UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR

RECONOCIMIENTO DE OBJETOS 2D BASADO EN ANALISIS DEL CONTORNO

-PROYECTO FIN DE CARRERA-

ULYA BAYRAM

JULIO 2009

RECONOCIMIENTO DE OBJETOS 2D BASADO EN ANALISIS DEL CONTORNO

-PROYECTO FIN DE CARRERA-

AUTOR: Ulya Bayram
TUTOR: Juan Carlos San Miguel Avedillo
PONENTE: José María Martínez Sánchez

Video Processing and Understanding Lab (VPU)
Dpto. de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Julio de 2009

Acknowledgments

Primarily I would like to thank Universidad Autonoma de Madrid for accepting me as an Erasmus Exchange student and taking care of all my problems. I would like to thank José María Martínez Sánchez for all his help on convincing me to complete my PFC on image processing field, Jesús Bescós Cano for introducing me a whole new field of Image Processing and Juan Carlos San Miguel Avedillo for all his help on this PFC with patience.

I also would like give my thanks to Bilkent University, where I felt as my home, for giving me the opportunity to study in UAM for a whole academic year; basically this opportunity opened my eyes for life. Special thanks to Nail Akar, Associate Prof. of Electrical and Electronics Engineering Department of Bilkent University for giving a special care for all his exchange students and for his advices on our academic lives.

I am grateful to all my friends from UAM, basically they made me feel home and be there with me when I felt homesick. Also thanks to my friends from Turkey for not forgetting me and not loosing the contact even when they were busy with their own problems.

Finally I want to thank to my family; my parents for their patience, my brothers for listening all my problems with patience and for their valuable advises, basically thank you for being my life coaches. Thanks you all for supporting all decision that I have made all along my life even if when are wrong. I am grateful for making me feel loved even from this long distance.

Ulya Bayram
July 2009

CONTENTS

FIGURE INDEX

TABLES INDEX

1 Introduction

1.1 Motivation

Surveillance is a French word that means *observing*. As it is clear from this translation, video-surveillance defines the observation made by electronic devices such as cameras. These devices transform the observed data into video content.

Nowadays video-surveillance is the most commonly used security system in public areas such as airports, train stations, schools and even street corners. However, it is impossible to monitor all the video content captured by using human operators and to detect if an event occurs at a place protected by a video-surveillance system. In order to detect the events or anomalous behaviors in the shortest and easiest way, the personal monitoring could be helped by proving automatic analysis and interpretation tools to make him/her be able to focus his/her attention.

During last years the research community focused their effort on the development of different techniques to automatically detect and classify the video content by analyzing different patterns. Features like color, texture and shape [1] are typically used to classify the video content.

In this context, object classification by using shape-based information is one of the most common approaches to classify objects in the monitored scene [2][3]. The success of this process depends on the description of the object features and the employed classification schemes which bring the necessity of determining the most appropriate shape-based features and matching techniques for each case.

Shape-based features are commonly used for object description and classification due to their easily extraction process, but they are one of the major challenges because of the wide range of poses, scales, illumination variations and camera position in the scene. The use of these features involves three aspects: their representation (feature selection), feature extraction and the matching procedure applied (feature or description matching).

Object classification process begins with feature selection. There are a lot of shape-based features in the literature; some of them use the contour of the detected object while the others use the region of the object. To have a successful classification, it is mandatory to select the optimal subset of features or each application.

After the features selection is completed, the process named feature extraction begins. When the selected features are extracted from the detected object, the result is a feature vector that basically consists of some values based on shape-properties of the contour or region.

Matching is the final process before completing the classification process. At this stage, previously extracted features of the training data are compared to the features of test data by using a certain classification method.

1.2 Objectives

The objective of the proposed project is the study and implementation of different classification schemes using 2D shape-features in the object classification stage of a typical video analysis system. The video-surveillance domain is proposed as a case of study for the object classification task. This main objective is divided into the following aspects:

- To study the state of the art in the following topics:
  - Shape-based object classification stage applied in video surveillance systems, analyzing the necessary resources (frame acquisition, object detection, object classification, databases…) to develop algorithms for object classification.
  - 2D shape-based features for object classification. This study covers three parts: extraction, representation and matching.
  - Techniques used for 2D shape-based object classification. This study surveys the existing approaches and their disadvantages and advantages.
  - Datasets used for training and testing the 2D shape-based object classification approaches of the state of the art.
- To select and implement the most relevant 2D shape-based features and the needed extraction methods for the video-surveillance domain.
- To select and implement the most relevant classification methods in order to test each implementation with the proposed shape-based description.
- To study in depth the different combinations between 2D shape-based features and classification methods in different contexts. The objective of this study is the identification of the optimal combinations that produce the best results in terms of accuracy and computational time. This study can be divided into the following aspects:
  - Define the different contexts (or scenarios) of application in the video-surveillance domain.
  - Define an appropriate shape-based dataset (if needed) for each scenario.

- Select the most appropriate set of features and classification techniques (from the implemented ones) in order to test the accuracy of each classification scheme proposed.
- Evaluate the results obtained in terms of accuracy and computational time.

1.3 Document structure

This document is structured by the following chapters:

- **Chapter 1** provides the introduction with basic concepts of the project.
- **Chapter 2** describes the state-of-the-art related to the topic of the project (2D shape-based object classification). It includes the most common approaches related to preprocessing, feature extraction and classification methods.
- **Chapter 3** presents a detailed description of the shape-based features and their extraction techniques selected for the project.
- **Chapter 4** presents a detailed description of the matching and classification techniques selected for the project with all details.
- **Chapter 5** describes the framework that integrates the different techniques developed in the project by defining each module used, the operations performed inside and available techniques in each module.
- **Chapter 6** describes the different application scenarios proposed, the corresponding selected datasets and the performance of the applied techniques in each scenario.
- **Chapter 7** briefly overviews the work done briefly, highlights the conclusions and describes the possible future work of this project.

Additionally two appendices have been added to overview some techniques used within the project:

- ♦ **Appendix A** Description of Background Subtraction Method
- ♦ **Appendix B** Description of Active Contour adjustment method (Snake's Method)

2 State of the art in shape-based classification

2.1 Introduction to object classification

According to [2][3], there are three main approaches to object classification: shape-based classification, motion-based classification and combined shape and motion based classification [3]. Shape-based object classification is the computation of the similarity between two shapes based on the selected features that describe the shapes. In general, a confidence measure of the similarity is computed as a result of the comparison process. This classification stage involves two main steps: feature selection/extraction and classification.

The proposed study reviews the shape-based features and classification techniques used in the video surveillance domain for the 2D object classification task. The main modules or processing stages of a typical video surveillance system that uses shape-based features for the object classification task are shown in Figure 1.

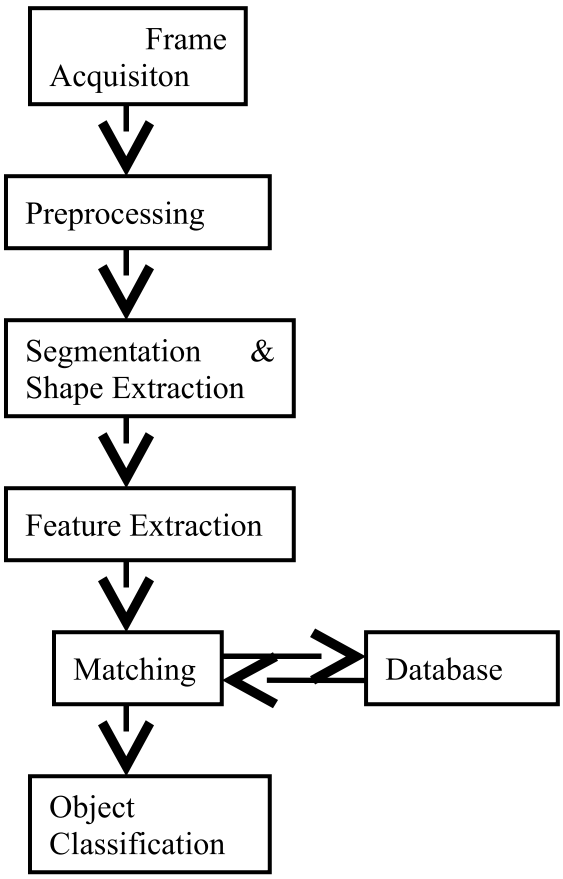


Figure-1: Data flow diagram in a surveillance system that uses shape-based features for the classification task

Generally, visual information is converted to digital signals (raw frames) in the frame acquisition module by using a video camera and the raw image is sent to the following modules. In the image preprocessing module, the acquired image is properly processed to use it in the following analysis stages. In the image segmentation and shape extraction module, the detected objects are extracted from frames and their shape is computed by applying different extraction techniques. In the following modules, shape features of the detected objects are extracted and in the classification stage, they are compared to the shape-based object models available in the database [5].

In the following sections, all existing and useful 2D shape-based features, their extraction methods and the most common classification techniques are briefly described in order to determine the optimal combination of features and matching techniques that will be used in the proposed work in the document.

2.2 Image pre-processing and shape extraction

2.2.1 Image Preprocessing

The operation performed after frame acquisition is object detection. The most common approaches in object detection are Optical Flow [6], Background Subtraction [7], Temporal Differencing [7] and Statistical Methods [8]. According to recent surveyed papers [4][5][6] object detection based on *Background Subtraction* followed by object tracking (in order to maintain the ID of the classified object) is the most common solution adopted due to its simplicity and effectiveness under certain constraints (e.g. stationary camera, slow background change, intermediate velocity of the objects). An example of *Background Subtraction* results can be seen at Figure 2.

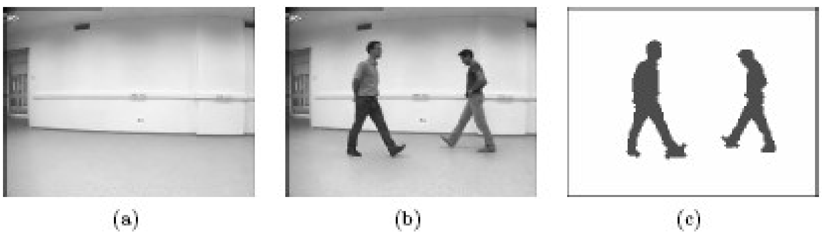


Figure-2: Background subtraction based object detection where (a) is the background, (b) is the current image and (c) is the detected objects [7]

After object detection, the next applied operation applied is preprocessing. Most shape-based classification approaches do not mention preprocessing techniques applied in order to help the extraction of the relevant shape data. However, there are some works that describe the most common preprocessing techniques [6][7]. According to these papers, there can be some defects in the detected objects such as camera noise, reflectance noise, background colored object noise or shadows and sudden illumination change. In order to remove these non-desirable effects; morphological operations, erosion and dilation, are usually applied [7]. The main drawback of these operations is the deformation produced in the contour of the object analyzed. More recently, more sophisticated morphological operations are applied in video-surveillance systems in order to remove these noisy effects without deforming the object contour Hata: Başvuru kaynağı bulunamadı.

Figure-3: Binary Blob samples

2.2.21 Shape Extraction

The following stage in the video analysis system described (see Figure 1) is the extraction of shape information. In the studied papers, the input data of this operation are binary blobs of detected objects (Figure 3). In order to extract shapes from binary blobs, a contour following algorithm is applied and boundary outlines are obtained [7] as in Figure 4.

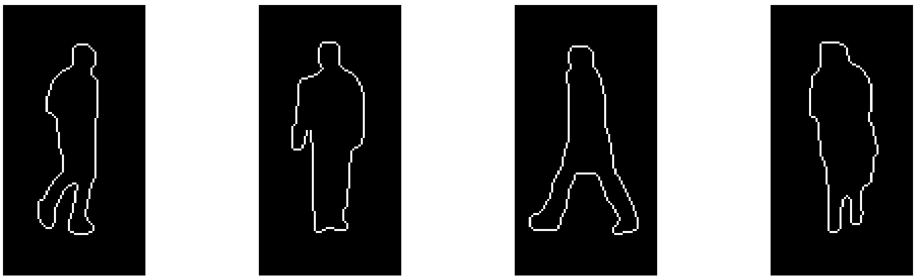


Figure-4: Binary blobs after shape extraction process

The output data is a sequence of pixel locations indicating the boundaries of the detected objects:

$$C = \{ p_1, p_2, ..., p_N \} \; where \; p_i = (x, y)$$

2.3 Shape-based Features

There are a lot of shape-based features available in the existing literature. In this section, the most common features and their extraction techniques are described. Due to the high amount of features available, a subset of them is usually used as a description of the object. Thus, the extracted feature data (or descriptor) is generally an N-dimensional vector which can be regarded as a point in an N-dimensional space [10].

$$Descriptor = [f_1, f_2, ..., f_N] \; where \; f_i \; is \; a \; feature$$

Features can be classified as contour-based or region based according to whether contour or region of the object is used in each specific technique. Also extraction techniques can be separated as structural or global if the shape is represented by sub-parts or as a whole [10]. According to this classification scheme, the existing shape-based features and extraction techniques are presented (Figure-4).

Shape features

- Contour based features
  - Structural
    - -Chain Code
    - -Polygon
    - -Smooth Curve
    - -Scale Space
    - -Syntactic Analysis
    - -Shape Context
  - Global
    - -Simple Descriptors
    - -Shape Signature
    - -Boundary Moments
    - -Wavelet Descriptor
    - -Scale Space
    - -Autoregressive
    - -Elastic Matching
    - -Distance Signal
    - -Hausdoff Distance
- Region based features
  - Structural
    - -Convex Hull
    - -Media Axis
  - Global
    - -Height by Width Ratio
    - -Fill Ratio
    - -Segment Convexity
    - -Convex Deviation
    - -Compactness
    - -Euler Number
    - -Eccentricity
    - -Geometric Moments
    - -Zernike Moments
    - -Pseudo Zernike
    - -Legendre Moments
    - -Shape Matrix

Figure-5: Features classification taxonomy proposed in [10]

2.3.1 Contour Based Features

These features use boundary information of the blob. There are two types of approaches to contour based features: global and structural. Global approaches use the continuous boundary information without dividing it into sub-parts. On the other hand, structural approaches divide the boundary information into segments and represent the boundary as a string or a graph (tree), so the matching operation becomes a string matching or graph matching [10].

2.3.1.11 Global Methods:

I. Simple Shape Descriptors:

Simple contour based features involve the use of the total number of border pixels, the binary blob size and some relations between them.

- **Perimeter:** This feature is the total number of border pixels of the blob. The result of this feature is an integer number represented it with *P* [8].
- **Eccentricity:** This feature is the value of the ratio of the length of the major axis over the length of the minor axis of the detected object. As it can be seen from Figure 6, this feature is the value of *a/b* [10].

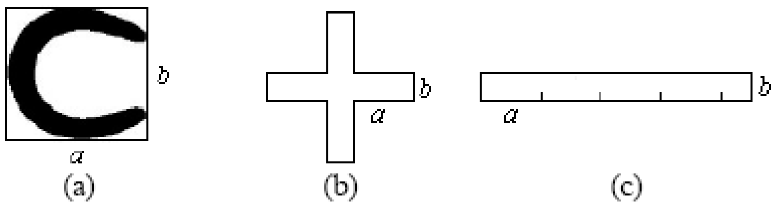


Figure-6: Image samples for eccentricity [10]

- **Height by Width Ratio:** This feature is the ratio of the Minimum Bounding Rectangle height and the width of the binary blobs. Minimum bounding rectangle samples can be seen in Figure 7. Therefore this feature is the value of *H/W* where *H* is the height and *W* is the width of these rectangles [4] [8].

Figure-7: Minimum bounding rectangle

II. Shape Signature:

By using this feature, a 2D shape is represented as a 1D function derived from its boundary. Boundary is represented as a sequence of coordinates (*x(t), y(t))* where *t=0,1, ...,N-1*. The Centroid distance is calculated from formula:

$r(t) = ([(x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2}$ where (x_c, y_c) are arbitrary points.

This *r(t)* function is the shape signature that can be observed from Figure 8. Shape signature feature is very sensitive to noise, so this method is not preferred to be used directly on image classification [11].

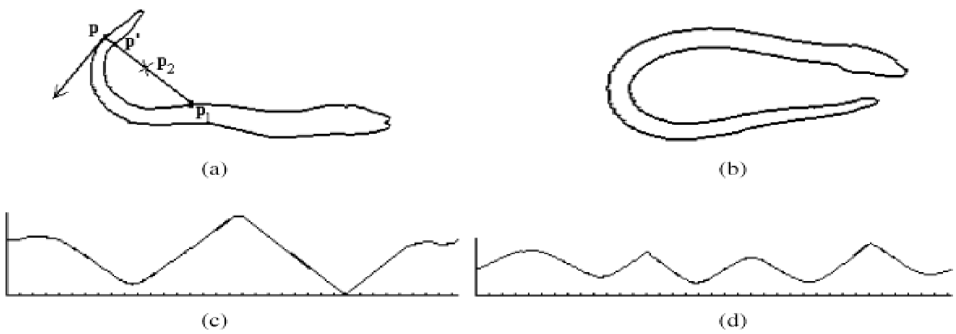


Figure-8: Shape signature samples where (a) and (b) are different shapes and (c) and (d) are shape signature results [8]

III. Boundary Moments:

In this feature, boundary is represented as $z(i)$ where *i* represents the pixel number of the boundary. According to this signature, the *rth* moment and central moments are calculated as:

$m_r = \frac{1}{N}\sum_{i=1}^{N} z(i)^r$ where m_r is the rth moment,

$\mu_r = \frac{1}{N}\sum_{i=1}^{N}(z(i) - m_1)^r$ where μ_r is the central moment and N is the total number of boundary points.

After a normalization procedure:

$$\bar{m}_r = \frac{m_r}{(\mu_2)^{r/2}} \text{ and } \bar{\mu}_r = \frac{\mu_r}{(\mu_2)^{r/2}}.$$

The output of this feature is a vector that contains the *rth* moment of the boundary pixels. The implementation of this feature is easy, but difficult to combine with higher order moments [10].

IV. Curvature Scale Space Method:

This feature is the representation of the curvature of the boundary as a scale space. If the boundary is represented as *C(s)=(x(s),y(s))*, where *x(s), y(s)* are the coordinates of the *s* boundary point, then this coordinate function can be convolved with a Gaussian kernel ϕ_σ :

$$x_\sigma(s) = \int x(s)\phi_\sigma(t - s)dt \text{ where } \phi_\sigma(t) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{t^2}{2\sigma^2}},$$

and the same calculation is done for *y(s)*. The peaks from the resulting convolved coordinates are used for matching. Before the matching operation based on peaks is applied, all the images are scaled into same size in order to have the same number of boundary points.

This peak matching is a point to point matching operation, so this method can be very complex according to possible high number of peaks [12].

V. Autoregressive Method:

This feature is based on the stochastic modeling of a 1D function *f* of boundary. The Autoregressive Method expresses the value of this function as a linear combination of the preceding values. Where the function is f , model coefficients are expressed as θ_j , m is order of the model and $\sqrt{\beta}\,w_t$ is the error term:

$$f_t = \alpha + \sum_{j=1}^{m}\theta_j f_{t-j} + \sqrt{\beta}\,w_t.$$

The feature vector of this method can be represented as:

$$\left[\theta_1,\ldots,\theta_m,\alpha/\sqrt{\beta}\right]^T$$

For complex boundaries, the autoregressive model is not sufficient for shape-based object description [10].

VI. Elastic Matching:

To use this feature, the original template τ (s) which is the boundary points of the original object is deformed by a warping deformation θ (s) as:

$\varphi(s) = \tau(s) + \theta(s)$ where $\varphi(s)$ represents the deformed template.

Similarity between the deformed template and the test image shape is measured by finding the values of functions that minimize the following function:

$$F = S + B + M = \alpha \int_0^1 \left[\left(\frac{d\theta_x}{ds}\right)^2 + \left(\frac{d\theta_y}{ds}\right)^2 \right] ds + \beta \int_0^1 \left[\left(\frac{d^2\theta_x}{ds}\right)^2 + \left(\frac{d^2\theta_y}{ds}\right)^2 \right] ds + \int_0^1 I_E(\varphi(s))ds$$

Where I_E is the original image, *S* and *B* are the strain and bending energies, *M* is the degree of deformation, *N* is the shape complexity value and *C* is the correlation, and therefore the feature vector is [13]:

(S, B, M, N, C).

VII. Distance Signal:

For the extraction of this feature, first the center of mass $C_m(x_m, y_m)$ of the blob is calculated as:

$$x_{C_m} = \frac{\sum x_i}{n} \text{ and } y_{C_m} = \frac{\sum y_i}{n}$$

where x_i and y_i are boundary point coordinates and *n* is the total number of boundary points.

$S=\{p_1,p_2,\ldots,p_n\}$ is the boundary representation of points that are ordered from the top center point that is found by C_m in clockwise direction. Then the distance signal is defined as:

$DS=\{d_1,d_2,\ldots,d_n\}$ found by function $d_i=dist(C_m,p_i)$

where "dist" function represents Euclidean distance from the center of mass to the boundary pixels. This operation and distance signal sample can be seen from Figure 9.

Since different shapes have different sizes, it is necessary to fix the size of the distance signal, for this purpose a scaling operation:

$$\overleftrightarrow{DS}[i] = DS\left[i * \frac{N}{C}\right]$$ is applied where C is a constant.

As final operation, this scaled distance signal is normalized as:

$$\overline{DS}[i] = \frac{\overleftrightarrow{DS}[i]}{\sum_{1}^{n} \overleftrightarrow{DS}[i]}$$

and this normalized distance signal vector is the feature vector of this method [7].

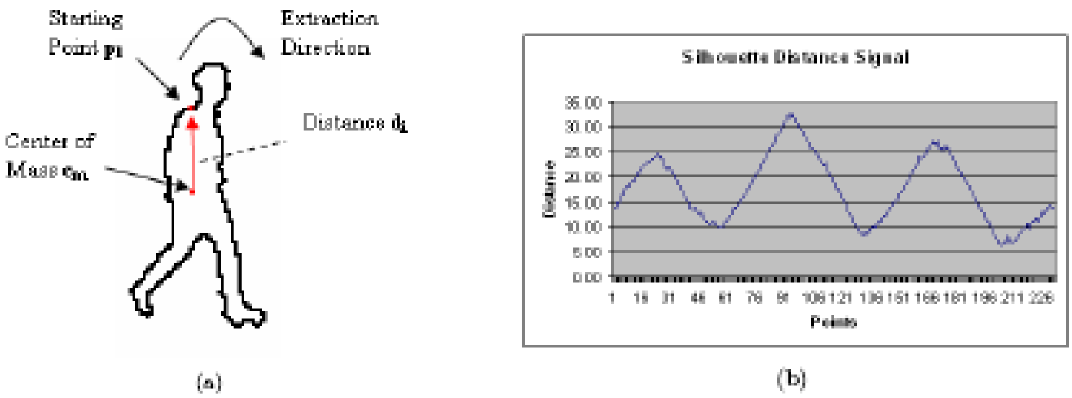


Figure-9: (a) Sample shape and (b) distance signal [7]

2.3.1.21 Structural Methods:

I. Chain Code Representation:

In order to use this feature, boundary is represented as a unit-size line segments sequence with a given orientation. A boundary is superimposed with a grid, and then approximated to the nearest grid point and a sampled image is obtained. Starting from an arbitrarily selected point, the chain code is generated by applying a 4-directional or 8-directional chain code.

In order to normalize the chain code, the pixel on the boundary that results the minimum coordinate values of x and y are taken as the starting point. Also the boundary can be represented as differences in successive directions in the chain code. For this case, computation is done by subtracting each element of the chain code from the previous one.

This chain code representation is generally not the preferred one to be used in shape matching operations due to its sensitivity to noise and variations [14].

II. Polygon Decomposition:

In the extraction of this feature, the boundary is divided into line segments by polygon approximation. For each vertices, a four-element string is expressed as a feature that consists of internal angle, distance to next vertex and *x-y* coordinates. In the matching stage, the distance between two strings is calculated as a similarity measure. By this feature, object models are generated, and then organized into a binary tree. Matching can be done as feature to feature or model to model matching. A polygon decomposition sample can be seen from Figure 10 [12].

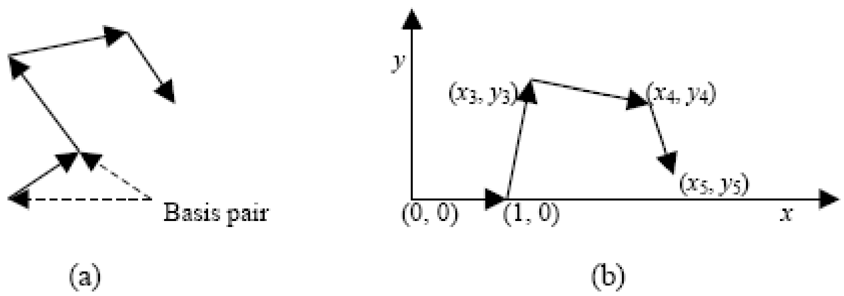


Figure-10: (a) boundary representation of the original shape and (b) polygon representation [12]

III. Smooth Curve Decomposition:

By using this feature, primitives that are called tokens are obtained from the boundary description. Maximum curvature and orientation are calculated for each token, and then similarity is measured by the Euclidean distance.

By indexing the features of these tokens, a feature database is created. For a test shape, first for N tokens of test shape are found, then similar tokens are found by searching N-times the database. Then by a model to model matching algorithm, a similar shape is found [12].

IV. Scale Space Method:

This feature is different than the Curvature Scale Space feature although their names seem to be the same. First, the shape primitives are obtained, and then the segment length, position and curvature tuning value are extracted for each segment descriptor. By these extracted features, a string is created such as:

$S=(s_1^S, s_2^S,.., s_N^S)$ where each *s* represents the set of these features.

In the matching step, a sophisticated model by model matching is employed between these strings [15].

V. Syntactic Analysis:

In this method, the boundary is represented with predefined primitives that are called *codewords*. For example, a *codeword* of a chromosome of a sample DNA is represented as string:

S= d b a b c b c b a b d a c b d b d c d

and a more detailed representation can be seen in Figure 11. For the matching stage, a string matching procedure is applied to find the minimum dissimilarity [10].

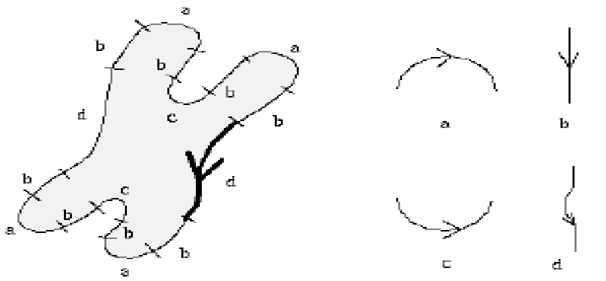


Figure-11: Representation of chromosome shape by syntactic analysis [10]

VI. Shape Context:

For the extraction of this feature, *N* samples are taken from the boundary. These samples do not have to correspond to key points of boundary. Mainly it is preferred to choose those samples uniformly spaced, but this is not mandatory. Then the vectors are considered that originate from a point to all other points are considered. This set of vectors represents the feature. In the matching stage, since it would be inappropriate and expensive to use this set of vectors, a histogram representation is used. The use of a log-polar coordinate system for the histogram gives a better representation. This representation is called shape context. From Figure 12, the operations and shape context can be seen [16].

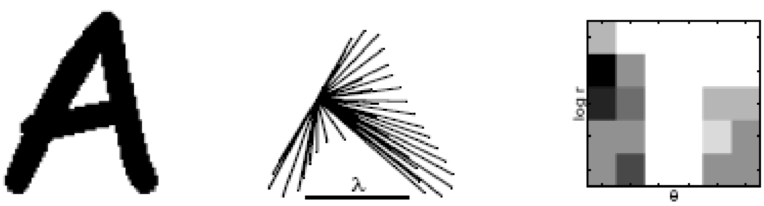


Figure-12: The first image represents detected object; the second image is the set of vectors that originate from one point to all others; and the third image is the shape context [16]

2.3.2 Region Based Shape Representation

Different than contour based shape representations, these features describe the shape region of the detected object of blob. As in the contour representation, region based representation types are also divided into two definitions: global and structural. Global approaches use the whole region of the object while structural methods divide the region into sub-parts.

2.3.2.1 Global Methods:

I. Simple Methods:

- **Fill Ratio[8]:** This feature is the value that is calculated by the ratio of size of the blob segment area (A) over the total minimum bounding rectangle area, defined by its Height *(H)* and Width *(W)*, as:

  $$F = A / (H * W)$$

  The area that is represented as *(A)* is the total number of pixels of the object region in the blob.

- **Compactness [4][8]:** This feature is calculated as the ratio of the square of the perimeter over the area, where the segment area A is the number of object pixels and perimeter P is the number of border pixels of the blob:

  $$C = P^2 / (4 * \pi * A)$$

- **Segment Convexity [8]:** This feature is the ratio of perimeter and square root of the area:

  $$\mathrm{SC} = \mathrm{P} / \sqrt{A}$$

- **Convex Deviation:** This feature is the ratio of segment shape factor which inverse of compactness C and segment convexity SC:

  $$CD = \arctan(1 / C * SC)$$

II. Geometric Moment Invariants:

This feature is usually called Hu Invariant Moment [17]. The published formula of Hu can be shown as;

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q p(x,y)\,dx\,dy$$ where $p, q = 0,1,2...$ and m is the moment value.

This feature is very popular, however lower order moments are not sufficient to describe the shape and higher order invariants are difficult to derive [12].

III. Orthogonal Moments:

The Orthogonal Moments description includes Legendre and Zernike moments. These moments are very similar to Hu invariant moments. The moments are calculated as;

$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_p(x)P_q(y)p(x,y)dxdy$ where P_p and P_q are Legendre and Zernike polynomials.

These polynomials conform complete sets of an orthogonal base, so these moments are called orthogonal moments. According to recent research [10], Zernike moments are more reliable than Legendre moments if the effects of noise are considered.

IV. Shape Matrix:

To compute this feature, the original image boundary is represented as a set of points $P=[p_1,p_2,..,p_n]$ where the deformed shape is represented as a set of points $Q=[q_1,q_2,..,q_n]$. At this stage two vectors are defined as:

$U=[u_1,u_2,..,u_n]$ where $u_i=p_i-g$ and $V=[v_1,v_2,..,v_n]$ where $v_i=q_i-g$.

For both equations, g is an arbitrary reference point. Also u_i can be described as linear combination of two neighboring points:

$$u_i = \alpha_i u_{(i-1)\bmod n} + \beta_i u_{(i+1)\bmod n} .$$

By applying this formula to all boundary points, shape equation:

$AU^T = 0$ is reached.

As U is described above, A is found by this formula. This A matrix is the feature that is called shape matrix. From Figure 13, deformation and matching operations on sample shapes can be seen [18].

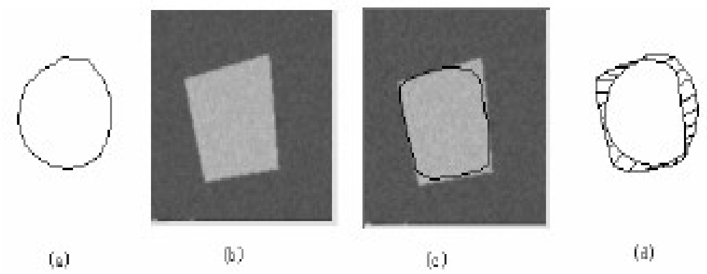


Figure-13: Matching and deformation by shape matrix [18]

2.3.2.2 Structural Methods:

I. Convex Hull:

When a convex region is represented as *R*, its convex hull is the smallest convex region that satisfies the condition:

$R \subset H$ where the difference $D=H-R$ is called convex deficiency.

Then shape can be represented as a string of concavities. From this result, a concavity tree can be extracted as in Figure 14. At this stage, the matching process is done by graph matching or string matching [10].

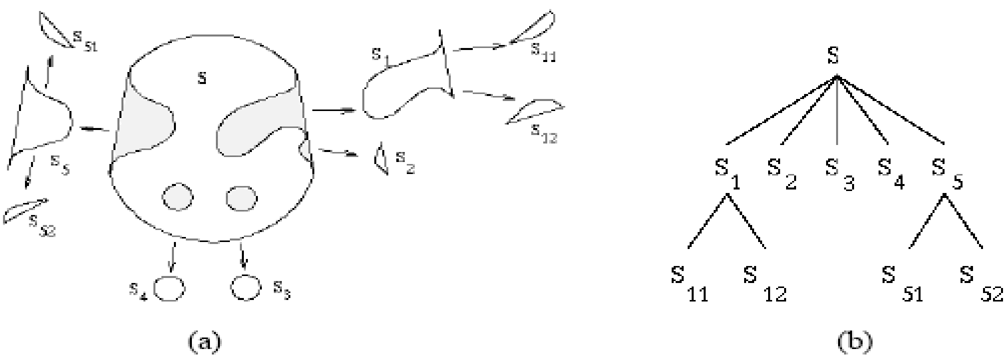


Figure-14: (a) Convex Hull and its concavities and (b) is the tree representation of the feature [10]

II. Medial Axis:

For the extraction of this feature, the region skeleton is obtained as a connected set of medial lines along limbs of the detected object. The idea in this representation is eliminating unnecessary information while taking the necessary one. Medial axis is basically the locus of centers of maximal discs that fit in the shape. A sample of this feature can be seen from Figure 15. After this feature is extracted, the matching operation is completed as a graph matching operation [10].

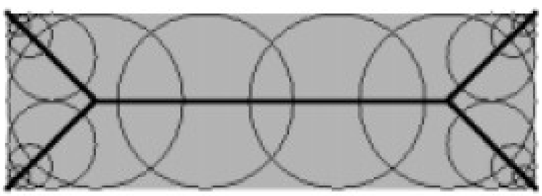


Figure-15: Sample of Medial Axis [10]

2.4 Matching and classification

Matching is the last stage before the classification of the object is complete. Once the selected features are extracted and the description vectors are constructed, matching is done by comparing the obtained description vectors (or measured patterns) with another description vector (or pattern) using a dissimilarity measure. Most of the features described use distance signal determination as a similarity measure. Similar images give smaller distance values while very different images give larger results [10]. Other methods try to model the variability of features by using statistical techniques (Support Vector Machines, Gaussian Mixture Model,…). The application of statistical techniques implies the necessity of large databases in order to build robust statistical descriptions of the object classes.

Then, classification is the stage in which the surveillance system objects are classified into categories [4]. Classification uses the result of the matching process; the objective in this stage is to assign an object type class from the available ones after the matching process. The common approach used is to apply the matching process using all the class models available in the database and then solve the classification problem using the maximum a posteriori criteria.

There are a lot of classifiers that use different methods. Most popular classifiers can be seen from Figure 16 and they are briefly described in this section.

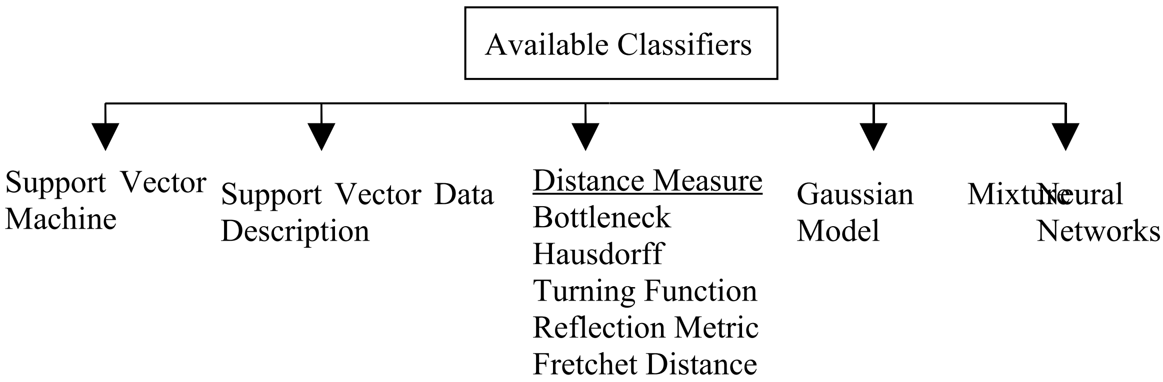


Figure-16: Classifier schemes employed in shape-based object recognition

2.4.1 Support Vector Machine Based Classification:

A Support Vector Machine (SVM) performs classification by constructing an *N*-dimensional hyper-plane that optimally separates the data into two categories. Specifically, a hyper-plane is created in a higher dimensional space and a margin is defined to separate the data. This classifier is very popular since it is more effective than other methods [4][6][7][8]. The result is a signed distance of the test data to the separating hyper-plane. In order to use this data in classification process, it must be converted to a belonging probability of an object to a class. A graphical description of this classifier can be seen in Figure 17.

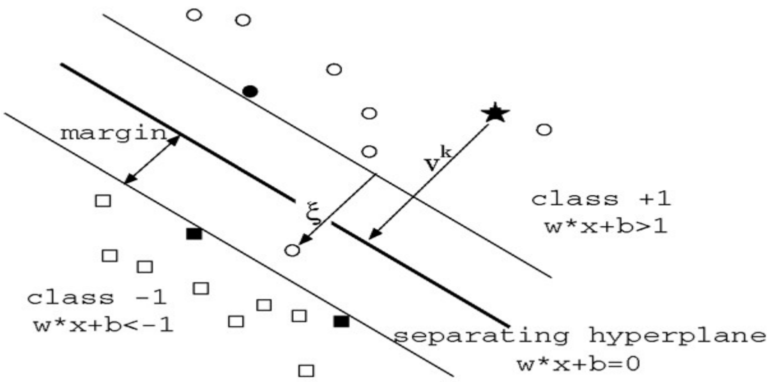


Figure-17: Support Vector Machine Based classifier behavior [19]

2.4.2 **Support Vector Data Description Based Classification:**

This classification method defines a spherically shaped boundary with minimal volume around the training data set; with this property it describes only one class of object. This behavior can be seen from Figure 18. As in SVM, there is a hyper-plane with a certain radius and center. The purpose is to minimize the error that is calculated by radius and center. Basically this classifier is inspired on SVM [8].

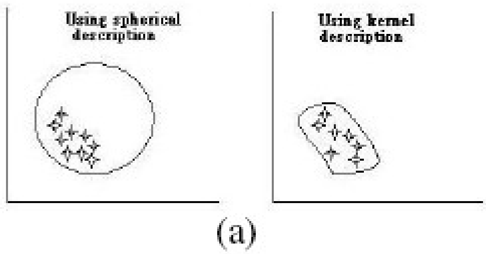


Figure-18: Class boundaries by Support Vector Data Description [8]

2.4.3 **Classification by Gaussian Mixture Model based Classification:**

The Gaussian Mixture Model classifier can use either expectation-maximization algorithm or Gaussian mixture vector quantization. This model assumes that the feature data can be described using a Gaussian mixture distribution and tries to fit a GMM to the training data. In order to understand this classification, graphs for different components can be seen from Figure 19 [20].

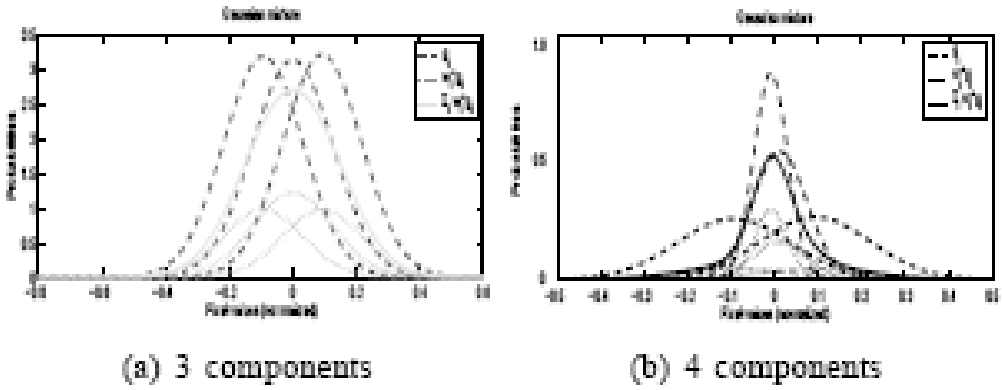


Figure-19: GMM based classification sample [20]

2.4.4 Neural Network based Classification:

A Neural Network (NN) or Artificial Neural Network (ANN) is a non-linear statistical data modeling tool inspired by the way biological nervous systems, such as the brain, process information. Its most important element is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. It can be used to model complex relationships between inputs and outputs or to find patterns in data.

The typical Neural Network classifier used in object classification is a feed-forward Back Propagation Neural Network classification [8]. The simple processing elements of this classifier (neurons), their connections and their activation function determine the output of the network depending on the inputs. Basically this determination can be defined as:

$$f(x) = K(\sum_i w_i g_i(x))$$

where *f(x)* is the composition of other functions, *K* is a predefined function and *g* is the data vector [22].

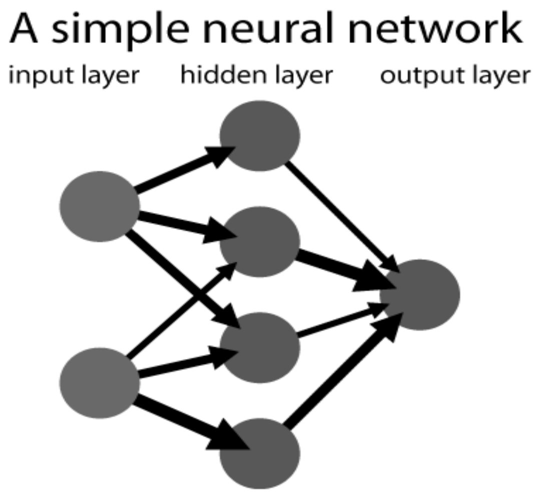


Figure-20: Simple Neural Network representation extracted from [22]

2.4.5 Distance Measure based Classification:

Several methods have been proposed depending on the feature used. In this section we describe the most relevant to our work:

- **Bottleneck Distance:** This method is a classifier based on the distance between two point sets. It searches for the minimum over all one-to-one correspondences of two data sets of the maximum distance. In the classification part, parametric search or nearest neighbor algorithm can be applied [12].
- **Hausdorff Distance:** This method is a dissimilarity measure; defined as maximum distance over all points' distances between training and template data [12].
- **Turning Function:** This method is a position function that can be derived as tangent, acceleration, tangent angle, cumulative angle…etc. The dissimilarity measure is the function of scale rotation and shift of polylines among other [12].
- **Reflection Metric:** This classifier is defined as unions of curves that are converted to real functions and then compared using integration; the result is the similarity measure [12].
- **Fretchet Distance:** This is the method that when walked among curves simultaneously the distance measure between corresponding points along the curves gives the result [12].

2.5 Databases/Datasets

In order to apply these features and classifiers, different databases and datasets have been proposed by different authors. In this section, a summary of the most important ones is described:

2.5.1 Database 1

This database consists of different object classes where each class folder contains rotated images of the class object. There are 17 object classes in this database where each object class folder contains 128 images. The image samples from different classes of this database can be observed from Figure 21 [23].

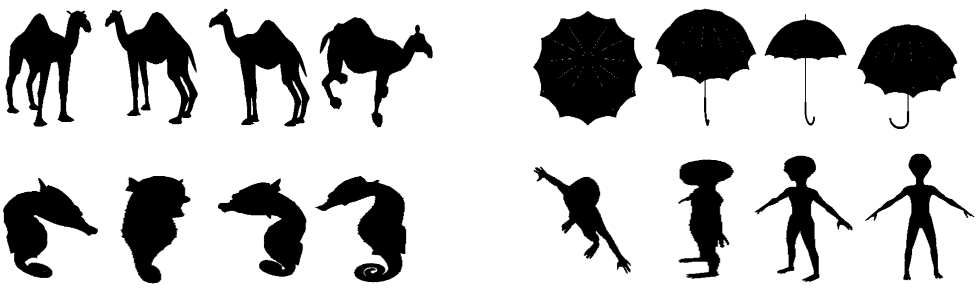


Figure-21: Database different object class image samples

2.5.2 Database 2

This database contains 23 object class folders and each folder contains at least 20 images. Different than the first database, this database contains different shapes of the same class object. The image samples from this database can be seen from Figure 22 [24].

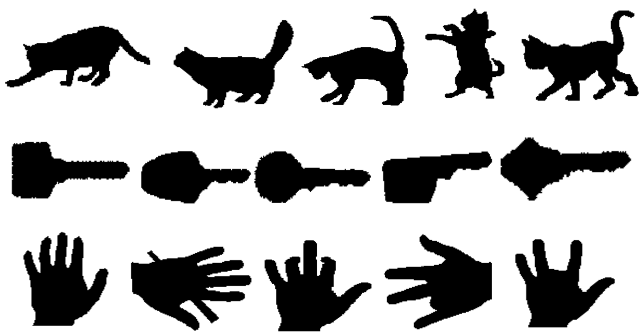


Figure-22: Image samples from second database

2.5.3 Database 3

With the purpose of detecting people from video-surveillance data, another database with two classes is used where the classes are Human and Object classes where the samples can be seen from Figure 23. The human shapes class contains 170 samples and the object shapes class contains 117 samples [25].

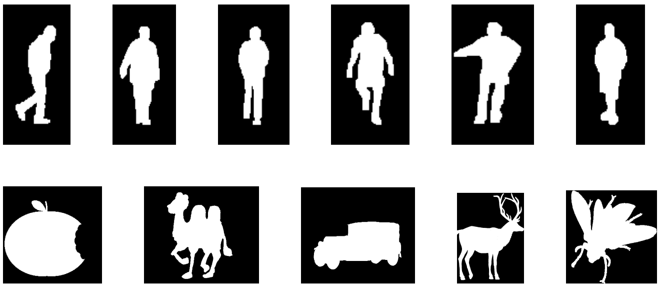


Figure-23: Human and Object Class database image samples

Then foreground mask data of video-surveillance system is selected to match with these two-object class databases that are obtained from IPPR Contest motion segmentation dataset[1]. The samples from this frame masks can be observed from Figure 24. This surveillance system data contains 299 frames where the used foreground masks are manually annotated.

Figure-24: Frame foreground mask samples of a video-surveillance system

2.5.4 Database 4

For the final scenario, the used frames are from a video-surveillance system that contains stolen or abandoned objects in the scene and for this detection, a background image and object masks are given. These datasets are taken from existing databases of VPU. Abandoned object database contains 26 frames, one background image and 26 object masks. Stolen object database contains 11 frames, one background image and 11 object masks. The samples from the abandoned and stolen object database can be observed from Figure 25 and 26.

[1]Available at http://media.ee.ntu.edu.tw/Archer_contest/

Figure-25: Background, two frames and two object masks of the abandoned object dataset

Figure-26: Background, two frames and two object masks of the stolen object dataset

3 Shape-based features selected

3.1 Introduction

In the previous section, all the available shape-based features are explained via their general properties. In this section, some features are selected in order to use them in the proposed project. In this selection process, the idea is to select features with different properties (from different categories) in order to compare them by analyzing their robustness and success in different object classification schemes (that is, combination of feature-matching technique). The selected features are simple-descriptors, boundary moments, distance signal and shape context. In the following sub-sections, these selected features are described in terms of the motivation yielding to their selection and their characteristics as advantages or disadvantages. Additionally, examples are given for some features in order to provide a better understanding of them.

3.2 Descriptors selected

3.2.11 Simple Descriptors:

As it is obvious from this name, these features are very simple to implement and their computational cost is very low; basically these are the reasons why they are chosen. Some features are contour based while the others are region based. Since the results of these features are simple quantities about the shape of the detected object, it is not reasonable to use these features alone. The expressiveness of these features is very low and they are useful in easy object classification problems. So it seemed logical to bring these features together to build a feature vector in order to test their accuracy in simple problems. Using a combination of contour based and region based features, it is considered that this feature vector could bring a powerful approximation to shape-based feature definition. As a result, for further operations such as matching and classification, this feature vector of simple features is applied.

In the extraction of these simple features, basically two properties of the detected object are used: area and perimeter. Area is a characteristic of detected object region of the detected object that refers to the total number of object pixels of the blob [8]. Area of a detected object changes as the size of the object changes, therefore the area of an object is not used alone as a feature. Similar to area, perimeter is a contour-based property of detected object where the total number of boundary pixels of the blob is the result [8]. Also the perimeter value depends on the size of the detected object. Therefore area and perimeter quantities are used with other size dependent values in order to make the features size independent.

Compactness:

Compactness is the feature that is the ratio of square of perimeter over area of the blob. Area and perimeter quantities in this definition are previously defined properties of detected object region and contour. Perimeter and area features are size dependent. Since their ratio is taken in this feature, it is not necessary to consider the size problem anymore, so this feature can be used for matching process effectively as a coefficient of the feature vector [4] [8]. This feature can be represented as a formula where the perimeter is *P* and area is *A:*

$$C = P^2/(4\pi A)$$

In order to understand the efficiency of this feature, compactness of different objects depicted in Figure 27 can be observed in Table 1.

Figure-27: Sample shapes and their compactness results

	$C = P^2/(4\pi A)$
Apple	1.3972
Chicken	2.8725
Cup	2.3374
Tree	3.0471

Table-1: Compactness results for shapes in Figure 27

Height by Width Ratio:

This feature can be found by the name elongation is some papers [6]. This feature cannot be classified as a contour based or region based feature, because minimum bounding rectangle of the detected object is used in this feature while a minimum bounding rectangle can be found from the contour or region of the detected object.

As it is clear from its name, this feature is the ratio of height and width of the minimum bounding rectangle of the blob. Where height is represented as H and width is represented as W; elongation or height by width ratio feature can be found as:

$$E=H/W$$

This feature can be useful while matching two shapes of same class. However two shapes from different classes can have the same elongation value, so it is for our advantage to use this feature in the feature vector instead of using it alone in the matching step [4][8]. Height by weight ratios of sample objects in Figure 27 can be observed in Table 2.

	$E=H/W$
Apple	1.0000
Chicken	1.0404
Cup	1.2878
Tree	1.0787

Table-2: Height by width ratio results for samples in Figure 27

Rectangularity:

Rectangularity is the ratio of the detected object area over area of the minimum bounding rectangle [10]. As the area features are used in this feature, this is one of the region-based features. Where H is the height and W is the width of the minimum bounding rectangle and A represents area, the formula that defines this feature is:

$$R = A/(H*W)$$

Rectangularity results for sample objects in Figure 27 can be found in Table 3.

	$R = A/(H*W)$
Apple	0.4278
Chicken	0.4255
Cup	0.3063
Tree	0.2855

Table-3: Rectangularity results for sample objects in Figure 27

Segment Convexity:

Similar to the Compactness feature, Segment Convexity is defined as the ratio of the perimeter of a blob over the square root of its area. The formula of this feature can be defined as:

$$SC = P/\sqrt{A}$$

This feature basically determines the convexity of the detected object [8]. Segment Convexity results for object samples in Figure 27 can be seen in Table 4.

	$SC = P/\sqrt{A}$
Apple	4.1985
Chicken	6.0157
Cup	5.4197
Tree	218.6413

Table-4: Segment Convexity results for sample objects in Figure 27

Convex Deviation:

The Convex Deviation feature uses the Segment Convexity feature and Compactness feature previously defined [8]. The formula for this feature can be defined as:

$$CD = \arctan(1/(C * SC))$$

Some convex deviation results can be found in Table 5 for the sample objects in Figure 27.

	$CD: \arctan(1/(C*SC))$
Apple	0.1688
Chicken	0.0578
Cup	0.0788
Tree	0.0015

Table-5: Segment Convexity results for sample objects in Figure 27

3.2.21 Boundary Moments:

This feature is determined as moments of the boundary points. While the boundary is represented as:

$z(i) = p_i$ where p represents boundary point coordinates for $i=1,2,..N$ and N is the number of boundary points.

Moments are calculated by the following formulas:

$$m_r = \frac{1}{N}\sum_{i=1}^{N} z(i)^r \text{ and } \mu_r = \frac{1}{N}\sum_{i=1}^{N} (z(i) - m_1)^r$$

where m_r is the rth moment of the boundary and μ_r is the central moment. Since these moment results are size dependent, it is necessary to normalize them for size invariance. The normalization procedure is applied as:

$$\overline{m}_r = \frac{m_r}{(\mu_2)^{r/2}} \text{ and } \overline{\mu}_r = \frac{\mu_r}{(\mu_2)^{r/2}}.$$

According to these moments, shape can be represented as:

$F_1 = (\mu_2)^{1/2} / m_1$, $F_2 = \mu_3 / (\mu_2)^{3/2}$, $F_3 = \mu_4 / (\mu_2)^2$ where these F values are the features.

One of the reasons to select this feature is its simple implementation while another reason is to desire to use a moment based feature to compare its performance with other features. Advantages of this method are its simplicity as it reduces the dimension of boundary representation and the simplicity of the operations performed while its disadvantage is its failure for implementation on higher order moments [10]. If it is not necessarily important to implement higher order moments in this feature, it can be applied successfully. For the object samples in Figure 27, the boundary moment results can be observed in Table 6.

	Boundary Moments Feature Vectors
Apple	(2, 3, 9, 24, 67, 196, 583, 1756, 533, 16303)
Chicken	(2, 5, 13, 38, 112, 339, 1042, 3244, 10192, 32288)
Cup	(2, 5, 15, 46, 149, 494, 1668, 5682, 19488, 67155)
Tree	(2, 5, 14, 44, 139, 456, 1545, 5354, 18918, 67946)

Table-6: Boundary Moments Feature Vectors extracted from sample objects in Figure 27

3.2.31 Distance Signal:

This feature is found by first determining the center of mass $C_m(x_m, y_m)$ as:

$$x_{C_m} = \frac{\sum x_i}{n} \text{ and } y_{C_m} = \frac{\sum y_i}{n}$$

where x_i and y_i are the boundary pixel coordinates. Then starting from the top center point on the boundary of the shape, boundary coordinates are ordered in clockwise direction by checking the 8-neighborhood of the current boundary point. Subsequently the distance signal is defined as:

$DS=\{d_1,d_2,...,d_n\}$ by applying the operation $d_i=dist(C_m,p_i)$ on the ordered boundary points where p_i represents those boundary points and *dist* function represents Euclidean distance.

In order to make this feature size independent, size of the distance signal is fixed to a constant value by operation:

$$\overleftrightarrow{DS}[i] = DS\left[i * \frac{N}{C}\right]$$

where C is a constant chosen by the user. Then the resulting scaled distance vector is normalized as:

$$\overline{DS}[i] = \frac{\overleftrightarrow{DS}[i]}{\sum_{1}^{n} \overleftrightarrow{DS}[i]}.$$

This normalized distance signal vector is the feature vector of the proposed shape [7]. Also, this distance signal vector can be represented as a graph and the matching operation can turn into a graph matching procedure. Human shape samples and corresponding distance signal graphs can be observed from Figure 28; the shapes above the plots correspond to human shapes and below the consequent distance signal graphs are shown.

Figure-28: Distance signals graphs for sample human shapes [7]

This feature belongs to the contour-based category; the idea under its selection is to implement a different and more complex feature in order to compare its success with other features. Since a boundary moment is a contour based feature too, the first comparison is based between distance signal and boundary moments methods results.

3.2.41 Shape Context:

Until this point, only global shape based features have been selected. Therefore in order to compare the efficiency of the structural shape-based features Shape Context is selected. For the extraction of this feature, *N* samples are taken from the boundary that does not necessarily correspond to key points of the boundary but they can be uniformly spaced for test and training shapes as in Figure 29. Then vectors originating from an arbitrary point to all other boundary points are considered as in Figure 30. This set of vectors represents the shape in this feature technique.

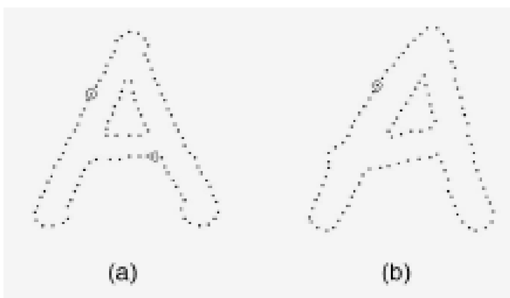


Figure-29: Sample shapes used for Shape Context representation

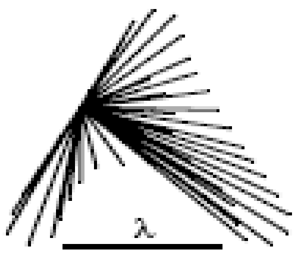


Figure-30: Representation of shape in Figure 29 by Shape Context as set of vectors

Nevertheless it is expensive to match those vectors with the test shape context vectors [16], so the histogram representation method is proposed for this feature. In order to have the best representation for the histogram, a log-polar coordinate system is applied. These log-polar histograms are the final representations by shape context feature. Sample log-polar representations of the proposed shapes in Figure 29 can be observed from Figure 31. Since this feature is a log-polar histogram, it cannot be used as a feature vector. So this feature has its own matching methods.

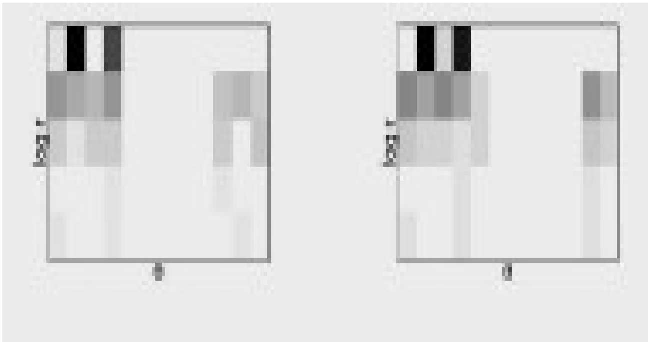


Figure-31: Log-polar Histograms (Shape Context Features) of sample shapes in Figure 29 [16]

Basically the matching operation is based on the distance between two histograms. However since this method uses a distance measure, and as based on the classes a class model cannot be created in order to make the matching process less expensive, training and test objects are compared one by one. So the disadvantage of this feature is its great cost [16]. The matching procedure applied for this feature is described in section 4.2.3.2.

4 Matching techniques selected

4.1 Introduction

In section 2, the common matching techniques (also known as classifiers in the literature) used for 2D object classification using shape-based features are explained by their general definitions. In this section, the selected matching/classification techniques to use in the project are described in detail.

According to the most popular techniques used in the literature and their compatibility with the selected features: Support Vector Machine Based, Gaussian Mixture Model and Distance Measure Based classifiers are selected for this project. Input data of each classifier is a feature vector determined by the different shape-based feature selected. In order to compare the performance of the classifiers, all the selected features are used in each classifier.

In the classification stage, the procedure applied is the same for each classifier. This procedure can be summarized in two phases: training and test. In the training phase by extracting the selected features and applying the selected classifiers, class models are created for each object class of the database. In the test phase, features of the test objects are extracted and basically the feature vector of the test shape is matched to the previously created models in the training phase. Subsequently the object class model with minimum dissimilarity or maximum similarity is chosen as the class of the proposed object, therefore the classification process is completed.

4.2 Matching/Classification techniques selected

4.2.11 Support Vector Machine Based Classifier

According to recent studies [19] [25], the performance of Support Vector Machine (SVM) classifier is better than other popular classifiers such as Neural Networks classifiers. In this approach, an optimal hyper-plane focuses on training data that separates different classes is created. These training data is called support vectors and the training data that do not fit into this definition are discarded by defining a margin where training data cannot be separated. Therefore Support Vector Machine based classifiers separate the maximum margin with a hyper-plane.

This classifier is based on separating two different classes; however this method can be applied to classify more than two classes by combining two-class classifications. In order to understand this method, the two-class classification problem is described. In the training stage of the classification process, input vectors are proposed as $\{x_i, y_i\}$ where x_i are the feature vectors of the class object shapes and $y_i \in \{-1, +1\}$ where i=1, 2... N and N is the number of shapes in the proposed class. y_i values are also called labelers since they label the input as class +1 object feature or class -1 object feature; first class (w_1) is labeled by y_i=+1 and second class (w_2) is labeled as y_i=-1.

According to previous descriptions, a separating hyper-plane is defined as:

$$f(x) = w.x + w_0 = 0$$

where w is the weight and w_0 is the bias. Separation of two classes by optimal hyper-plane can be seen from Figure 32. In order to find w and w_0, two inequalities are proposed as the defined hyper-plane separates +1 and -1 classes:

$y_i(w.x_i + w_0) \geq +1$ for w_1 labeled as y_i=+1,

$y_i(w.x_i + w_0) \leq -1$ for the second class w_2 labeled as y_i=-1.

The models are created in the training phase by using input feature vectors as the proposed weight and bias values; these values represent the property of that class model.

In the test phase, the test shape feature is extracted and sent for matching to the following operations; the two previously described equations create two hyper-planes while the needed result is an optimal hyper-plane and to make the decision about the test object as which side of the hyper-plane does it belong. In order to reach that optimal hyper-plane, the maximum margin is found by:

$$w.x_i + w_0 = \pm 1.$$

The margin can be expressed as $2/\|w\|$, so optimal margin is found by solving the equation:

$$\min \{\frac{1}{2}\|w\|^2\} \text{ where } y_i(w.x_i + w_0) - 1 \geq 0.$$

If this definition is converted to the Lagrange formula and inner product is applied, then SVM classifier becomes:

$$f(x) = \sum_{i \in S} \lambda_i y_i K(x_i x) + w_0 .$$

The decision about the test shape is done according to this function where input value is the feature vector of the test shape [19][27].

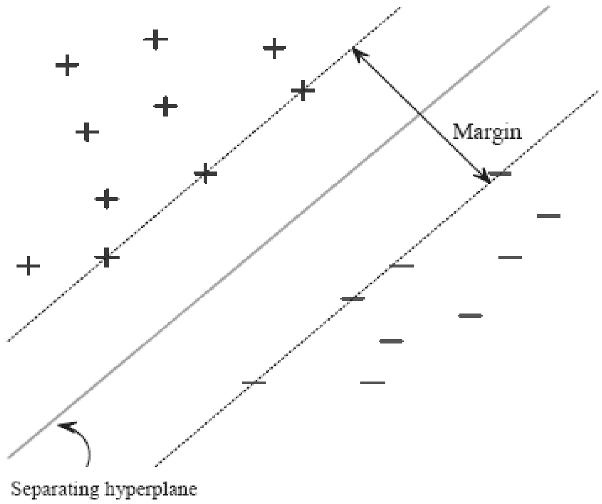


Figure-32: Support Vector Machine Based Classifier and optimal hyper-plane [19]

In case of multi-class classification, the acyclic graph method is used and the algorithm of this method can be observed in Figure 33 for a three object classes' classification case. The method followed in the figure can be described as follows: in the beginning, the first and second classes are matched by SVM, then their result (represented as 1/2) is matched to the result of SVM matching result of second and third classes (which is 2/3 in the figure), therefore final SVM matching result of (1/2) and (2/3) is the decision result of the three-class matching problem [19].

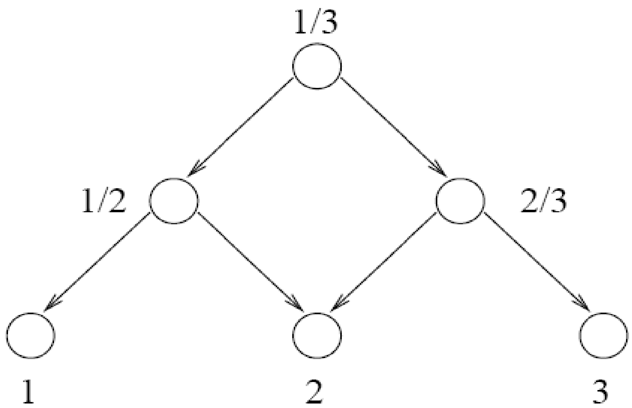


Figure-33: Acyclic Graph approach to multi-class classification in SVM [19]

There are many advantages of this classification method; the primary advantage is its effective model generalization property. Since the classification process is dependent on the input data, the reached classification is clear. Also SVM classification can interpret the confidence measure with the proposed decision. A disadvantage of this classification can be observed in more than two object class classification, since if the matching is done by the acyclic graph method for more than two classes, then this classification process takes more time to give the final decision [19].

4.2.2 Classification by Gaussian Mixture Model

As in the SVM classifier, the Gaussian Mixture Model (GMM) classification consists of training and test phases. In the training phase, feature vectors of each database class shapes are extracted and hold in the feature vector:

$$X_i = \{x_1, x_2, ..., x_m\}$$

where *i* represents the class number and x_m is the feature vector and *m* is the number of features in that class. Then, models for each class are defined as:

$$p(X_i \mid \theta_i) = \prod_{m=1}^{M} \sum_{j=1}^{J} p(z_i) b_j(x)$$

where *J* is the number of Gaussians, and θ_i defines model parameters as:

$$\theta_i = \{p(z_i), \mu_i, \sum_i\}$$

where $p(z_i)$ is the weight, μ_i is the mean vector and $\sum_i$ is the covariance matrix of *i*th class. The *b(x)* term used in the previous formula is defined as:

$$b_i(x_m) = \frac{1}{(2\pi)^{2/D} \cdot |\Sigma_i|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x_m - \mu_i)' \, \Sigma_i^{-1}(x_m - \mu_i)\right\}$$

where *D* represents the dimension of the feature vector x_m. After the models are created, each model is represented by a Gaussian Mixture Model:

$$\lambda = \{p(z_i), \mu_i, \sum_i\} \text{ for } i\text{th class.}$$

In the test phase of the GMM classifier, features of the test shapes are extracted and as in the training phase, the features are collected in vector:

$$X = \{x_1, x_2, ..., x_T\}$$

where *T* is the number of test shapes. Where *S* is the total number of class model numbers that are created in training phase and λ_k is the GMM model of the *k*th class:

$$\hat{S} = \arg\max P(\lambda_k \mid X) \text{ for } 1 \le k \le S.$$

As is clear from the previous formula, each class models λ_k are compared to the feature vectors of the test shapes and the class type with maximum probability is selected as $\hat{S}$

[15] [19]. GMM based representation of the created models can be observed from Figure 34.

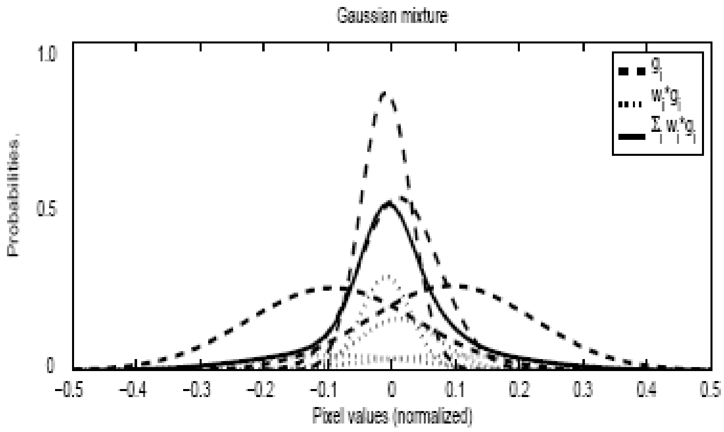


Figure-34: GMM based object class model representation [15]

Since the statistical measure (probability functions) is used in the GMM based classification procedure, the results detected by this method are more accurate.

4.2.3 Distance Measure Based Classifier

4.2.3.1 **Generic Distance Signal**

In this classifier, the operation is based on the similarity measure between object shapes. In the classification of distance signals DS_A and DS_B, similarity between these two shapes can be found as:

$$Dist_{AB} = \sum_{i=1}^{n} |\overline{DS}_A[i] - \overline{DS}_B[i]|.$$

Classification is done by calculating *Dist* between test shape (*A*) and database (*B*). If the tested object is represented as *O*, and two template database shapes are *I* and *P*; object *O* is classified as *P* when:

$$Dist_{OP} \leq Dist_{OI}.$$

A classification result by this classifier can be observed from Figure 35. Most left object is the test object in the figure and all other twelve objects are database shapes. According to the distance signals shown under each shape, the matching shape is found by distance based classification and represented in the box [5].

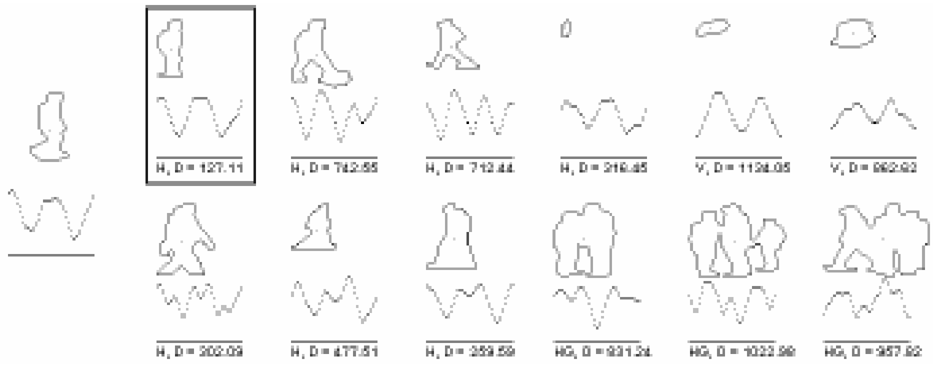


Figure-35: Distance measure based classification sample [5]

4.2.3.2 Shape Context

The Shape Context feature is classified by a different procedure where the idea is to measure the distance between two shapes. The Shape Context feature vector set is converted to log-polar object shape histograms, so the matching becomes the measure of the similarity between two histograms. If two histograms are defined as *g(k)* and *h(k)*, then their similarity is measured by:

$$Cs = \frac{1}{2}\sum_{k}^{K}\frac{[g(k) - h(k)]^2}{g(k) + h(k)}$$

where the value of *Cs* ranges from 0 to 1. In addition to this distance measure, also tangent angle dissimilarity is needed to be measured. Where θ_1 and θ_2 are angles of unit vectors that connect the center of the unit circle with half length of the core, local appearance is measured as:

$$C_A = \frac{1}{2}\left\| \begin{pmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{pmatrix} - \begin{pmatrix} \cos(\theta_2) \\ \sin(\theta_2) \end{pmatrix} \right\|.$$

The matching cost is calculated as:

$$C = (1 - \beta)Cs + \beta C_A$$

where β is a weight constant. The final decision is made by the result of matching cost *C;* if this value is lower between a test shape and a database shape and higher for all other database shapes, then they are similar shapes. In Figure 36, on the left of the figure database objects and on the right corresponding log-polar histograms can be observed. Each row contains similar shapes, so the similarity or difference between log-polar histograms of similar shapes can be observed [13].

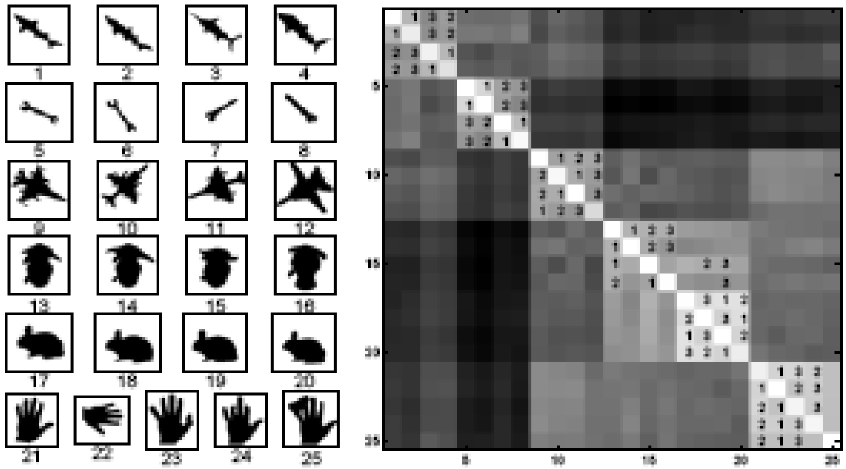


Figure-36: Database and corresponding log-polar histograms [13]

5 System integration

5.1 Introduction

In this section, the framework that integrates the different techniques developed in the project is described by defining each stage, the operations performed inside and the available techniques in each stage. Then, it is used to compare the implemented techniques in the defined experiments in chapter 6.

The proposed framework is focused on 2D object classification in a video analysis general scenario and it is not designed for a specific application. The final purpose of the framework depends on the object detection method applied (e.g. object detection using Background Subtraction for video-surveillance application).

As a general description, the shape-based object classification procedure begins with object detection. In this stage, raw frames are given as input and after the detection is complete, binary foreground masks are produced. Since these binary foreground masks can contain noise, wrong detected pixels or discontinuities on object shapes, it is important to pre-process these binary masks to make them ready for further operations. Then object blobs are extracted from binary masks to make the feature extraction easier. Afterwards shape-based features of the blob are extracted. In the matching phase, these features of the detected object are compared with feature models based on database shapes. Finally a decision is made according to the result of the matching operation. The block diagram of the proposed framework is depicted in Figure 37.

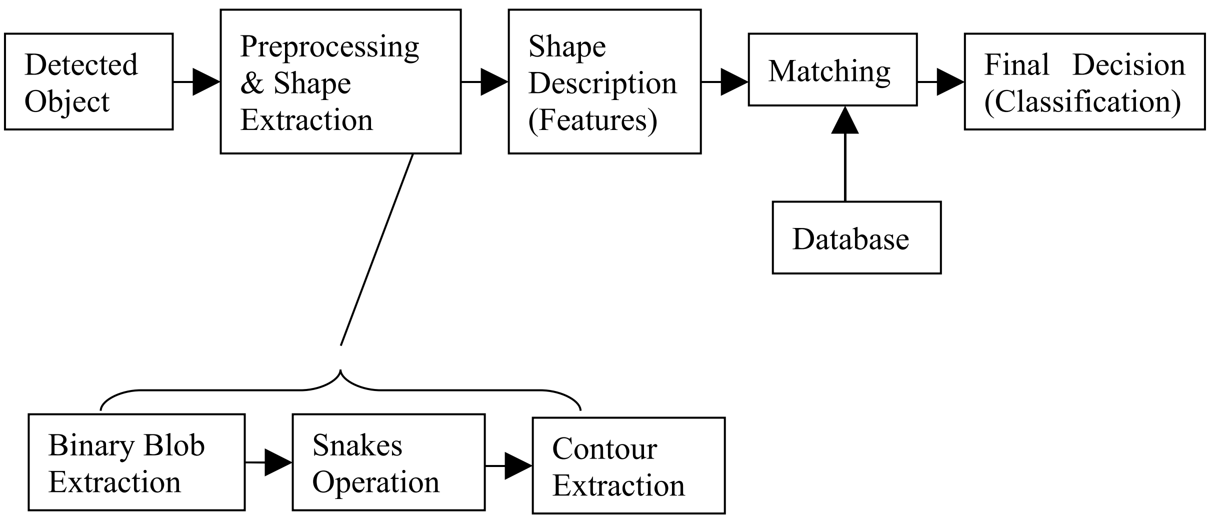


Figure-37: Information flow diagram of the proposed project

5.2 Object Detection

The input of this stage is a raw frame (Figure 38) that can contain moving objects. The objective in this stage is to detect the existing objects if there are any in the raw frame so this object can be used for further classification stages and it can be categorized depending on the available classes in the database. For this purpose, the Background Subtraction method is the most common method used in video-surveillance systems under two constraints: stationary camera and low illumination changes of the environment. The technique used in this project is based on the noise introduced by the camera [28]. It has been provided by the Video Processing and Understanding Lab and it is briefly described in Appendix 1.

The result of this process is depicted in Figure 39. As it can be observed in this figure, the objects in the frame are detected and represented in a binary mask. However as it can be seen from the figure, after this subtraction operation, there are some wrong detected pixels and noise effects. Since this binary mask cannot be used in classification stages with noise, it has to be pre-processed.

Figure-38: Raw frame and background image

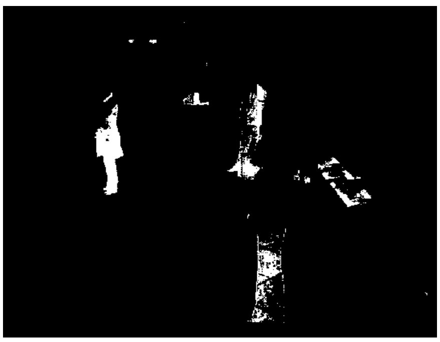


Figure-39: Binary mask after Background Subtraction

5.3 Pre-processing & Shape Extraction

The pre-processing stage takes the binary foreground mask with noise, shadows and wrong detected pixels as input in order to operate on the detected objects in further stages. In the pre-processing stage, the first applied operations are morphological operations: namely, dilation and erosion. By starting with erosion and then dilating the resulting binary image, noise effects and wrong detected pixels are reduced. After this process, applying more dilation operations can make the object pixels more connected and better, but the object contour is deformed. In order to correct this deformation, other pre-processing techniques should be applied. Binary foreground image mask after pre-processing can be seen in Figure 40.

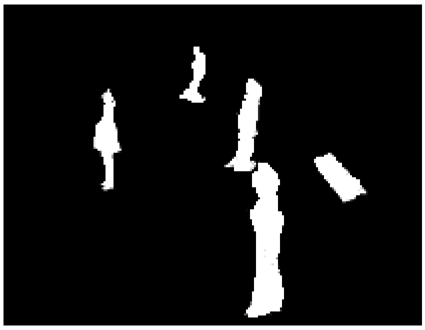


Figure-40: Binary image mask after pre-processing

Afterwards blob extraction is applied to the binary foreground image mask in order to analyze each object shape independently. For this purpose, the image mask is analyzed to infer how many objects does it contain, and then each detected object is converted into binary blobs. These binary blobs are the input of the feature determination process. Some binary blob samples can be seen from Figure 41.

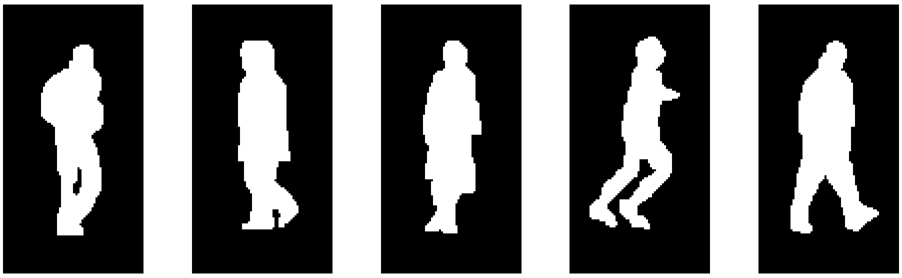


Figure-41: Binary blob samples

After these phases, it is possible for the contour of the detected object to be deformed, but for a good classification it is important to have a clearly presented contour. To correct these deformations, a Snakes operation is applied on the bounding box contours for a certain iteration value. More information about the Snakes operation can be found in Appendix 2.

In the contour extraction process, there are many methods defined. These methods are based on a 4 or 8 neighborhood of the object pixels with background, where the background and detected object shape coincide; the contour of the object is extracted. The result of this extraction process is a 2D vector of coordinates of boundary pixels.

5.4 Shape Description

This is the stage where the features of shapes of the detected objects are extracted in order to be used in the matching operation. There are different features based on different properties of the objects that can be extracted in the proposed framework. These features belong to two different types: region based and contour based. Also the techniques used in the features separate them as global and structural; while global techniques use the proposed information as a whole, structural features divide the shape or the region into segments and operate on these segments.

With the purpose of comparing different kinds of features in order to decide which one is better under certain conditions, different kinds of features can be selected in this module. From structural features, the shape context feature is selected. From region based features, compactness, rectangularity, segment convexity and convex deviation are selected. From contour-based features, elongation, boundary moments and distance signal are selected. While compactness, rectangularity, elongation, segment convexity and convex deviation features are very simple to compute, boundary moments and distance signal methods are chosen because their nature is very different than other simple features.

Simple features take the region or contour information of the binary blob and by simple mathematical calculations between perimeter, area or minimum bounding rectangle size, simple values are reached that are not enough to be used to match a shape with another alone. So these resulting values are put together to build a feature vector and this vector is used in the matching stages.

In the boundary moments, a vector that contains boundary pixel coordinates is used as input. As it is obvious from its name, this feature calculates moments of this contour vector by different properties such as normal moment or central moment, also the order of moments can be changed.

Another different feature is distance signal. As for the boundary moments, the input of this feature is also a vector of contour pixel coordinates. First, the center of mass of the object is determined, and then by finding the distance from the center to all boundary points and after some normalization operations, a signal that describes the shape is reached. So basically the distance from the center of mass to all boundary points are the feature for this model.

The most different feature selected is shape context. In the determination of this feature, *N* samples are taken from the boundary of the binary blob, and then by complex mathematical operations, histograms are obtained from boundary information. Basically these histograms are the features for this method.

5.5 Matching

This is the stage before the final decision is taken by the system. The inputs of this stage are the previously extracted features and the available models in the database. By using these features, as the primary operation, models are created from database object classes. Then in the matching stage, the features of detected objects are compared to the available object models.

In this framework, Support Vector Machine (SVM), Gaussian Mixture Model (GMM) and Distance Measure (DM) based classifiers have been implemented. SVM and GMM based classifiers have been selected due to their popularity of success. The distance classifier has been selected because there are some features, like Shape Context, that can only be used with this type of classifiers if we want to adjust to the metrics proposed by the authors [16].

In SVM and GMM based classifiers, training and test phases are very similar. Both create models based on description vectors from the available classes in the database as described in chapter 4. Then, the similarity between the corresponding description vectors of the detected object is compared as described in chapter 4. However, although the training and test phases are similar in the SVM and GMM, the internal operations involved to build these classifiers are very different. This difference creates the difference between the results of these two classifiers.

The Distance Measure based classifier is basically used for shape context, but it can be easily extended to compare other feature description vectors. This classifier presents a high computational cost as the distance signal comparison is done by using all points of the shape description (no models are created). The detected object feature result is compared to the results of each object in each class of the database one by one.

5.6 Classification

In the classification stage, the results of the matching stage are detected for each model. The objective of this stage is to select the model with maximum similarity with the tested shape feature. The maximum a posteriori criterion has been selected to choose the class of the object under evaluation [29]. This criterion is based on finding the class that produces maximum likelihood in the matching process from the available ones. Primarily, likelihoods for each model are calculated, then the model which maximizes the likelihood is found and a decision is made.

This classification method is applied to the GMM classifier, however in the SVM classifier case; there is no need for an extra algorithm since the classifier itself selects the class models that match with the test shape with maximum similarity. In the Shape Context based classification procedure, since there are no models, test shapes are matched with all database shapes, and the result of the matching procedure is the difference between two images. Therefore the reverse of the previous algorithm is used for classification, which is the minimum likelihood is searched. The result of this search points to the similar shapes of the database where the objective is to find the object class that matches. So the number of minimum likelihoods of each object class are taken, the class with maximum number of minimum likelihoods is selected as the class model of the test shape.

6 Experimental results

6.1 Test scenarios

The final objective of the developed project is to compare the performance of the selected features and classifiers with the final objective of finding the best feature-classifier combination for three different scenarios. According to each scenario, accuracies by different feature-classifier combinations are computed.

These scenarios are based on different object detection techniques and consequently the application results of features and classifiers vary for each scenario. In order to test the selected features and classification techniques on different objects, the first scenario is determined as the typical database classification problem. Basically this scenario consists of a database composed of different objects classes, the data is divided into training and test sets, training data is used to acquire the class models, test data is compared against all available class models (by matching their features) and finally the maximum a posterior criteria is used to decide the class of the objects under evaluation.

The second scenario is People Detection in video-surveillance video. This scenario is very similar to first one; it contains a two-class database that has human and other shape classes. Primarily, the models of the classes are computed. Then, the frames taken from video-surveillance video are analyzed to determine the objects of interest and they are matched using the model classes to classify them as people or non-people blobs.

The final scenario is slightly different than the first two scenarios. The application of shape-features to classify stationary objects in video-surveillance into Abandoned Object or Stolen Object classes is proposed. The main difference with the previously defined scenario is that it does not use any kind of database or object class. The objective in this scenario is to discriminate the stationary objects detected in the video-surveillance frames between abandoned and stolen objects. Primarily, the frames are analyzed in order to determine the objects of interest. Then, shape features of the objects of interest in the mask, current frame and background are extracted. Finally, if the shape-feature of the objects in the current image matches with the mask ones, the object is determined as abandoned or stolen as the matching is performed between the mask and background features.

6.1.1 Database classification

The objective in this scenario is to compare the efficiency of features and classifiers and to find the best combination of feature-classifier where the best accuracy is reached without any time constraints. Therefore, two databases are selected where one database contains several object images that are rotated in each class. The other database contains more clear shapes of the same class objects with small differences in each class.

The selected classifiers are Gaussian Mixture Model and Support Vector Machine except the classifier of shape context method. With each classifier, three feature vectors are used; simple features combination vector, boundary moments and distance signal. Shape context is one of the chosen features however since this feature is not compatible with other classifiers, it is explained as another classification technique. The classification process is repeated for each feature, so the process is done three times for each classifier.

Basic operations in Gaussian Mixture Model classifier can be observed from Figure 42. In the GMM classification for this scenario, images under each class of the training dataset are processed by morphological operations in the training phase. Then shape data of the objects are extracted and the proposed feature technique is applied. For each object class, feature vectors are collected and finally a class feature matrix is created. By applying GMM based operations on the proposed class feature matrix, a class model is created. The same operation is repeated for each class, and these models are collected. In the test phase, the same pre-processing and shape extraction techniques are applied over the test dataset for each object. Then features of the proposed test shape are extracted. Finally by applying GMM based matching operations, the feature vector of the test shape is compared to previously created models. The model with maximum similarity is chosen as the class of the proposed shape. SVM classification steps are the same as the GMM classification ones except the label vector. In the training phase of SVM classification, each class model is labelled by a variable. Therefore in the test phase, after SVM based matching operations between feature vector and models, a decision is given by the model label of the test shape.

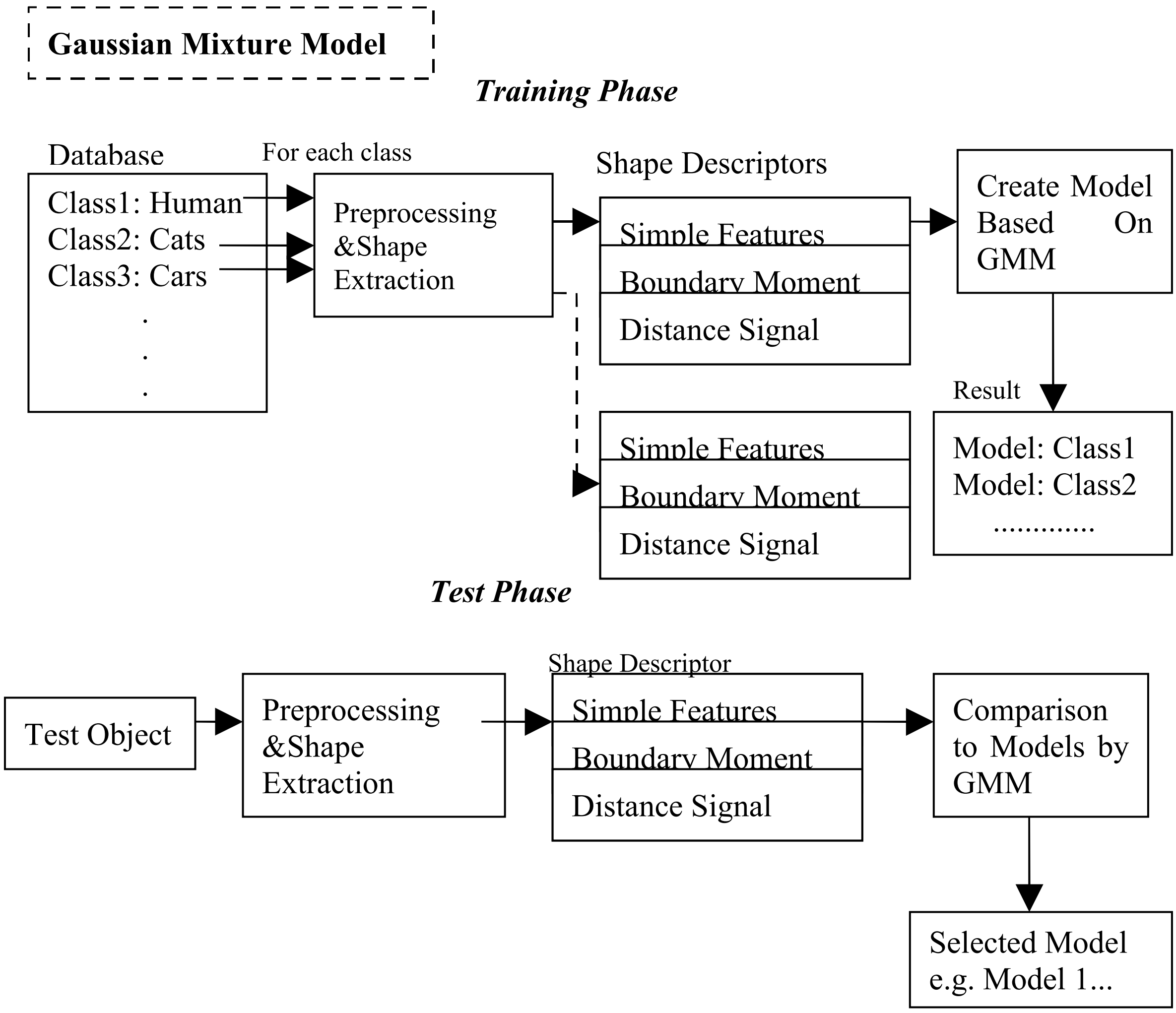


Figure-42: GMM based classification procedure scheme for training and test phases

The final classification technique used in this scenario is Shape Context. Basically, shape context is a feature, but it has its own classification technique, so it is explained as a different classification method. As shape context algorithm can be observed in Figure 43, different than all other classifiers, the procedure begins with the test object. The test image is pre-processed by morphological operations and shape is extracted by the contour following algorithm. Then the test object region and the database class object region are given to shape context procedure as input, test image is compared to all the objects in each class in the database one by one and the result for each image comparison is an error vector that measures the dissimilarity between two images. These vectors are hold, and then the class with biggest minimum-error vector-mean value result is chosen as the class of this test object.

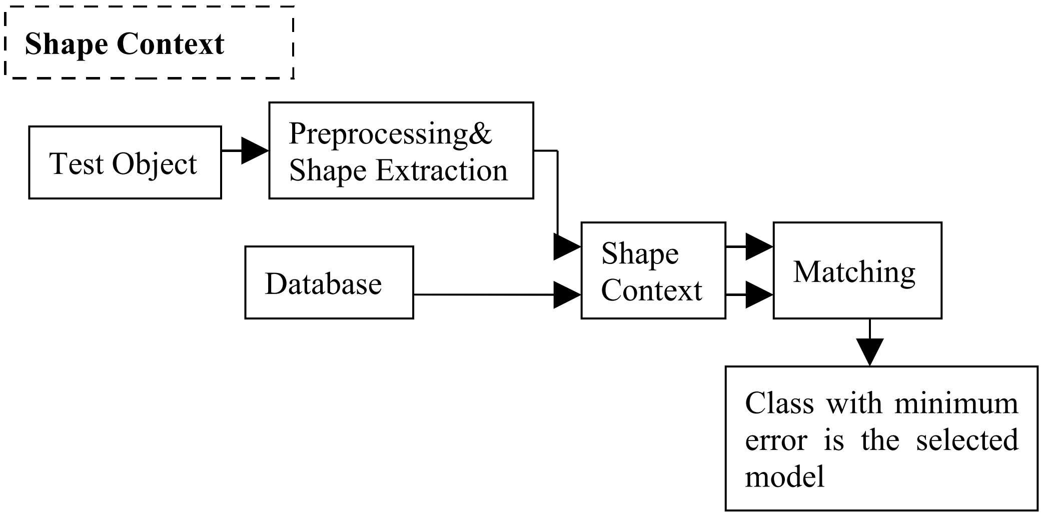


Figure-43: Shape Context based classification procedure scheme for the first scenario

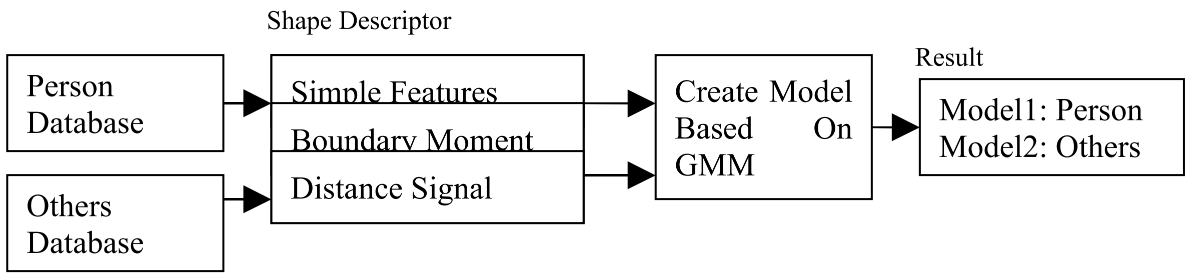
6.1.2 People detection in video-surveillance

The second scenario is based on human detection in video-surveillance that uses a two-class database where one class contains human shapes and the other one contains non-human shapes. Different than the first scenario, two models are created from the two available classes. The difference from the first scenario is that the test images are video-surveillance frames. As it is out of the scope of the project, foreground analysis is simulated by manually annotating the relevant foreground data in video-surveillance sequences.

As in the first scenario, the primarily applied classification is based on GMM. GMM based operations can be observed in Figure 44. In the training phase, the person database and other objects database image features are extracted one by one and models are created based on GMM operations. Mean, variance vectors and a weight value for each database are kept as model data. In the test phase, operations are a little bit different than in the first scenario. In order to make the calculations easier, additionally to the frame foreground masks, a text file with the information about the coordinates of the detected objects in each frame are given as second input data for this phase. Since the foreground masks are manually annotated, there is no need for a pre-processing stage since the data is provided without any noise effects. By using the information given in the text file, blobs are extracted from the frames, and then the shape of the frame object is extracted. In the following, the feature vector of the shape is extracted in order to use it in matching operations. GMM based operations are done on the feature vector and previously reached model data to compare them and two values are found for the two models. The model with the maximum similarity value is chosen as the class of the frame object. As in the first scenario, the SVM classifier follows a similar algorithm as the GMM classifier. In this scenario, two models are created for the training dataset and labelling vector (labelled human class by +1 and other object shapes class by -1), and the SVM classification becomes a binary classification problem. In the test phase, the extracted features are matched to models by SVM operations and if the test shape feature is found similar to the human model, result is given as +1, while otherwise the test shape does not belong to a human and result is given as -1.

Gaussian Mixture Model

Training Phase



Test Phase

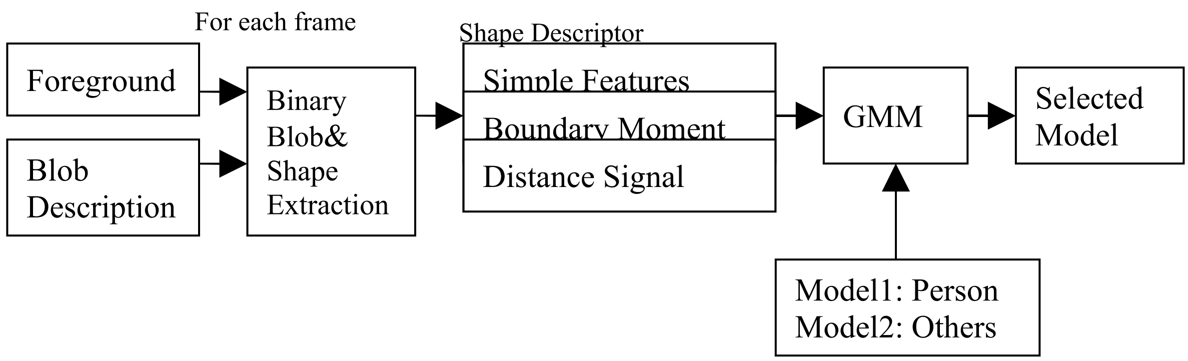


Figure-44: GMM based classification process for the second scenario

As in the first scenario, shape context for this scenario is explained separately since this feature has different classification procedure. The basic operations for this method can be observed in Figure 45. By reading the text file, blobs are extracted from frame masks and each blob is compared to each object in the databases, then the database with maximum number of minimum error valued object shapes is chosen as the class of the frame object.

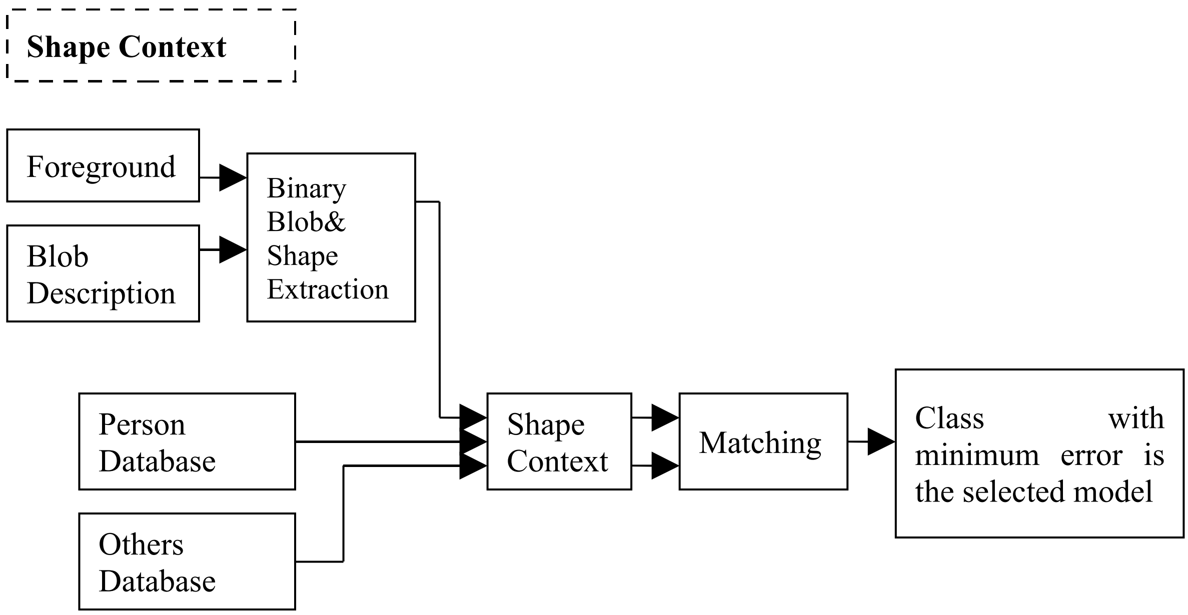


Figure-45: Shape Context based classification procedure scheme for the second scenario

6.1.3 Abandoned/Stolen Object detection in video-surveillance

This scenario is different than the first two scenarios with respect to both the training and test phases. In this scenario there are no databases, and the objective is to detect in the current frames if an object from previous frames is stolen or abandoned. Therefore test inputs are video-surveillance frames and a background image in order to test for the existence of the object, and training data are foreground masks. The general procedure in classification is to compare the foreground mask object with background and current frame; if the mask object does not exist in the background but exists in the current frame, then the decision is that this object is abandoned. However if the mask object exists in background but does not exist in frames, then the decision is that the object is stolen. In order to determine the existence of this mask object, the selected features and classifiers are used as in the other scenarios.

The first classifier applied is GMM based classifier as in all other scenarios. The algorithm applied for this classifier can be observed in Figure 46. In the training phase, a foreground mask object feature vector is extracted, and then an object model is created by GMM. In the test phase, a Background Subtraction algorithm is executed over the background and the current frame. Afterwards, pre-processing techniques are applied in order to get rid of noise effects. Then object blobs are extracted from frame and background if there is a detected object in the background or in the current frame. Subsequently, the shape-based feature vector is extracted from the blobs and a GMM based matching operation is done between mask model and blob features. If the mask object is found similar to the object determined in the current frame but is not similar to the background object, then the decision is given as the detected object is abandoned in the scene. If the mask object is not similar to the frame object but similar to the background object, then it is decided that the object is stolen from the scene. The same procedure is repeated for each selected feature as in all other scenarios.

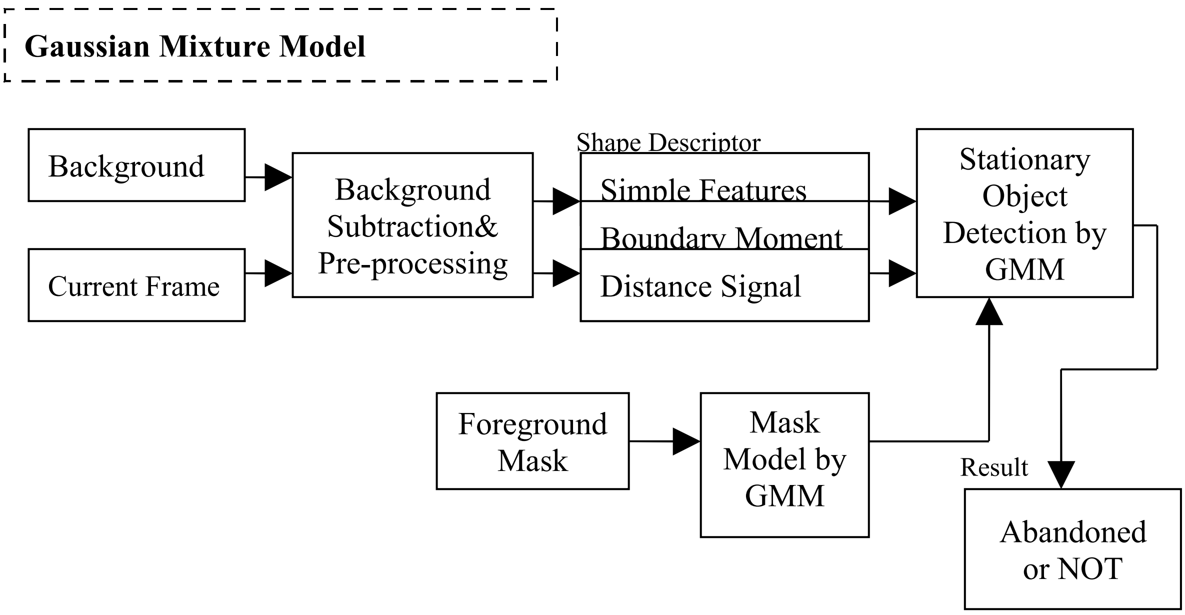


Figure-46: GMM based classification for the third scenario

SVM based classification is very similar to GMM based classification; after basic operations on background and current frame and existing objects are detected, a model is created by processing the feature vector of the mask object and the model is compared to current frame objects and background objects by SVM based matching operations. If the mask object is more similar to the current frame object than to the background object, then the detected object is detected as abandoned. If the mask object is more similar to the background object than to the current frame object, then the object is detected as stolen. This procedure is repeated for all three feature techniques (Simple Features, Boundary Moments and Distance Signal).

The Shape Context based feature and classification algorithm can be observed in Figure 47. The foreground mask object and the detected frame object regions are matched by the shape context feature and classified by the classifier techniques proposed in shape context definition and finally the error vector is reached. Then the same foreground object is matched to the background object by shape context and the second error vector is reached. Finally these two error vectors are compared and the minimum error vector is proposed; if the minimum error is found from the matching of mask and foreground, then the decision is made as the detected object is abandoned. If the minimum error is found from the matching of background object and mask, then it is proposed that the detected object is stolen for the scene.

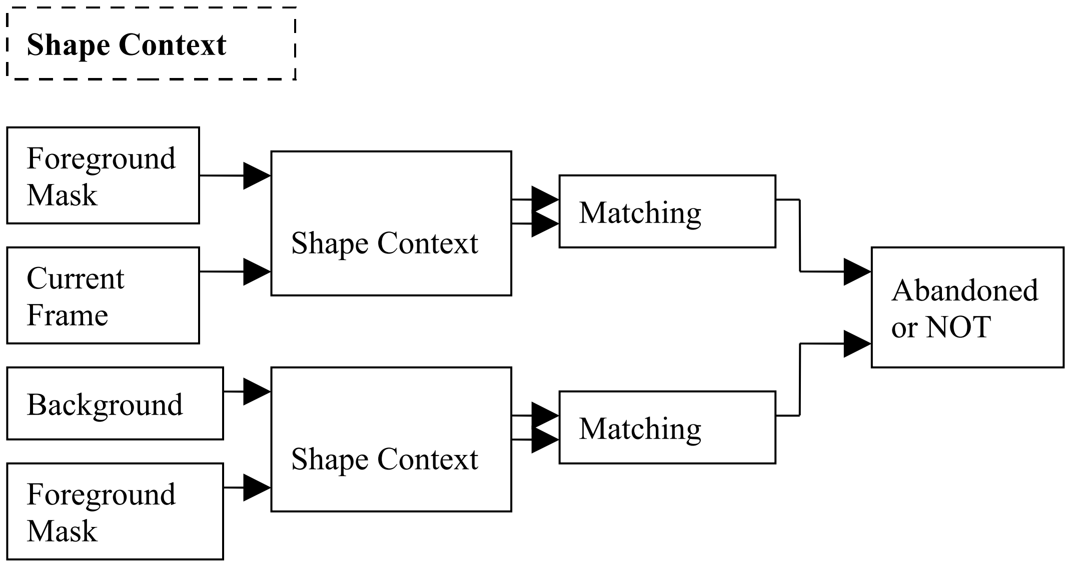


Figure-47: Shape context based classification procedure scheme

6.2 Dataset description

What makes each scenario special and distinct is the dataset used within these scenarios. The first scenario, database classification, is based on finding the best feature-classifier combination for the most accurate classification. Two different database sets are used where one database contains rotated object shapes and the other database contains same objects with small differences.

The second scenario, people detection in video-surveillance, is based on reaching an optimum result with respect to accuracy and time cost. Therefore in order to detect people in video-surveillance video, a database with two object classes is selected for the training phase and video-surveillance video frames are selected for the test phase.

Finally the third scenario, abandoned and stolen object detection in video-surveillance, is based on finding the best feature-classifier combination where the best accuracy is reached with least time cost. Therefore in order to detect the objects in a scene if the object is stolen or abandoned, two datasets are selected where one dataset contains abandoned objects and the other dataset contains stolen objects. Each dataset contains a background image of the scene, a frame sequence from scene and a foreground mask sequence of the detected object.

In order to understand the classification results reached for each scenario, it is important to understand the contents of the datasets used within proposed scenarios.

6.2.1 Database classification

In this scenario, the main goal is to find the best feature-classifier combination where the best accuracy is reached. Therefore a database is selected for this scenario with 17 object classes where the image samples from this database can be observed in Figure 48. Each object class under this database contains 128 rotated images of the same object. Database images are binary in order to operate on them easily. Since the database is based on rotated views of the same object, the objective under using this database is to observe if the feature-classifier combination is rotation invariant.

As this scenario consists of training and test phases, it is necessary to have datasets for each phase. The test dataset consists of randomly selected images from the database by 10% from each class and the training dataset consists of the rest of the images which are 90% for each class. The selected test image shapes are matched to the models created from the training dataset by following selected features and classifiers and the best accuracy is searched.

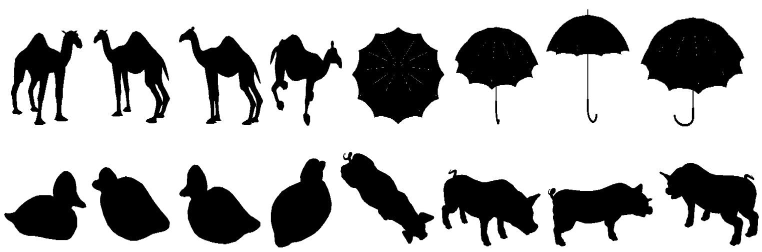


Figure-48: Sample images from first database

Since the detection of rotation invariance is not the main objective of this scenario, a second database is applied for database classification. The second selected database contains 23 object classes and each class contains 20 images while some classes contain more. Different from the first database, each image has small differences however still represents a certain shape of the object. Shape samples of this database can be observed in Figure 49. This database is used for better accuracy since each database image shapes represent the same object.

Similar to the first database, 10% of the each class images are randomly selected for training dataset and the rest of the images formed test dataset. The selected test image shapes are matched to the models created from the training dataset by following selected features and classifiers and the best accuracy is searched.

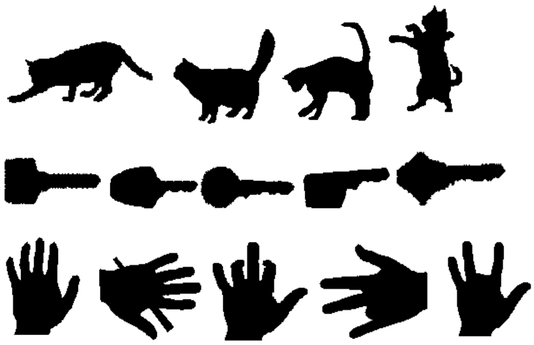


Figure-49: Sample images from the second database

6.2.2 People detection in video-surveillance

This scenario is based on finding the best feature-classifier combination for the optimum case of good accuracy and small time cost. Therefore a database with little number of classes and a dataset where the matching operation results in a good accuracy needs to be used. Thus, for people detection a database with two object classes is selected for the training phase. Since the scenario is people detection in video-surveillance, then the test data must be surveillance video sequences.

The selected database images can be seen in Figure 50. Human database images can be observed in the first row and the other objects in the database can be seen in the second row. Since the purpose is to detect in each frame if the current object is human or an object such as an animal, a car, or some other type of object. So, the objects database contains all kind of object shapes in order to have a successful matching. The human database contains 170 different human shapes while object database contains 117 shapes. Since the number of database images is smaller than in the first scenario, time cost is important.

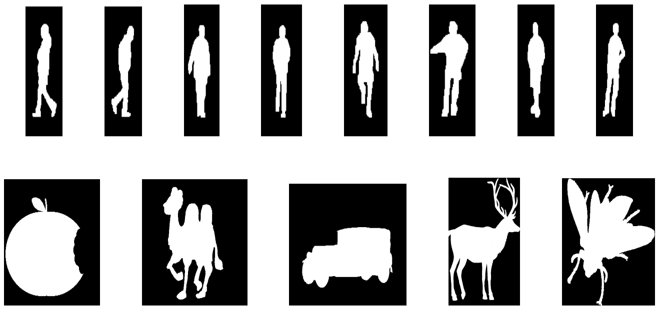


Figure-50: Human and Object Class database image samples

In test phase, foreground masks of a video-surveillance are selected as dataset to compare with the two class models created from human and object databases. The selected sequence contains 299 manually annotated frame masks. Samples from the sequence used in second scenario can be seen in Figure 51.

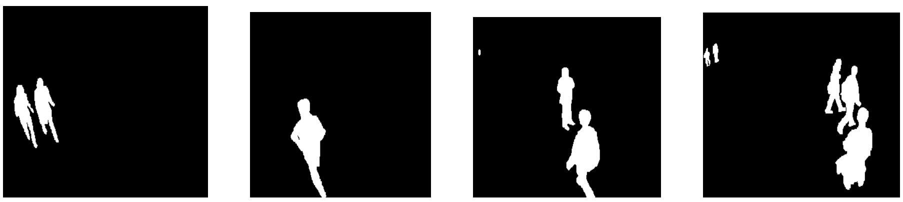


Figure-51: Frame foreground mask samples of the annotated video-surveillance frames.

6.2.3 Abandoned/Stolen Object detection in video-surveillance

This final scenario is based on finding the best feature-classifier combination for minimum time cost with high accuracy. Since the case in this scenario is abandoned or stolen object detection in video-surveillance, a video-surveillance video is the input data. Primarily, a sequence where an object is abandoned is selected. For the matching procedure of this scenario, a background image, corresponding frame sequence and object mask of the scene are needed for each object to test. The abandoned object sequence contains 23 frames and 23 masks with one background. Frame and mask samples of the selected video-surveillance video can be seen in Figure 52 with the background image. Unlike the previous scenario, the frames and background images are coloured and further processing techniques are needed for feature extraction and classification. In this scenario, the training dataset is the corresponding masks and the test datasets are background image and current frames.

Figure-52: Background image, two frames and two foreground object masks of the abandoned object dataset

Since it is not reasonable to make a decision by just taking the result of abandoned object frames, a second video-surveillance video is selected where an object is stolen from the scene. As in the abandoned object dataset, this sequence contains a background image, current frames and corresponding object masks. Image samples from this dataset can be observed in Figure 53. This sequence contains 11 annotated frames, one background coloured and 11 binary masks.

Figure-53: Background, frame samples and foreground object masks of stolen object dataset

6.3 Performance evaluation

In this section, performance evaluation of the selected features and classifiers is done by following the previously described approaches for each scenario. Basically this section is where the experimental results are described. As the objective in this project is to measure the efficiencies of selected features and results for different classifiers; these results are observed from these experimental results. Therefore, in one way to look at the case, this section is the cornerstone of the project.

In the experiment stage, for simplicity of computation and to have more accurate classification results, the number of Gaussians used for GMM classifier for each scenario is selected as 1, and then same procedure is repeated for 3 and 5 Gaussians. As in the GMM model, also in the SVM classifier needs some parameters. For all scenarios and datasets, default parameters for probability estimation of SVM classification are used.

In order to test the shape-based features and classification techniques selected in the proposed scenarios, the implementations of the GMM model and Active Contour adjustment from the Video Processing and Understanding Lab have been used. The LIBSVM library [31] has been used as the SVM implementation. The implementations of the extraction method and the matching procedure for the Shape Context feature have been taken from the author's website [16]. Additionally, some modifications of the original implementations have been done in order to use them in the project.

6.3.1 Database classification

In this case, experimental results of the database classification are examined. The first used database contains more than hundred images under each object class and each class contains rotated samples of class objects, so it can be proposed that this database basically measures the efficiency of features for rotation invariance. According to previously made descriptions of the selected features, some simple features are not rotation invariant since the perimeter and the area of the object are used; under 2D rotations, areas and perimeters of objects vary. For this reason, the rotation invariance is also detected with the use of this database. Shape context is expected to be rotation invariant since in its own algorithm it seeks for modifying the test shape to make it similar to the training object shape where the feature is the amount of modification made. In this stage, these expectations are compared to the experimental results for each classifier.

First experimental results can be observed from Table 7 for the GMM based database classification. These results are taken by 72 test images with 18-object class databases where under each class there are 128 images. When these results are examined, it can be observed that the feature with maximum cost is distance signal where the most effective feature is distance signal too. The feature with minimum cost and minimum correct matches is the Simple Features method.

Gauss 1	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*4.918%*	*892.9*
Boundary Moments	*24.59%*	*969.5*
Distance Signal	*31.15%*	*1403.5*
Gauss 3	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*4.918%*	*993.9*
Boundary Moments	*27.87%*	*1064.5*
Distance Signal	*40.98%*	*1500.6*
Gauss 5	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*4.918%*	*1003.2*
Boundary Moments	*27.87%*	*1154.6*
Distance Signal	*47.54%*	*1605.3*

Table-7: Experimental results for GMM based classification for database classification

The second operation is done for SVM based classification where the experimental results can be observed from Table 8. For this classifier, the same training database and test object images are used, so these results basically shows the difference between GMM and SVM based classifiers. While the costs for the simple features method increase, costs for the boundary moments and distance signal techniques do not change much. However the differences can be observed from the number of correct matches. For SVM based classifier, the best matching result is taken by using the Boundary Moments method, while the Distance Signal feature is the least successful.

	% of Correct Matches	**Time Cost (seconds)**
Simple Features	14.75%	1024.26
Boundary Moments	68.85%	899.6
Distance Signal	9.836%	1307.5

Table-8: Experimental results for SVM based database classification

The experimental results from the application of shape context technique can be observed from Table 9. As it has been studied before, this technique is a feature method with its own matching technique (in this feature images are compared one by one), so the computing time cost is expected to be very large. According to observed results, computing time cost is 18 hours which is not acceptable in this case.

	% of Correct Matches	**Time Cost**
Shape Context	*10.56%*	*18 hours*

Table-9: Experimental results for Shape Context method for database classification

Since the use of one database is not enough to reach conclusions, a database with different object classes that contains different objects of the corresponding class is applied. This database contains 23 object classes where each class contains generally 20 images while some classes contain more images. The same procedure is applied for this database and the following experimental results are reached.

In the GMM based classification for the second database; the reached experimental results can be observed from Table 10. Basic differences with respect to the first database for the GMM classification can be observed: cost in this database is larger because of the increase in used data. In this result, Distance Signal feature has the largest number of correct matches while other feature methods fail.

Gauss 1	% of Correct Matches	Time Cost (seconds)
Simple Features	*70.77%*	*1710*
Boundary Moments	*23.08%*	*1524*
Distance Signal	*35.38%*	*1532.9*
Gauss 3	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*47.62%*	*1801*
Boundary Moments	*38.10%*	*1630*
Distance Signal	*66.67%*	*1674.3*
Gauss 5	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*55.56%*	*1835*
Boundary Moments	*44.44%*	*1668*
Distance Signal	*77.78%*	*1702*

Table-10: Experimental results of GMM based classification for second database

The second matching operation done on the proposed database is SVM based classification; experimental results for this case can be seen from Table 11. According to these results, the best matching technique so far for this database can be defined as SVM based classification. Boundary moments and simple features have the perfect matching results while distance signal clearly fails. The methods with high correct matching results have larger costs.

	% of Correct Matches	Time Cost (seconds)
Simple Features	*98.46%*	*1698.7*
Boundary Moments	100%	1645.4
Distance Signal	10.77%	1347.7

Table-11: Results for SVM based classification for second database

The final matching operation applied for the database is Shape Context based operations, the result of this operation can be observed from Table 12. Matching percentage is pretty high for this case while time cost is really huge.

	% of Correct Matches	Time Cost (seconds)
Shape Context	*70.13%*	*21 hours*

Table-12: Experimental result for Shape Context method for second database

6.3.2 People detection in video-surveillance

Different from the first scenario, this scenario is based on classifying the video-surveillance frames to detect if the moving objects are human or an object. With this objective, a two-class database is used for the training phase where the Human class contains human shapes with different position and the Other Objects class contains different objects such as apples, cars, animals…etc. The human class consists of 170 shapes and the Other Objects class contains 117 images. Video-surveillance frames used in the test phase consist of 299 frames with a text file that carries the detected object coordinates in frames.

After application of GMM based classification on the proposed database classes and frames, the reached experimental results can be observed from Table 13. According to the table, clearly the simple features technique fails the classification while boundary moments successfully match the frames with database classes correctly with a small cost. Also distance signal method has a higher cost than other features, so it cannot be accepted for an optimum result.

Gauss 1	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*0.00%*	*60*
Boundary Moments	*80.01%*	*60*
Distance Signal	*33.78%*	*135.3*
Gauss 3	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*0.00%*	*73*
Boundary Moments	*20.40%*	*75*
Distance Signal	*51.08%*	*142*

Gauss 5	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*0.00%*	*73*
Boundary Moments	*20.40%*	*78*
Distance Signal	*54.19%*	*155*

Table-13: GMM based classification results for Human detection

SVM based classification results for the human detection procedure can be seen from Table 14. According to these experimental results, boundary moments method has the biggest percentage of correct matches. Time costs for this classifier are larger than the results of the GMM based operation, while boundary moments method has the biggest percentage of match, it has the biggest cost.

	% of Correct Matches	**Time Cost (seconds)**
Simple Features	36.21%	605.2
Boundary Moments	70.42%	*899.6*
Distance Signal	55.77%	*516.13*

Table-14: SVM based classification results for Human detection

The same procedures applied in the first scenario are applied for the shape context based feature and matching; experimental result can be observed from Table 15.

While the percentage of correct match is acceptable, still the cost of the technique is very high, so this method cannot be used for this scenario.

	% of Correct Matches	Time Cost
Shape Context	*81.42%*	*3 hours*

Table-15: Shape Context based method results for Human detection

6.3.3 Abandoned/Stolen Object detection in video-surveillance

This scenario is different from the other two scenarios as it is not using any database. In this scenario, object masks, video-surveillance frames and a background image are used. In the determination of abandoned or stolen objects; background and current frames are detected if an object disappears from frames or an object is abandoned in the scene in current frames. A first test is done by using a frame sequence where an object is abandoned in the scene. For this case, there are 26 frames and 26 object masks.

For the GMM based classification, experimental results reached can be observed from Table 16. According to this test, the most successful feature is the simple features method while its cost is the maximum at the same time. After this method, the boundary moments method is the second more successful feature used and this method has the minimum cost value. For this case, the distance signal method is the worst successful feature. Percentages of matches are higher for this case unlike other scenarios and costs are very low; the reason of this result is the used number of frames. Since there is lower number of frames as used in other scenarios, the percentages and costs changed.

Gauss 1	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*100%*	*41.28*
Boundary Moments	*92.31%*	*37.41*
Distance Signal	*65.38%*	*41.04*
Gauss 3	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*100%*	*43.02*
Boundary Moments	*95.31%*	*40.10*
Distance Signal	*75.38%*	*42.54*
Gauss 5	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*100%*	*43.82*
Boundary Moments	*95.31%*	*41.20*
Distance Signal	*75.38%*	*43.65*

Table-16: GMM based classification results for Abandoned/Stolen Object detection

In the SVM based classification for the detection of abandoned/stolen object, the experimental results can be observed from Table 17. These results are basically the reverse of the results observed in the GMM based classification while costs are very similar. In these results, the most successful feature is the distance signal, while the simple features matching percentage is the lowest.

	% of Correct Matches	Time Cost (seconds)
Simple Features	*61.5%*	*38.61*
Boundary Moments	*84.61%*	*36.25*
Distance Signal	*88.46%*	*42.67*

Table-17: SVM based classification results for Abandoned/Stolen Object detection

The results of the shape context method can be observed from Table 18. Since the less number of frames are used for this scenario, shape context method did not cost as much as it did in previous scenarios. As the shapes are matched one by one in this method, the number of correct matched are higher than previously applied feature results, but still the cost of this technique is higher than other techniques.

	% of Correct Matches	Time Cost (seconds)
Shape Context	*92.31%*	*81.11*

Table-18: Shape Context method results for Abandoned/Stolen Object detection

Classification results for each feature are proposed for the abandoned object based video sequence. Since this scenario includes the stolen object case, a second surveillance-video sequence where an object is stolen from the scene is used. This sequence consists of 11 frames and a background image with 11 object masks. Since the lowest number of frames is used in this case, the costs of the operations are very low.

As in other cases, the first classifier applied is the GMM based classifier where the experimental results can be seen from Table 19. Since the number of frames is very small, simple features and boundary moments methods give the perfect matching results where in the distance signal method; one detected frame result is wrong. Also the costs are very similar and small.

Gauss 1	% of Correct Matches	Time Cost (seconds)
Simple Features	*100%*	*4.36*
Boundary Moments	*100%*	*4.26*
Distance Signal	*90.90%*	*4.97*
Gauss 3	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*100%*	*4.50*
Boundary Moments	*100%*	*4.43*
Distance Signal	*92.95%*	*5.02*
Gauss 5	**% of Correct Matches**	**Time Cost (seconds)**
Simple Features	*100%*	*4.50*
Boundary Moments	*100%*	*4.43*

Distance Signal	*92.95%*	*5.02*

Table-19: GMM based classification results for a stolen object sequence

In SVM based classification; the simple features method gives a wrong decision for a frame where boundary moments and distance signal methods have perfect matching results and the costs are very low for this case as in GMM based classification results.

	% of Correct Matches	**Time Cost (seconds)**
Simple Features	*90.90%*	*4.34*
Boundary Moments	*100%*	*4.14*
Distance Signal	*100%*	*4.52*

Table-20: SVM based classification results for a stolen object sequence

Final results are taken for the shape context method as in Table 21. Cost is higher than for the previous classifiers as it is expected since each image is compared one by one. Percentage of correct matches can be seen lower than other methods, however it must be considered that the number of frames used is very small.

	% of Correct Matches	**Time Cost (seconds)**
Shape Context	*81.81%*	*12.51*

Table-21: Shape Context method results for a stolen object sequence

7 Conclusions and future work

7.11 Summary of work

In this work, shape-based feature and some existing classification techniques have been studied for the classification of 2D objects. The most popular shape-based features and classification techniques in the surveyed studies have been selected. One representative feature and classification technique for each category listed in the study of the state of art have been selected. Additionally, their implementation difficulty has been taken into account in the selection procedure.

After the selection of features and classifiers, different combinations of feature-classifier are proposed in order to achieve the main objective of the project. The idea under these different feature-classifier combinations is to find the best feature-classifier combination (or optimal) in three cases. The first case is based on finding the best combination where the accuracy is the main objective (without any time constraints). In order to find this combination, it is necessary to define a scenario that has a great amount of shape data for the training and test phases. The high availability of data ensures the confidence of the results obtained. The selected scenario for this case is chosen as the typical database classification problem where there are a lot object classes and each class is characterized by several shapes with different scale, rotation and size. This data is divided into the training and test sets. Due to the heterogeneity of the database shapes, this scenario is the best way to find the optimal feature-classifier combination.

The idea under the second case is to select the best combination taking into account the accuracy and computational time. The difference from the first case is that the number of database classes has to be reduced for this case and the accuracy has to be detected for video-surveillance data. The best scenario to fit in this case is chosen as People detection in video-surveillance video where there are only two database classes in the training phase (people and non-people classes). In the test phase, the video-surveillance sequences are manually annotated obtaining a foreground mask frame sequence for the test phase. The blobs of these foreground masks are analyzed by using the available feature-classifier combinations in order to compute their likelihood of being people or non-people.

Different from the first two cases, the third case is focused on finding the best feature-classifier combination taking into account the best accuracy and the minimum computational time. In this case, the computational time has more priority than the accuracy. Another difference with the other cases is the unavailability of an external database with object classes to recognize or detect. The best scenario to fit in this case is decided as Abandoned or Stolen Object Detection in video-surveillance where the necessary data is composed by the background of sequence, the video sequence frames and the foreground masks of the candidate objects to be detected as abandoned or stolen. The idea of this scenario is to extract shape-features in the foreground mask, current image and background image for each frame. Then, apply similarity measures between the shape features of foreground mask and the current/background images respectively. Finally, the abandoned or stolen object decision is taken by checking these similarities. If it is more similar to the current image, then the object is abandoned and if it is more similar to the background, the object is stolen.

For each scenario, three classifiers are used which are GMM, SVM and Shape Context based classifiers. With GMM and SVM classifiers, the same operations are repeated for the three feature descriptors which are Simple Features, Boundary Moments and Distance Signal. Shape Context is one of the feature descriptors; however it is not compatible with classifiers, so it has its own matching techniques that are generally based on distance measure between two shapes. For each scenario, the accuracy and time cost reached by each feature-classifier combination are detected and some conclusions are reached for each scenario as in the following section.

7.2 Conclusions

In the previous chapter, experimental results are represented in tables for each scenario and each classifier. Each classifier and feature combination gives different accuracy results for each scenario, because in each one the used data is different. By considering the use of different input data; the best feature and classifier technique combination is found by analyzing their accuracy and computational cost results. In this section, these results are evaluated according to each scenario and proposed classifiers.

7.2.1 Database classification

In this scenario, the experiments to recognize the objects classes have been performed using two different databases. The first database has been used to test rotation invariance of the selected features. According to the results, for GMM based classification basically all feature techniques obtain low accuracy results; maximum correct match is found better by Distance Signal than Boundary Moments feature. Then SVM based classification is applied on the same features; for this case Distance Signal feature failed and Boundary Moments method passed the classification by 68.8% where Distance Signal have the greatest time cost for this case. Finally the test is done for the same database by Shape Context, since one-by-one matching is applied in this technique, this procedure costs 18 hours with 10.56% correct matches. The basic results are found but it is not proper to give a decision by just one database classification result since the used database is based on rotation invariance and rotated object shapes are not exactly the best dataset for similarity measure. Therefore a second database is applied for the same feature-classifier combinations.

For the second database, in the GMM based classification the best accuracy is reached as 70.7% by Simple Features method where Distance Signal gives 35.38% and Boundary Moment gives 23.08% success. When the same process is applied on SVM based classification, two successful results are reached by Simple Features and Boundary Moments where Distance Signal fails. Finally Shape Context is applied and accuracy is reached as 70.13%.

As the conclusion of this scenario, without any time constraints, the best accuracy can be reached by Boundary Moments feature technique combined with SVM classifier and Shape Context methods.

7.2.2 People detection in video-surveillance

This scenario is based on people shape detection and classification in a video-surveillance frame sequence. Therefore a two-class database is used where one class has people shapes and other one consists of different non-people shapes. Since the idea under this scenario is to find the best feature-classifier combination where the combination gives the best accuracy and the optimum time cost, the experimental results are detected for optimum results.

For the GMM based classifier, the best matching result with 80% of correct match is found for the Boundary Moments while it also has the minimum cost. The same procedure is applied for the SVM based classifier and the result shows that for this classifier, also Boundary Moments is the most successful feature technique to use with this scenario while it has 70.42% of success. The procedure is also done for Shape Context; however it costs 3 hours, since the time cost is important for this scenario, this feature is not considered.

Since the purpose is to find optimum result with best accuracy and least time cost, according to experimental results, Boundary Moments with GMM classifier gives the optimum result for this scenario.

7.2.3 Abandoned/Stolen Object detection in video-surveillance

This final case is based on abandoned or stolen object detection in video-surveillance without using any kind of databases. This scenario is detected for two different datasets where one dataset has an abandoned object in the scene and the other dataset has a stolen object from the scene. The objective in this scenario is to find the optimum accuracy where computational time is minimal. First applied frame dataset is the abandoned object case. The object model is created from the foreground object mask in GMM classification and in the test phase, background and current frame objects are matched with the model, if the object is detected more similar to background than current frames, and then the object got stolen from the scene. However if the current frames are more similar than the background, then the object is abandoned. According to this definition, the classification results for GMM classifier are reached as; Simple Features gave the best accuracy with 100% match and Boundary Moments followed with 92.31%. Since the time cost is considered in this case, it should be noted that the minimum time cost is reached by Boundary Moments.

For the SVM based classifier, the best matching with least time cost is reached by Boundary Moments. Finally 92.31% of correct match is obtained by the Shape Context method, but this method has the largest time cost. Since the idea under this scenario is to select the best combination with least time cost, Shape Context is not considered.

For the second dataset, a frame sequence where an object got stolen from the scene is applied. However this frame sequence has 11 frames, so almost all methods give correct matches. The GMM classifier gave the best matching for Simple Features and Boundary Moments as in the abandoned frame sequence. Then SVM is applied, and the best matching percentages are reached by Boundary Moments and Distance Signal. For Shape Context, the result shows that this method is still applicable with a good matching percentage. Since the matching time cost is important, for GMM and SVM classifiers, minimal time cost and best matching percentage are reached by Boundary Moments where the percentage is higher for GMM classifier. So, the best combination for this scenario is proposed as Boundary Moments with GMM classifier.

7.3 Future work

In this work, a search for the optimal combination of feature-classifier is performed by analyzing the results for different scenarios. The possible lines of future work are:

- Extend the study to other types of features and classifiers (like Neural Networks).
- Principal Component Analysis (PCA) should be used with shape-based features in order to get rid of useless results. After such operation, the classification process would have less computational time which means less cost for the user.
- Study the inclusion of an Online Feature Selection module (OFS) (similarly to Hata: Başvuru kaynağı bulunamadı) and an Online Classification Technique Selection module (OCTS) in the proposed framework. This module will allow the system to select the optimum combination of feature-classifier depending on the objective of the classification task (e.g., scenario). Additionally, the usefulness of the features and techniques has to be characterized more precisely for each scenario.
- Additionally, specific classifiers can be constructed for each problem (e.g., scenario) by combining different classifier types (hybrid classifiers) in order to increase the robustness of the classification.

References

[1] B. Sami, L. Jorma, O. Erkki, "Statistical Shape Features in Content-Based Image Retrieval", Proc. of 15th International Conference on Pattern Recognition, volume II, 2000, pp. 1062-1066.
[2] S. Loncaric, "A survey on shape analysis techniques", *Pattern Recognition*, Vol. 31, pp. 983-1001, 1998.
[3] R. C. Veltkamp, M. Hagedoorn "State-of-the-art in shape matching", Technical report UU-CS-1999-27, Utrecht University (The Netherlands), 1999.
[4] F. A. Otoom, H. Gunes, M. Piccardi, "Automatic Classification of Abandoned Objects for Surveillance of Public Premises", Proc. of Congress on Image Processing, 2008, pp. 542-549.
[5] M.M. Mahmoud, E.Y. Rammadan, "2D Object Recognition Using Its Contour Feature", Informatics Research Institute, University of Bahrain.
[6] B. Bose, E. Grimson, "Improving Object Classification in Far-Field Video", Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004 (CVPR'04).
[7] Y. Dedeoglu, "Moving Object Detection, Tracking and Classification for Smart Video Surveillance", Msc. Thesis, Bilkent University, 2004.
[8] R. N. Hota, V. Venkoparao, A. Rajagopal, "Shape Based Object Classification for Automated Video Surveillance with Feature Selection", Proc. of 10th International Conference on Information Technology, (ICIT 2007), 2007, pp. 97-99.
[9] J.C. San Miguel, J.M. Martinez, "Robust unattended and stolen object detection by fusing simple algorithms", Proc. of AVSS 2008, pp. 18-25.
[10] D. Zhang, "Image Retrieval Based on Shape", Doctorate Thesis, Faculty of Information Technology, Monash University, 2002.
[11] Shape Signature Matching for Object Identification Invariant to Image Transformations and Occlusion
[12] R. Veltkamp, M. Hagedoorn, "Shape Similarities, Properties, and Constructions", *Proceedings of the 4th International Conference* on *Advances in Visual Information Systems, VISUAL 2000, Lyon, France, November 2000* (pp. 467-476).
[13] J. Lötjönen, T. Mäkelä, "Elastic Matching Using a Deformation Sphere", Proc. of 4th International Conference on Medical Image Computing and Computer Assisted Intervention, October 14-17, 2001, pp. 541-548.
[14] X. Dai, S. Khorram, "A Feature-Based Image Registration Algorithm Using Improved Chain-Code Representation Combined with Invariant Moments", IEEE Transactions on Geoscience and Remote Sensing, volume-32, September 1999.
[15] N. Rougon, F. Preteux, "Understanding the Structure of Diffusive Scale-Spaces", CONF 1996, pp. 884-889.
[16] S. Belongie, J. Malik, "Matching with Shape Context", IEEE Transactions on Pattern Analysis and Machine Intelligence, 27 (11):1832-1837, Nov. 2005. Code implementation available at http://www.eecs.berkeley.edu/Research/Projects/CS/vision/shape/sc_digits.html.
[17] M. K. Hu, "Visual Pattern Recognition by Moment Invariants", IRE Trans. Info. Theory, vol. IT-8, pp.179–187, 1962.
[18] C. Sheng, Y. Xin, "Shape-Based Image Retrieval Using Shape Matrix", International Journal of Signal Processing, 2005, pp. 163-166.
[19] C. Campell, "Introduction to Support Vector Machines", Bristol University.

[20] C. Campell, “Introduction to Gaussian Mixture Model”, Bristol University.
[21] I. Aleksander, H.Morton, An introduction to neural computing. 2nd edition, EDITORIAL.
[22] NN at Wikipedia- http://en.wikipedia.org/wiki/Neural_network
[23] Database, http://www.cs.toronto.edu/~dmac/ShapeMatcher/index.html
[24] Database, http://lmb.informatik.uni-freiburg.de/people/haasdonk/datasets/distances.en.html
[25] L.J. Latecki, “Shape Data for the MPEG-7 Core Experiment CE-Shape-1,” http://www.cis.temple.edu/~latecki/TestData/mpeg7shapeB.tar.gz, 2002.
[26] K. Ozonat, M. R. Gray, “Gauss Mixture Model-Based Classification for Sensor Networks”, Proc. of Data Compression Conference, 2006, pp. 322- 331.
[27] A. Tzotsos, “A Support Vector Machine Approach For Object Based Image Analysis”, Laboratory of Remote Sensing, School of Rural and Surveying Engineering.
[28] A. Cavallaro, O. Steiger, T. Ebrahimi, “Semantic Video Analysis for Adaptive Content Delivery and Automatic Description”, IEEE Transactions of Circuits and Systems for Video Technology, 15(10):1200-1209, Oct. 2005.
[29] MAP at Wikipedia- http://en.wikipedia.org/wiki/Maximum_a_posteriori
[30] M. A. Garcia, “Gamma-Based Video-Object Segmentation”, November 2007.
[31] Chang, C.-C., Lin, C.-J.: LIBSVM : a library for support vector machines (2001), software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm

Glossary

ANN	Artificial Neural Network
DM	Distance Measure
GMM	Gaussian Mixture Model
NN	Neural Network
OCTS	Online Classification Technique Selection
OFS	Online Feature Selection
PCA	Principle Component Analysis
SVM	Support Vector Machine

Annex

A. Background subtraction based on Gamma Function

Application of this method consists of two stages: properly updating the reference background and suitably subtracting background and current image. Object segmentation procedure is applied between current frame and reference background and frame is detected for any changes. If a person comes to the frame, its shape is detected by background subtraction and the data got ready for further processing such as shadow removal and blob extraction. If the current frame is represented as *I[x, y]* and background frame is represented as *B[x, y]*, background subtraction method can be defined by the operation:

$$b[x,y] = 1 \Rightarrow B[x,y] = \alpha\, B[x,y] + (1-\alpha)\, I[x,y]$$

where α is the constant of the operation.

Another type of background subtraction is Gamma based segmentation which is not very different from background subtraction. The operation applied in this method is:

$$foreground\,(I[x,y]) \Leftrightarrow (I[x,y] - B[x,y])^2 > \beta$$

In this method, subtraction is done for each pixel as in the previous method and a threshold β is applied. The problem with this method is that the result is usually very noisy. However this problem can be reduced by subtracting a square window around every pixel such as:

$$foreground\,(I[x,y]) \Leftrightarrow \sum_{i=-W}^{W} \sum_{j=-W}^{W} (I[x+i, y+j] - B[x+i, y+j])^2 > \beta$$

Further subtraction operations can be defined as:

$$X_{i,j}(x,y) = \frac{I[x+i, y+j] - B[x+i, y+j]}{\sigma} \approx N(0,1)$$

The result of this operation can be optimized as:

$$Q(x,y) = \sum_{i=-W}^{W} \sum_{j=-W}^{W} \left(\frac{I[x+i, y+j] - B[x+i, y+j]}{\sigma} \right)^2$$

The performance of the subtraction depends on the selected constants, they can be found by numerical root finding methods such as bisection method [30].

B. Active Contours adjustment (Snakes method)

This method is used to adjust the object shapes of the binary foreground masks of the frames. It is important to apply because discontinuities on the shape of the object can cause false results, which is not desired.

Active contour can be described as an energy minimizing function that moves some points on the image in order to minimize its energy. Active contour determination can be applied on each point as:

$$E = \alpha E_{continuity} + \beta E_{curvature} + \lambda E_{image}$$

where α , β and λ are parameters of the operation. By currently iterating, active contour searches for more points to minimize, when there is no more possible minimization, then the operation is complete [9].

In the proposed project, implementations that are previously done by VPU are used.

PRESUPUESTO

- **Ejecución Material**

- Compra de ordenador personal (Software incluido)... 2.000 €
- Alquiler de impresora láser durante 6 meses..50 €
- Total de ejecución material... 2.050 €

1) Gastos generales

- 16 % sobre Ejecución Material... 352 €

2) Beneficio Industrial

- 6 % sobre Ejecución Material.. 132 €

3) Material fungible

- Gastos de impresión.. 60 €
- Encuadernación... 200 €

4) **Subtotal del presupuesto**

- Subtotal Presupuesto... 12060 €

5) I.V.A. aplicable

- 16% Subtotal Presupuesto.. 1929.6 €

6) Total presupuesto

7) Total Presupuesto... 13989,6 €

Madrid, Julio de 2009

El Ingeniero Jefe de Proyecto

Fdo.: ULYA BAYRAM
Ingeniero Superior de Telecomunicación

PLIEGO DE CONDICIONES

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un sistema de Reconocimiento de objetos 2D basado en el análisis de contorno. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es

obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad “Presupuesto de Ejecución de Contrata” y anteriormente llamado ”Presupuesto de Ejecución Material” que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.