDO, A.; SAPIRO, G. Simultaneous object classification and segmentation with high-order multiple shape models. **IEEE transactions on image processing : a publication of the IEEE Signal Processing Society**, v.19, n.3, p.625–35, 2010.

LI, Y.; GU, L.; KANADE, T. Robustly aligning a shape model and its application to car alignment of unknown pose. **IEEE transactions on pattern analysis and machine intelligence**, v.33, n.9, p.1860–1876, 2011.

LING, H.; JACOBS, D. W. Shape classification using the inner-distance. **IEEE transactions on pattern analysis and machine intelligence**, v.29, n.2, p.286–299, 2007.

LIU, M.-Y. et al. Fast directional chamfer matching. **IEEE Conference on Computer Vision and Pattern Recognition**, p.1696–1703, 2010.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v.60, p.91–110, 2004.

LU, C. et al. Shape guided contour grouping with particle filters. **IEEE 12th International Conference on Computer Vision**, p.2288–2295, 2009.

LU, G.; SAJJANHAR, A. Region-based shape representation and similarity measure suitable for content-based image retrieval. **Multimedia Systems**, v.7, n.2, p.165–174, 1999.

MA, T.; LATECKI, L. J. From partial shape matching through local deformation to robust global shape similarity for object detection. **IEEE Conference on Computer Vision and Pattern Recognition**, p.1441–1448, 2011.

MACRINI, D. et al. Bone graphs: medial shape parsing and abstraction. **Computer Vision and Image Understanding**, v.115, n.7, p.1044–1061, 2011.

MALLAT, S. A wavelet tour of signal processing: the sparse way. **Academic press**, 2008.

MANDAL, S.; Mahadeva Prasanna, S. R.; SUNDARAM, S. Curvature point based HMM state prediction for online handwritten assamese strokes recognition. **IEEE Twenty First National Conference on Communications (NCC)**, p.1–6, 2015.

MARVANIYA, S.; GUPTA, R.; MITTAL, A. Adaptive Locally Affine-Invariant Shape Matching. **arXiv**, 2015.

MCNEILL, G.; VIJAYAKUMAR, S. Hierarchical procrustes matching for shape retrieval. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, v.1, p.885–894, 2006.

MEGER, D. et al. Explicit Occlusion Reasoning for 3D Object Detection. **BMVC**, p.1–11, 2011.

MERHY, M. et al. An optimal elastic partial shape matching via shape geodesics. **IEEE International Conference on Image Processing (ICIP)**, p.4742–4746, 2014.

MICHEL, D.; OIKONOMIDIS, I.; ARGYROS, A. Scale invariant and deformation tolerant partial shape matching. **Image and Vision Computing**, v.29, n.7, p.459–469, 2011.

MOKHTARIAN, F. et al. Efficient and robust retrieval by shape content through curvature scale space. **Series on Software Engineering and Knowledge Engineering**, v.8, p.51–58, 1997.

MOKHTARIAN, F.; MACKWORTH, A. Scale-based description and recognition of planar curves and two-dimensional shapes. **IEEE transactions on pattern analysis and machine intelligence**, n.1, p.34–43, 1986.

MOKHTARIAN, F.; MACKWORTH, A. K. A theory of multiscale, curvature-based shape representation for planar curves. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.14, n.8, p.789–805, 1992.

MORI, G.; BELONGIE, S.; MALIK, J. Efficient shape matching using shape contexts. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.27, n.11, p.1832–1837, 2005.

OZDEMIR, B. et al. Performance measures for object detection evaluation. **Pattern Recognition Letters**, v.31, n.10, p.1128–1137, 2010.

RAVISHANKAR, S.; JAIN, A.; MITTAL, A. Multi-stage contour based detection of deformable objects. **European Conference on Computer Vision**, p.483–496, 2008.

ROTHER, C.; KOLMOGOROV, V.; BLAKE, A. Grabcut: interactive foreground extraction using iterated graph cuts. **ACM Transactions on Graphics (TOG)**, v.23, n.3, p.309–314, 2004.

RUBLEE, E. et al. ORB: an efficient alternative to sift or surf. **International Conference on Computer Vision**, p.2564–2571, 2011.

SABER, E.; XU, Y.; Murat Tekalp, A. Partial shape recognition by sub-matrix matching for partial matching guided image labeling. **Pattern Recognition**, v.38, n.10, p.1560–1573, 2005.

SEBASTIAN, T. B.; KLEIN, P. N.; KIMIA, B. B. Recognition of shapes by editing their shock graphs. **IEEE Transactions on pattern analysis and machine intelligence**, v.26, n.5, p.550–571, 2004.

SEKITA, I.; KURITA, T.; OTSU, N. Complex autoregressive model for shape recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.14, n.4, p.489–496, 1992.

SHARVIT, D. et al. Symmetry-based indexing of image databases. **IEEE Worshop on Content-Based Access of Image and Video Libraries**, p.56–62, 1998.

SIDDIQI, K. et al. Shock graphs and shape matching. **International Journal of Computer Vision**, v.35, n.1, p.13–32, 1999.

SMITH, T. F.; WATERMAN, M. S. Identification of common molecular subsequences. **Journal of molecular biology**, v.147, n.1, p.195–197, 1981.

SONKA, M.; HLAVAC, V.; BOYLE, R. Image processing, analysis, and machine vision. **Cengage Learning**, 2014.

TANG, M. et al. Grabcut in one cut. **IEEE International Conference on Computer Vision**, p.1769–1776, 2013.

TAUBIN, G.; COOPER, D. B. Recognition and positioning of rigid objects using algebraic moment invariants. **SAN DIEGO,'91, SAN DIEGO, CA**, p.175–186, 1991.

TEAGUE, M. R. Image analysis via the general theory of moments. **JOSA**, v.70, n.8, p.920–930, 1980.

TIENG, Q. M.; BOLES, W. Recognition of 2D object contours using the wavelet transform zero-crossing representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.19, n.8, p.910–916, 1997.

UNGER, M. et al. TVSeg-Interactive Total Variation Based Image Segmentation. **BMVC**, v.31, p.44–46, 2008.

VAN OTTERLOO, P. J. A contour-oriented approach to shape analysis. **Prentice Hall International (UK) Ltd.**, 1991.

VICENTE, S.; KOLMOGOROV, V.; ROTHER, C. Joint optimization of segmentation and appearance models. **IEEE 12th International Conference on Computer Vision**, p.755–762, 2009.

VU, N.; MANJUNATH, B. Shape prior segmentation of multiple objects with graph cuts. **IEEE Conference on Computer Vision and Pattern Recognition**, p.1–8, 2008.

WANG, X. et al. Fan shape model for object detection. **IEEE Conference on Computer Vision and Pattern Recognition**, p.151–158, 2012.

WANG, X.; HAN, T. X.; YAN, S. An HOG-LBP human detector with partial occlusion handling. **IEEE 12th International Conference on Computer Vision**, p.32–39, 2009.

XU, C.; LIU, J.; TANG, X. 2D shape matching by contour flexibility. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.31, n.1, p.180–186, 2009.

YANG, H. S.; LEE, S. U.; LEE, K. M. Recognition of 2D object contours using starting-point-independent wavelet coefficient matching. **Journal of Visual Communication and Image Representation**, v.9, n.2, p.171–181, 1998.

YANG, M.; KPALMA, K.; RONSIN, J. A survey of shape feature extraction techniques. **Pattern recognition**, p.43–90, 2008.

YOUNG, I. T.; WALKER, J. E.; BOWIE, J. E. An analysis technique for biological shape. I. **Information and control**, v.25, n.4, p.357–370, 1974.

ZAHN, C. T.; ROSKIES, R. Z. Fourier descriptors for plane closed curves. **IEEE Transactions on Computers**, v.100, n.3, p.269–281, 1972.

ZHANG, D.; LU, G. Generic Fourier descriptor for shape-based image retrieval. **IEEE International Conference on Multimedia and EXPO (ICME)**, v.1, p.425–428, 2002.

ZHANG, D.; LU, G. Review of shape representation and description techniques. **Pattern Recognition**, v.37, n.1, p.1–19, 2004.

ZHANG, D.; LU, G. et al. A comparative study on shape retrieval using Fourier descriptors with different shape signatures. **International Conference on Intelligent Multimedia and Distance Education (ICIMADE01)**, 2001.

ZHANG, D.; LU, G. et al. A comparative study of Fourier descriptors for shape representation and retrieval. **Asian Conference on Computer Vision (ACCV)**, p.646–651, 2002.

ZHANG, G.; XU, J.; LIU, J. A New Method for Recognition Partially Occluded Curved Objects under Affine Transformation. **IEEE International Conference on Intelligent Systems and Knowledge Engineering (ISKE)**, n.1, p.456–461, 2015.

ZHOU, B. et al. Learning deep features for scene recognition using places database. **ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS.**, p.487–495, 2014.

A

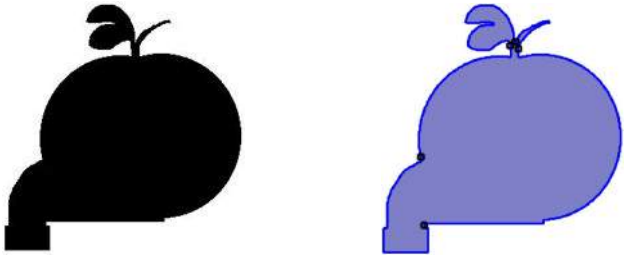
Appendix A - Method walkthrough in examples

Example 1:

Addresses only cases when high curvature points do not generate good shape division. This means intersection between objects is not a high curvature point.

1. High curvature points detection

High curvature points detection as shown in the paper. In this example, one intersection between the occluded objects was not detected as a high curvature point. Occlusion was divided into 5 parts and representation and retrieval was executed as described in dissertation. The original image here corresponds to occlusion between an apple and a hammer.



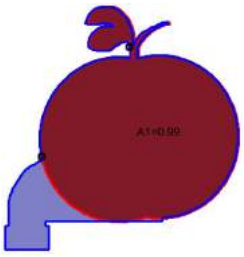
(a) Original image (b) High curvature points definition

2. Appling Restriction 1 to retrieved hypotheses

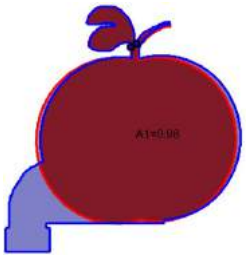
Preventing undesirable retrieved objects as being accepted as valid. Retrieved objects with large areas outside the occlusion are likely to correspond to invalid objects. The equation used for Restriction 1 is occlusion likelihood:

$$p(O|y) = \frac{area(y \bigcap O)}{area(y)}$$

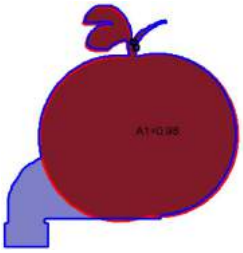
Blue contour represents occlusion O and red contour is analyzed hypothesis y. Dark red area is intersection $y \bigcap O$, and light red area represents exterior area to occlusion $y \setminus O$. $A1$ refers to $p(O|y)$ and in this case, all hypothesis y with $p(O|y) > \tau_2$ are taken as valid. In this example, $\tau_2 = 0.9$ and no hypotheses were discarded in this step.



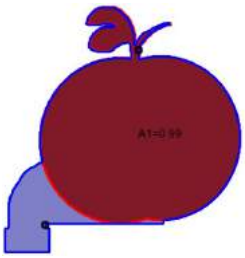
Hypotheses 1: A1=0.99 (valid)



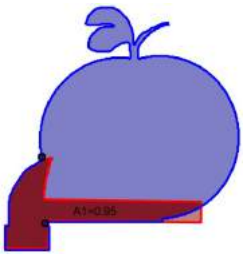
Hypotheses 2: A1=0.98 (valid)



Hypotheses 3: A1=0.98 (valid)



Hypotheses 4: A1=0.99 (valid)



Hypotheses 5: A1=0.95 (valid)

3. Sorting valid hypotheses by A1: $H = \{H5, H2, H3, H1, H4\}$

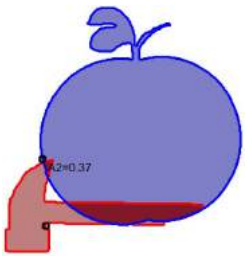
4. Appling Restriction 2 to valid hypotheses

Eliminating multiple valid hypotheses. Duplicated hypotheses are removed until only one remains. This means invalid hypotheses are removed each iteration. Equations used for Restriction 2 are:

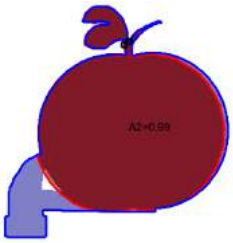
$$A_k = \frac{area(y_k \bigcap O_k^{'})}{area(y_k)}$$

$$O_k^{'} = O \bigcap \left(\bigcup_{\forall j \neq k} y_j^* \right)$$

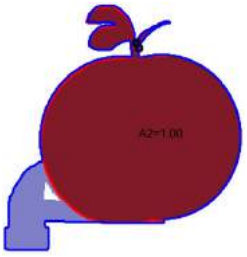
Blue contour represents generated occlusion $O_k^{'}$ and red contour is analyzed hypothesis y. Dark red area is intersection $y \bigcap O_k^{'}$, and light red area represents non-occluded area of hypothesis y. All hypothesis y with $A1 < \tau_3$ are valid. In this example $\tau_3 = 0.85$ and duplicated hypotheses are discarded. The remaining valid hypotheses set H is shown at each step, eliminating one duplicated hypothesis each time.



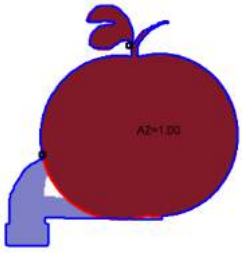
Hypotheses 5: A2=0.37 (valid) $H = \{H5, H2, H3, H1, H4\}$



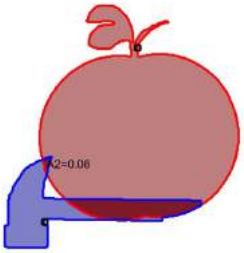
Hypotheses 2: A2=0.99 (invalid) $H = \{H5, H3, H1, H4\}$



Hypotheses 3: A2=1.00 (invalid) $H = \{H5, H1, H4\}$



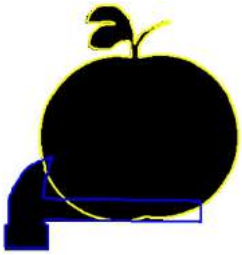
Hypotheses 1: A2=1.00 (invalid) $H = \{H5, H4\}$



Hypotheses 4: A2=0.08 (valid) $H = \{H5, H4\}$

5. Showing valid hypotheses

Valid hypotheses of last step correspond to recognition result.

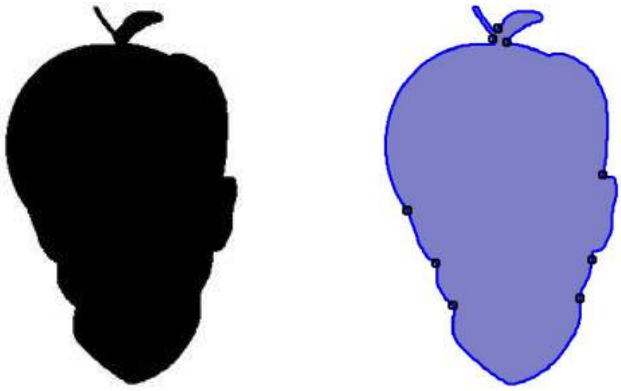


Example 2:

Addresses only cases when high curvature points do not generate good shape division. This means intersection between objects is not a high curvature point. Original image of this example corresponds to occlusion of an apple and a face.

1. High curvature points detection

High curvature points detection as in the paper. Occlusion is divided into 9 parts (6 hypotheses are shown) in this example and representation and retrieval is done as proposed in dissertation.



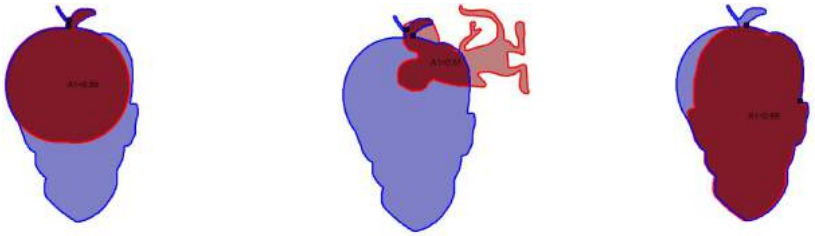
(a) Original image (b) High curvature points definition

2. Appling Restriction 1 to retrieved hypotheses

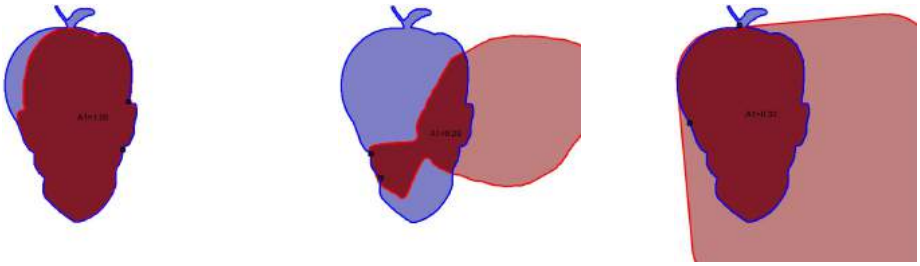
Preventing undesirable retrieved objects as being accepted as valid. Retrieved objects with large areas outside the occlusion are likely to correspond to invalid objects. The equation used for Restriction 1 is occlusion likelihood:

$$p(O|y) = \frac{area(y \bigcap O)}{area(y)}$$

Blue contour represents occlusion O and red contour is analyzed hypothesis y. Dark red area is intersection $y \bigcap O$, and light red area represents the exterior area to occlusion $y \setminus O$. $A1$ refers to $p(O|y)$ and all hypothesis y with $p(O|y) > \tau_2$ are accepted as valid. In this example, $\tau_2 = 0.9$ and hypotheses with too much exterior area are discarded in this step, as in the cases 2, 7 and 9.



Hypotheses 1: A1=0.99 (valid) Hypotheses 2: A1=0.51 (invalid) Hypotheses 3: A1=0.99 (valid)



Hypotheses 4: A1=1.00 (valid) Hypotheses 7: A1=0.29 (invalid) Hypotheses 9: A1=0.33 (invalid)

3. Sorting valid hypotheses by A1: $H = \{H1, H3, H6, H4\}$

4. Appling Restriction 2 to valid hypotheses

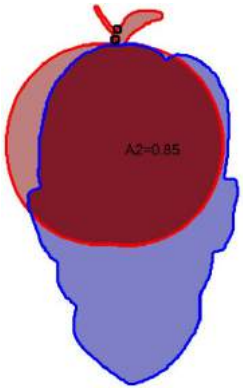
Eliminating multiple valid hypotheses. Duplicated hypotheses are removed until only

one remains. This means invalid hypotheses are removed each iteration. Equations used for Restriction 2 are:

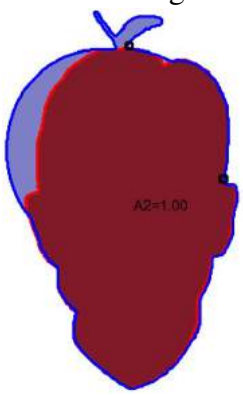
$$A_k = \frac{area(y_k \bigcap O_k^{'})}{area(y_k)}$$

$$O_k^{'} = O \bigcap \left(\bigcup_{\forall j \neq k} y_j^* \right)$$

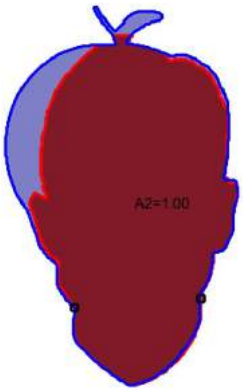
Blue contour represents generated occlusion $O_k^{'}$ and red contour is analyzed hypothesis y. Dark red area is intersection $y \bigcap O_k^{'}$, and light red area represents non-occluded area of hypothesis y. $A2$ refers to A_k and all hypothesis y with $A1 < \tau_3$ are valid. In this example, $\tau_3 = 0.85$ and duplicated hypotheses are discarded. The remaining valid hypotheses set H is shown each step.



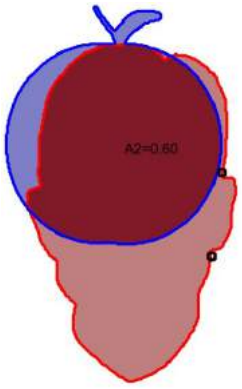
Hypotheses 1: A2=0.85 (valid)
$H = \{H1, H3, H6, H4\}$



Hypotheses 3: A2=1.00 (invalid)
$H = \{H1, H6, H4\}$



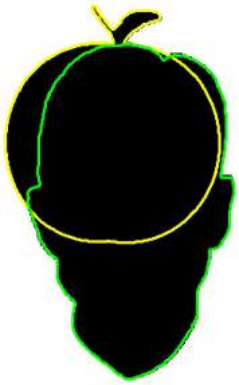
Hypotheses 6: A2=1.00 (invalid)
$H = H1, H4$



Hypotheses 4: A2=0.60 (valid) $H = H1, H4$

5. Showing valid hypotheses

Valid hypotheses of last step correspond to recognition result.

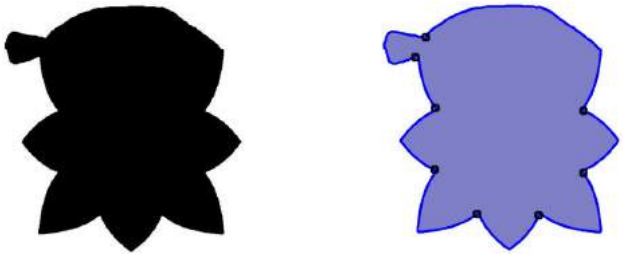


Example 3:

Addresses only cases when high curvature points do not generate good shape division. This means intersection between objects is not a high curvature point. Original image of this example corresponds to occlusion of a fish and device2 class of object from MPEG7 dataset.

1. High curvature points detection

High curvature points detection as in the paper. Occlusion is divided into 8 parts (6 hypotheses are shown) in this example, and representation and retrieval is done as described in dissertation.



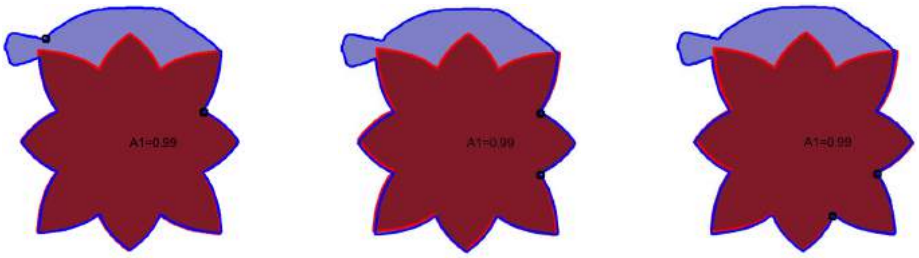
(a) Original image (b) High curvature points definition

2. Appling Restriction 1 to retrieved hypotheses

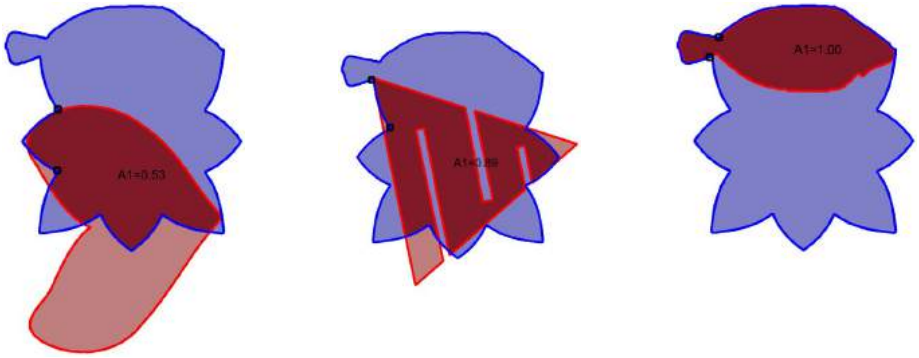
Preventing undesirable retrieved objects as being accepted as valid. Retrieved objects with large areas outside the occlusion are likely to correspond to invalid objects. The equation used for Restriction 1 is occlusion likelihood:

$$p(O|y) = \frac{area(y \bigcap O)}{area(y)}$$

Blue contour represents occlusion O and red contour is analyzed hypothesis y. Dark red area is intersection $y \bigcap O$, and light red area represents exterior area to occlusion $y \setminus O$. $A1$ refers to $p(O|y)$ and all hypothesis y with $p(O|y) > \tau_2$ are accepted as valid. In this example, $\tau_2 = 0.9$ and hypotheses with too much exterior area are discarded in this step, as in case 6.



Hypotheses 1: A1=0.99 (valid) | Hypotheses 2: A1=0.99 (valid) | Hypotheses 3: A1=0.99 (valid)



Hypotheses 6: A1=0.53 (invalid) | Hypotheses 7: A1=0.89 (valid) | Hypotheses 8: A1=1.00 (valid)

3. Sorting valid hypotheses by A1: H={H7,H1,H2,H3,H4,H5,H8}

4. Appling restriction 2 to valid hypotheses

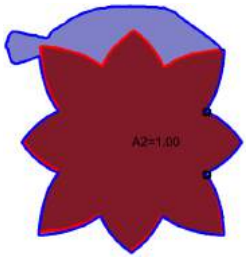
Eliminating multiple valid hypotheses. Duplicated hypotheses are removed until only one remains. This means, invalid hypotheses are removed each iteration. Equations used for Restriction 2 are:

$$A_k = \frac{area(y_k \bigcap O_k^{'})}{area(y_k)}$$

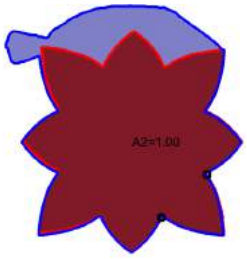
$$O_k^{'} = O \bigcap \left(\bigcup_{\forall j \neq k} y_j^* \right)$$

Blue contour represents generated occlusion $O_k^{'}$ and red contour is analyzed hypothesis y. Dark red area is intersection $y \bigcap O_k^{'}$, and light red area represents non-occluded area of hypothesis y. $A2$ refers to A_k and all hypothesis y with $A1 < \tau_3$ are valid. In this example $\tau_3 = 0.85$ and duplicated hypotheses are discarded. Valid hypotheses set H is show in each step.

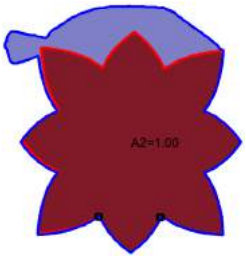
Hypotheses 7: A2=1.00
(invalid)
$H = \{H1, H2, H3, H4, H5, H8\}$



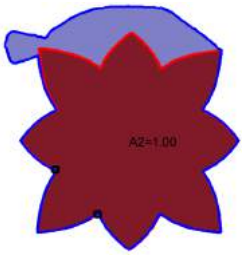
Hypotheses 1: A2=1.00
(invalid)
$H = \{H2, H3, H4, H5, H8\}$



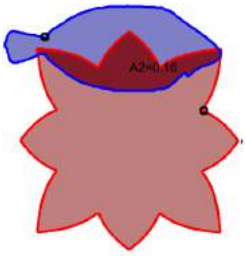
Hypotheses 2: A2=1.00
(invalid)
$H = \{H3, H4, H5, H8\}$



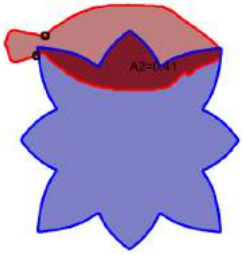
Hypotheses 3: A2=1.00
(invalid)
$H = \{H4, H5, H8\}$



Hypotheses 4: A2=1.00
(invalid)
$H = \{H5, H8\}$



Hypotheses 5: A2=0.16
(valid)
$H = \{H5, H8\}$



Hypotheses 8: A2=0.41
(valid) $H = \{H5, H8\}$

5. Showing valid hypotheses

Valid hypotheses of last step correspond to recognition result.