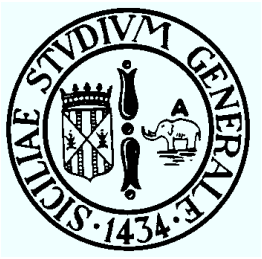



UNIVERSITÀ DEGLI STUDI DI CATANIA

Facoltà di Scienze Matematiche Fisiche e Naturali

Dottorato di ricerca in Informatica

---

# SHAPE MODELING AND DESCRIPTION FROM 2D IMAGES

GIUSEPPE CLAUDIO GUARNERA

A dissertation submitted to the Department of Mathematics and Computer Science and the committee on graduate studies of University of Catania, in partial fulfillment of the requirements for the degree of doctorate in computer science.

ADVISOR
Prof. Giovanni Gallo

---

XXV Ciclo

# Abstract

The visual ability of humans and other animals is the result of the complex interaction of light, eyes and brain. We are unaware of how extremely complex is the shape analysis performed by the brain, since it is mainly done subconsciously, with no need to involve the higher level of cognition. Hence, although "to see and understand" seems natural and straightforward, the design of versatile and robust computer vision system is a tough task.

In computer era, the attempt to imitate the human ability to understand shapes led to the fields of computer vision and pattern recognition, having been motivated by a wide spectrum of important applications such as in robotics, biology, security and so on. In computer vision, accordingly to the great variety of applications, there is a wide range of possible computer's "eyes", which can be very different from a human eye (e.g. tomography devices, ultrasonic range sensors).

Among all different aspect underlying visual information, the shape of objects plays a very important role, which will be emphasized in this dissertation showing how such a fundamental feature can be efficiently used to solve a wide range of relevant problems in Computer Vision, Pattern Recognition and Computer Graphics, in both the 2-D and 3-D realms. In almost all cases the input data is a set of one or more images, demonstrating that they convey enough information about the shape, if a proper setup is used. The input devices we use range from DSLR cameras to MRI devices, with great differences in the imaging techniques and in the quality of produced images. Starting from the input images, we show how to accurately model the surface of the depicted objects using polarization cues or how to parameterize shapes using state of the art descriptors, based on statistical properties of the object classes or simply on the individual objects surface or silhouette.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

### 1.0.1 Motivation

One of the most stimulating research fields, from both the scientific and technological perspectives, is related to vision, the main human sense and source of most significant information for most people about the world. The visual ability of humans and other animals is the result of the complex interaction of light, eyes and brain. We are able to see because light from an object can move through space and reach our eyes. Once light reaches our eyes, signals are sent to our brain, and our brain deciphers the information in order to detect the appearance, location and movement of the objects we are sighting at. The whole process, as complex as it is, would not be possible if it were not for the presence of light. When we see a picture, we understand its meaning by recognising the objects which are depicted and relating them to what we have learned in our lives: among all different aspect underlying visual information, the shape of objects plays a very important role, since shapes can be considered as being the words of the visual language.
We are unaware of how extremely complex is the shape analysis performed by the brain, since it is mainly done subconsciously, with no need to involve the higher level of cognition. Hence, although "to see and understand" seems natural and straightforward, the design of versatile and robust computer vision system is a tough task. In computer era, the attempt to imitate the human ability to understand shapes led to the fields of computer vision and pattern recognition, having been motivated by a wide spectrum of important applications such as in

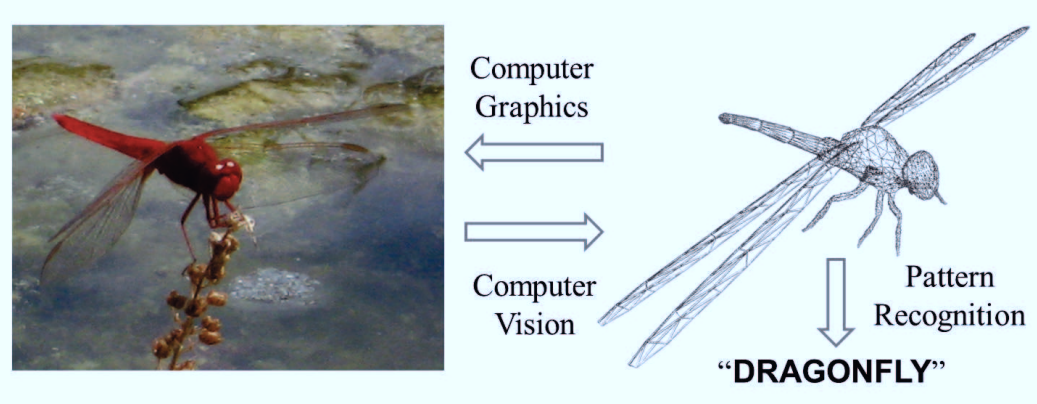



Figure 1.1: Conceptual representation of the relations between the fields of computer vision, computer graphics, and pattern recognition.

robotics, biology, security and so on. What we see is a 2-D picture perceived by the eyes, so it is common to think of a shape as a 2-D projection of a 3-D object (a geometric model), but in computer vision, accordingly to the great variety of applications, there is a wide range of possible computer's "eyes", which can be very different from a human eye (e.g. tomography devices, ultrasonic range sensors), hence shapes can be represented as parametric surfaces, triangular meshes, etc. other than the common 2-D shapes. As stated before, computer vision deals with extracting information about objets from their visual representation, whereas computer graphics addresses the converse problem: how to realistically and aesthetically render an image from a geometric model. There are many other fields which involve shapes, such as pattern recognition, i.e. to assign a label to the representation of a shape. Fig. 1.1 describes the relations between computer vision, computer graphics and pattern recognition.

In this dissertation the key role of the shape in all these fields will be emphasized, demonstrating how such a fundamental feature can be efficiently used to solve a wide range of relevant problems, in both the 2-D and 3-D realms. In almost all cases the input data will be a set of one or more 2-D images, showing that they convey enough information about the shape, if a proper setup is used.

### 1.0.2 Chapter Overview

The remainder of this thesis is organized as follows. Chapter 2 discusses the relevant background for the topics covered in this dissertation.

In Chapter 3 we propose a novel method for estimating surface orientation from the Stokes polarization vector under a *single* spherical incident illumination condition that is either circularly polarized or unpolarized. Accurate shape and appearance estimation is a crucial component in many computer vision and computer graphics applications. For instance, digitally reproducing the appearance of physical objects such as cultural artifacts, consumer products, material samples, and human faces is a long-standing goal of computer graphics. Despite being at the focal point of intense research, shape and appearance estimation remains a challenging problem, especially under uncontrolled real-world conditions. We illustrate the practicality of our technique by estimating surface normals under uncontrolled outdoor illumination from just four observations from a fixed viewpoint. Estimation of both small-scale surface shape as well as surface reflectance greatly benefits from accurate knowledge of surface orientation and we describe how the proposed method can be combined with a novel data-driven surface reflectometry technique which is able to measure, per visible surface point, diffuse and specular albedo, specular roughness of isotropic BRDFs, and index of refraction, assuming circularly polarized incident lighting.

In Chapter 4 we present a novel technique to describe the shapes of women breasts in a low dimensional parameter space. The parametrization is obtained from a sample of about 40 patients MRI taken in prone position. The data have been cleaned from noise and disturbances and has been dimensionally reduced using Principal Component Analysis. The resulting principal modes may be used as clinical indicator and have a direct medical interpretation. If two references relative to extremal shapes (one of a reconstructed breast and one of a severely aged breast) are taken, all the other shapes span a continuum space that provides an objective way to classify and describe the variability observed in the common clinical practice.

We then discuss the feasibility of estimating the Body Condition Score (BCS) of cows from digital images by employing statistical shape analysis and regres-

sion machines in Chapter 5. The shapes of body cows are described through a number of variations from a unique average shape. Specifically, Kernel Principal Component Analysis is used to determine the components describing the many ways in which the body shape of different cows tend to deform from the average shape. This description is used for automatic estimation of BCS through regression approach. The proposed method has been tested on a new benchmark dataset available through the Internet, demonstrating that both our semi-automatic and automatic approaches outperform the state-of-the-art approaches proposed in the context of dairy cattle research.

The use of contour-based shape descriptors to deal with articulated and deformed shapes is discussed in Chapter 6, describing how Computer Vision and Pattern Recognition can provide a great support in Archaeometry. An important feature of the Minoan culture is the pottery of Kamares style, that documents the Cretan cultural production between the first half of the II millennium BC. This high level painted production, characterized by the combination of several diverse motifs, presents an enormous decorative repertoire. The extraordinary variety of combinations between elementary motifs according to a complex visual syntax makes interesting the automatic identification of the motifs, particularly upon potsherds. A complete pipeline to accomplish this task is still a challenge to Computer Vision and Pattern Recognition, however the co-occurrence of the different shapes in a specimen can be used to help the archaeologists in the cultural and even chronological setting.

In Chapter 7 we describe hardware and software components to detect counterfeits of Euro banknotes, demonstrating the use of low cost hardware and consequently a very simple shape classification technique to effectively solve this problem. The system is also able to recognize the banknote values. The proposed method makes use of images acquired with near infrared camera and works without motors or any other moving parts. The proposed system is robust to changes in ambient lighting and banknote positioning. The effectiveness of the proposed solution has been properly tested on a dataset composed by genuine and fake euro banknotes.

Finally, we conclude with a discussion on the scope and contribution of the dissertation in Chapter 8.

# Chapter 2

# Background and related work

Our work relates to a significant body of work in photometric stereo, polarization-based analysis of reflected light,shape from shading and 2D shape description. We start by describing the principles of photometric stereo and their application to surface normals estimation, describing also the current state of the art. A recent advance in computer vision, which is able to recover low frequency details of shape, albedo and illumination from a single grayscale image will then be described. We then review some shape 2D descriptors, used to model statistical properties of the shape contour or to address the typical deformation observed in shape analysis.

## 2.1 Shape Appearance

Shape recovery is a classic problem in computer vision. The goal is to obtain a 3-D scene description from one or more 2-D images. The recovered shape can be expressed in several ways: depth $Z(x,y)$, surface normal $(n_x,n_y,n_z)$, surface gradient $(p,q)$, and surface slant $\phi$ and tilt $\theta$. The depth can be considered either as the relative distance from the camera to surface points, or the relative surface height above the $x-y$ plane. The surface normal is the orientation of a vector perpendicular to the tangent plane on the object surface. The surface gradient, $(p,q) = (\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y})$, is the rate of change of depth in the $x$ and $y$ directions. The surface slant $\phi$ and tilt $\theta$ are related to the surface normal as

$(n_x, n_y, n_z) = (l \sin\phi \cos\theta, l \sin\phi \sin\theta, l \cos\theta)$, where $l$ is the magnitude of the surface normal.

### 2.1.1 Intrinsic Images

The problem of describing a scene in terms of intrinsic characteristics, such as range, orientation, reflectance, and incident illumination of the surface element visible at each point in the image, was first posed to the computer vision community as the "intrinsic images" problem [1]. The central problem in recovering intrinsic scene characteristics is that the information is confounded in the original light-intensity image: a single intensity value encodes all the characteristics of the corresponding scene point. Recovery depends on exploiting constraints, derived from assumptions about the nature of the scene and the physics of the imaging process. Over time the intrinsic images problem has been simplied to the problem of separating an image into shading and albedo, which has seen some recent progress [2] mostly by relying on color as a cue. However, none of these algorithms allow shape to be directly recovered.
A competing approach to single-image techniques is to better constrain the problem with additional data. Instances of this approach are photometric stereo [3], structure from motion [4], and inverse global illumination [5]. All these techniques depend on multiple observations.

**Shape From Shading.** In computer vision, the techniques to recover shape are called shape-from-$X$ techniques, where $X$ can be shading, stereo, motion, texture, etc. The classic shape-from-shading problem [6] deals with the recovery of the shape of a smooth opaque object from a monocular image, given a knowledge of the surface photometry and the position of the light-source. This method in some sense can be considered as complementary to the use of stereoscopy, which relies on matching up sharp detail and will fail on smooth objects.
To solve the shape-from-shading problem, it is important to study how the images are formed. A simple model of image formation is the Lambertian model, in which the gray level at a pixel in the image depends on the light source direction and the surface normal. In shape-from-shading, given a gray level image, the aim is to recover the light source and the surface shape at each pixel in the

image. However, real images do not always follow the Lambertian model. Even if we assume Lambertian reflectance and known light source direction, and if the brightness can be described as a function of surface shape and light source direction, the problem is still not simple. This is because if the surface shape is described in terms of the surface normal, we have a linear equation with three unknowns, and if the surface shape is described in terms of the surface gradient, we have a non-linear equation with two unknowns. Therefore, finding a unique solution to shape-from-shading is difficult; it requires additional constraints. Most shape-from-shading algorithms assume that the light source direction is known. In the case of the unknown light source direction, some assumptions about the surface shape are required, such as the local spherical surface, uniform and isotropic distribution of the surface orientation. Once the light source direction is known, 3-D shape can be estimated.

**Shape and Albedo from a Single Image.**

None of the "intrinsic images" algorithms allow shape, or illumination to be directly recovered. Recently a superset of the shape-from-shading problem has been addressed by [7], in which shape, albedo, and a spherical harmonic model of illumination are estimated from a single grayscale image of an object, using shading as primary cue. This technique is the first unified model for jointly estimating all the mentioned parameters together. Since this "inverse optics" problem is terribly underconstrained, the authors construct statistical models of albedo and shape, and define an optimization problem that searches for the most likely explanation of a single image. The model has been called SAIFS (shape, albedo, and illumination from shading) and it produces reasonable results on arbitrary grayscale images; SAIFS has been further extended to recover chromatic illumination [8]. Cast shadows and specularities are not addressed by the SAIFS model. The key observation is that the space of albedos, shapes, and illumination that reproduce an image is immense but not all albedos and shapes are equally likely. The SAIFS algorithm is an optimization problem in which the most likely shape, albedo, and illumination are recovered under to the statistical model of natural albedo and shape, such that a single image is exactly reproduced. The SAIFS model relates to the classic shape-from-shading problem since it uses shading as

a cue for depth.

Assuming Lambertian reectance and orthographic projection, given a log-albedo map $A$, a depth-map $Z$, and illumination $L$, the log-intensity image $I$ is dened as $I = A + S(Z,L)$. The function $S(Z,L)$ is called the "logshading image" of $Z$ with respect to $L$: it linearizes $Z$ into a set of normals and renders those normals using $L$, a model of spherical harmonic illumination. Assuming that the image $I$ and illumination $L$ have been observed, but $Z$ and $A$ are unknown, the problem is still underconstrained, so priors on Z and A are imposed. The term "prior" in the context of SAIFS is actually an abuse of terminology, since it actually refers to loss functions or regularizers on $Z$ and $A$. Two priors on albedo are used, which encourage local smoothness and global sparsity, and three priors on shape which encourage flatness, outward-facing orientation at the occluding contour, and local smoothness. More in detail $g(A)$ are two priors on albedo: one which encourages piecewise-smoothness by placing heavy-tailed distributions on the multiscale gradient norm of log-albedo, and one which encourages a low entropy in the marginal distribution of log-albedo across all scales of an entire image. $f(Z)$ are the three priors on shape: one which encourages fronto-parallel flatness (primarily to address the bas-relief ambiguity [4]), one which informs the surface orientation near an objects occluding contour, and one which encourages a novel measure of smoothness by placing heavy-tailed distributions on the multiscale gradient norm of the mean curvature of shape. The algorithm search for the most likely shape and albedo that explain image $I$ and this is equivalent to minimizing the sum of $g(A)$ and $f(Z)$, which is defined as being loosely equivalent to the negative log-likelihoods of $Z$ and $A$ respectively.

The optimization problem is:

$$\begin{aligned} \underset{Z,A}{\text{minimize}} \quad & g(A) + f(Z) \\ \text{subject to} \quad & I = A + S(Z,L). \end{aligned} \tag{2.1.1}$$

The problem formulation in Equation 2.1.1 can be extended to present a novel framework for recovering illumination in addition to shape and albedo. The technique can be modified to handle unknown illumination, reformulating the problem

as a MAP estimation:

$$\underset{Z}{\text{maximize}} \quad \log(P(A|Z,L)P(Z)) \tag{2.1.2}$$

$$P(A|Z,L) = \exp(-g(I - S(Z,L))) \tag{2.1.3}$$

$$P(Z) = \exp(-f(Z)). \tag{2.1.4}$$

$L$ can be introduced as a latent variable, and marginalizing over it:

$$\underset{Z}{\text{maximize}} \quad \log(P(Z)\sum_{L} P(A|Z,L)P(L)) \tag{2.1.5}$$

### 2.1.2 Photometric Stereo and Normals Estimation

The seminal work by Woodham [3] on photometric stereo, which proposes to estimate surface normals from single viewpoint images of diffuse surfaces lit from different lighting directions using a simple linear system, has inspired much research since its introduction in 1978. The idea of photometric stereo is to vary the direction of the incident illumination between successive images, while holding the viewing direction constant. Since the imaging geometry is not changed, the correspondence between image points is known a priori. This technique is called "photometric" because it uses the radiance values recorded at a single image location, in successive views.

Let $(x,y,z)$ be the coordinates of a surface point. If the equation of an object surface is given explicitly as $z = f(x,y)$, then a surface normal is given by the vector: $\left[\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y}, -1\right]$. We can define $p$ and $q$ by: $p = \frac{\partial f(x,y)}{\partial x}$ and $q = \frac{\partial f(x,y)}{\partial y}$, then $(p,q)$ is the gradient of $f(x,y)$ and the surface normal can be written as $[p,q,-1]$. Let $R(p,q)$ be the reflectance map, which determines image intensity as a function of $p$ and $q$. A reflectance map captures the surface reflectance of an object material for a particular light source, objet surface and viewer geometry. Using the reflectance map, the basic equation describing the process of image formation can be written as:

$$I(x,y) = R(p,q). \tag{2.1.6}$$

If three images $I_1(x,y)$, $I_2(x,y)$ and $I_3(x,y)$ are obtained by varying the direction of incident illumination with no change in the imaging geometry, then each pixel $(x,y)$ in the three images corresponds to the same surface point and gradient $(p,q)$. The effect of varying the illumination is to change he reflectance map in each image. The three views are characterized by three independent equations:

$I_1(x,y) = R_1(p,q)$

$I_2(x,y) = R_2(p,q)$

$I_3(x,y) = R_3(p,q)$ .

These equations in general are non linear with more than one solution. If the directions of incident illumination are not collinear in azimuth, assuming a lambertian surface three equations are sufficient to uniquely determine the surface orientation and the reflectance factor $\rho$ at each image point.

Let $\underline{I} = [I_1, I_2, I_3]'$ be the column vector of intensity values recorded at a point $(x,y)$ in each of the three views and let

$\underline{n_1} = [n_{11}, n_{12}, n_{13}]'$

$\underline{n_2} = [n_{21}, n_{22}, n_{23}]'$

$\underline{n_3} = [n_{31}, n_{32}, n_{33}]'$

be unit column vectors defining the three directions of incident illumination, which hence have to be known beforehand.

We can construct the matrix $\underline{N}$ where

$$\underline{N} = \begin{bmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \\ n_{31} & n_{32} & n_{33} \end{bmatrix}.$$

Let $\underline{n} = [n_1, n_2, n_3]'$ be the column vector corresponding to a unit surface normal at $(x,y)$. Then $\underline{I} = \rho \cdot \underline{N} \cdot \underline{n}$, so $\rho \cdot \underline{n} = \underline{N}^{-1} \cdot \underline{I}$, if $\underline{N}^{-1}$ exists, which is the case if and only if $n_1$, $n_2$, $n_3$ do not lie in a plane. In this case the reflectance factor and unit surface normal at $(x,y)$ are given by: $\rho = \left|\underline{N}^{-1} \cdot \underline{I}\right|$

and

$\underline{n} = \frac{1}{\rho} \cdot \underline{N}^{-1} \cdot \underline{I}$.

It should be noted that since the sun's path across the sky is nearly planar, the described technique cannot be used with outdoor images taken at different times during the same day.

Photometric stereo variants have been successfully used in applications as diverse as capturing surface normals of micron-level accurate details [9] to surface normals of large outdoor scenes [10]. Extensions of photometric stereo have allowed orientation to be recovered in the presence of interreflections [11] as well as specularity [12, 13]; a recent survey of photometric stereo work in this area is reviewed in [14]. Another line of work has endeavored to estimate surface orientation specifically from its specular reflection. Many of its successors differ from traditional photometric stereo use extended lighting patterns in order to reliably observe the specular reflection for a certain range of angles. [15] uses light from fluorescent tube lights reflecting onto diffuse surfaces to estimate surface orientation assuming mirror-like specular reflection. [16] uses the eyes specular reflection of a pattern of rings and spline-based surface fitting to model the shape of the cornea. More recently, [17] have combined specular normals obtained using many lighting conditions with positional information to create highly detailed geometry. While recent advances do not require the incident lighting to be known beforehand [18], these methods still require multiple lighting conditions to accurately estimate surface orientation. In contrast, out work requires just a *single* lighting condition, making it better suited for normal estimation in uncontrolled outdoor environments.

Shape estimation under fixed lighting can be obtained from specular cues. Roth and Black [19] introduce the concept of specular flow and illustrate how such specular flow relates to shape under general environmental illumination for surfaces that exhibit (binary) mixtures of diffuse and specular materials. Vasilyev et al. [20] extend specular flow to general unconstrained surface shapes, but they require exact knowledge of the content of the environment and assume that the specular flow is know. However, estimating specular flow from multiple viewpoints is a difficult and complex problem. While the proposed method also relies on specular reflections, it only needs four observations from a *single* viewpoint under constant illumination, avoiding complex specular flow and/or correspondence computations. Furthermore the proposed method naturally handles per-pixel mixtures of diffuse and specular.

Polarization cues have also been used extensively for estimating surface normals. Most methods exploit the property that the *angle of polarization* relates to the surface normal direction perpendicular to the plane of incidence. The direction within the plane of incidence is found by either observing the surface from multiple viewpoints [21, 22, 23, 24], or from the *degree of polarization* [25, 26, 27, 28, 29]. The majority of the above methods infer surface information from a series of photographs of the surface while rotating a linear polarizer in front of the camera. This either requires a large number of photographs, or a careful calibration. The proposed method does not only rely on linear polarization cues, but instead infers surface information from the full characterization of the reflectance (i.e., the Stokes reflectance field), that can be captured in just four photographs with different polarizers in front of the camera (i.e., three linear polarizers rotated 0, 45, and 90 degrees, and a (left) circular polarizer). Furthermore, we show that even under unpolarized incident lighting, some non-negligible circularly polarized reflectance is present, which impacts the accuracy of the estimated normals if ignored.

Ma et al. [30] propose a photometric stereo variant that makes use of four lighting patterns and employs either linear or circular polarization for separating diffuse and specular reflections. Diffuse and specular reflection components differ in that specular reflection alters polarization in a deterministic manner in accordance with the Fresnel equations, while diffuse reflection produces outgoing light which is generally close to unpolarized. They show that the centroids of the diffuse and specular reflectance each yield an estimate of the pixel's surface normal, which is very accurate for the specular component.
The four spherical illumination patterns $P_i(\tilde{\omega})$ needed are:

- a constant pattern $P_c(\tilde{\omega}) = 1$,
- a linear gradient along the $x$-coordinate $P_x(\tilde{\omega}) = \omega_x$,
- a linear gradient along the $y$-coordinate$P_y(\tilde{\omega}) = \omega_y$,
- a linear gradient along the $z$-coordinate $P_z(\tilde{\omega}) = \omega_z$

and , where $\tilde{\omega}$ is the direction defined on a spherical domain $\Omega$. Note that $\tilde{\omega} = [\omega_x, \omega_y, \omega_z]$ are the components of the normalized vector corresponding to the direction $\tilde{\omega}$. Unlike the proposed method, they rely on photometric cues for estimating the surface normal and they need an expensive lighting apparatus (i.e. the USC-ICT *Light Stage*) to create the gradient lighting patterns.

Koshikawa [31] estimate the surface normal of a *single* surface point from the Stokes vector, measured using a ellipsometer, under a single directional circularly polarized light source. They assume known index of refraction, and estimate the normal from the Fresnel equations. However, they ignore the ambiguity in the azimuthal angle of the surface normal. In contrast, the proposed method employs a standard DSLR camera instead of an ellipsometer and considers both circularly polarized *and* unpolarized incident lighting, and explicitly handles the azimuthal ambiguity.

## 2.2 2D Shape Representation

Shape is an important visual feature and it is one of the basic features used to describe image content. Although 3D objects are now commonly used in a number of areas such as games, mechanical design for CAD models, archaeology and cultural heritage, medical research studies and many more, most image processing applications work with 2D images and the majority of shape descriptors developed are built for such images. However, 2-D shape representation and description is a difficult task. This is because when a 3-D real world object is projected onto a 2-D image plane, one dimension of object information is lost. As a result, the shape extracted from the image only partially represents the projected object. To make the problem even more complex, shape is often corrupted with noise, defects, arbitrary distortion and occlusion. Shape representation generally looks for effective and perceptually important shape features based on either shape boundary information or boundary plus interior content. Shape descriptors are computational tools used for analysing image shape information, which consist of mathematical functions which when applied to image produce numerical values, representative of a specific characteristic of the given shape. The nature and meaning of such

values depends on the definition of the shape descriptor. After shape features have been extracted, they can be used as input features for an image processing application. The standard MPEG-7 has set several principles to measure a shape descriptor, that is, good retrieval accuracy, compact features, general application, low computation complexity, robust retrieval performance and hierarchical coarse to fine representation [32]. Good retrieval accuracy requires a shape descriptor be able to effectively find perceptually similar shapes from a database. Perceptually similar shapes usually means rotated, translated, scaled shapes and affinely transformed shapes. The descriptor should also be able to find noise affected shapes, variously distorted shapes and defective shapes, which are tolerated by human beings when comparing shapes. This is known as the robustness requirement. Compact features are desirable for indexing and online retrieval. If a shape descriptor has a hierarchical coarse to fine representation characteristic, it can achieve a high level of matching efficiency. This is because shapes can be matched at coarse level to first eliminate large amount dissimilar shapes, and at finer level, shapes can be matched in details. A desirable shape descriptor should be application independent rather than only performing well for certain type of shapes. Low computation complexity is an important characteristic of a desirable shape descriptor. For a shape descriptor, low computation complexity means minimizing any uncertain or ad hoc factors that are involved in the derivation processes, hence it implies clarity and stability.

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

There are some properties which are desirable for shape descriptors in general, although they are not, strictly speaking, requirements for all shape descriptors.

The following is a list of some of these desirable properties which by no means intends to be extensive:

- ***Rotation invariance.*** The rotation of an object does not affect its shape; therefore it would be expected that a shape descriptor should produce the same measure for a shape $S$ and for the same shape rotated by $\theta$ degrees, $R(S;\theta)$,

- ***Translation invariance.*** The shape of an object is independent of the coordinate axes used; therefore it would be expected that a shape descriptor should produce the same measure for the shape $S$ regardless of its location in the coordinate plane.

- ***Scale invariance.*** Because the shape of an object is independent of its representation, the scale of an object should not affect the measure produced by a shape descriptor.

- ***Well defined range.*** Having an idea of the range of values produced by a shape descriptor can be important when interpreting the meaning of the values produced by the descriptor and it is useful to know the range produced by a descriptor in particular when designing an application, in particular if normalisation is required.

### 2.2.1 Statistical Shape Models

Most real-world applications need an automated system to recover image structure and understand its meaning. This necessarily involves the use of models which describe and label the expected structure of the world. Real applications, such as in medical image interpretation, are also typically characterised by the need to deal with complex and variable structure. To deal with structural complexity, some prior knowledge of the problem can be used, which could also provide tolerance to noisy or missing data, and provide a means of labelling the recovered structures. Generative models, if complete enough to be able to generate realistic images of target objects, are of particular interest. In fact, using such a model, image interpretation can be formulated as a matching problem: given an image to

interpret, structures can be located and labelled by adjusting the model's parameters in such a way that it generates an image which is as similar as possible to the real thing.

In real applications there is a need to deal with variability, because they have to handle classes of objects which are not identical, for example female breasts. Hence, a suitable model should maintain the essential characteristics of the class of objects it represents and it should be able to deform to fit a range of examples. Such *deformable model* should be *general*, i.e. being able of generating any plausible example of the class it represents and it should only generate legal examples, for instance it could only deform in ways found in a given training set. The latter characteristic, being *specific*, is very important to correctly interpret and label objets.

The "Point Distribution Model" [33] is a computational tool for representing the mean geometry of a shape and some statistical modes of geometric variation inferred from a training set of shapes, represented by a set of labelled points. The technique determines the statistics of the points over the training set: the mean positions of the points give an average shape and a number of modes of variation are determined describing the main ways in which the example shapes tend to deform from the average, thus allowing to include variation in shape in the model. The method produces a compact and flexible point distribution model, with a small number of linearly independent parameters.

In order to be modeled, a shape need to be represented as a set of points, for instance chosen around the boundary. The method works by modelling how different labelled points tend to move together as the shape varies, examining the statistics of the co-ordinates of the labelled points over the training set. Shapes must be aligned in the same way with respect to a set of axes by scaling, rotating and translating the training shapes so that they correspond as closely as possible, using for instance the "Generalised Procrustes Analysis" [34].

Let $x_i$ be a vector describing the $n$ points of the $i^{th}$ shape in the set, $x_i = (x_{i0}, y_{i0}, x_{i1}, y_{i1}, \cdots, x_{in-1}, y_{in-1})'$.

Let $M_j\left[x_j\right]$ be a rotation by $\theta_j$, and a scaling by $s_j$. Given two similar shapes $x_i$ and $x_j$ we can choose $\theta_j$, $s_j$ and a translation $(t_x, t_y)_j$ mapping $x_i$ onto $M_j\left[x_j\right]$ so as to minimise the weighted sum

$$E_j = (x_i - M_j(x_j))' W (x_i - M_j(x_j)) \tag{2.2.1}$$

where $M_j \binom{x_{jk}}{y_{jk}} = \binom{(s_j \cos\theta) x_{jk} - (s_j \sin\theta) y_{jk} + t_{jx}}{(s_j \sin\theta) x_{jk} + (s_j \cos\theta) y_{jk} + t_{jy}}$ and $W$ is a diagonal matrix of weight for each point. The weights can be chosen to give more significance to those points which move "less" than other points in a shape. Once a set of aligned shapes is available the mean shape and variability can be found. The mean shape $\bar{x}$ is calculated using

$$\bar{x} = \frac{1}{N_s} \sum_{i=1}^{N_s} x_i. \tag{2.2.2}$$

The modes of variation, that are the ways in which the points of the shape tend to move together, can be found by applying *principal component analysis* to the deviations from the mean: for each shape in the training set the deviation from the mean $dx_i$ is calculated as: $dx_i = x_i - \bar{x}$.

The next step is to calculate the $2n \times 2n$ covariance matrix $S$ using

$$S = \frac{1}{N_s} \sum_{i=1}^{N_s} dx_i dx_i' \tag{2.2.3}$$

The modes of variation of the points of the shape are described by the unit eigenvectors $p_i$ of $S$, $i = (1 \ldots 2n)$ such that $Sp_i, = \lambda_i p_i$, $p_i' p_i = 1$ where $\lambda_i$ is the $i^{th}$ eigenvalue of $S$, $\lambda_i \geq \lambda_{i+1}$.

It can be demonstrated that the eigenvectors of the covariance matrix corresponding to the largest eigenvalues describe the most significant modes of variation in the variables used to derive the covariance matrix, and that the proportion of the total variance explained by each eigenvector is equal to the corresponding eigenvalue [35].

The total variance $\lambda_T$ of all variables is given by $\lambda_T = \sum_{i=1}^{2n} \lambda_i$. Usually only a small number $t$ modes is sufficient to explain most of the observed variation, and the error discarding the remaining $2n - t$ modes is mathematically bound and easy to estimate. A pratical way to determine a good value for $t$ is to chose the smallest number of modes such that the sum of variance explained was a sufficiently large

proportion of $\lambda_T$.

The $i^{th}$ eigenvector affects point $k$ in the model by moving it along a vector parallel to $(dx_{ik}, dy_{ik})$, which is obtained from the $k^{th}$ pair of elements in $p_i$, $(dx_{i0}, dy_{i0}, \ldots, dx_{ik}, dy_{ik}, \cdots, dx_{in-1}, dy_{in-1})$ Any shape in the training set can be approximated using the mean shape and a weighted sum of these deviations obtained from the first $t$ modes $x = \bar{x} + Pb$ where $P = (p_1 p_2 \ldots p_t)$ is the matrix of the first $t$ eigenvectors, $b = (b_1 b_2 \cdots b_t)$ is a vector of weights for each eigenvector. The eigenvectors are orthogonal, $P'P = I$ so $b = P'(x - \bar{x})$.

The above equations allow us to generate new examples of the shapes be varying the parameters $(b_i)$ within suitable limits. The parameters are linearly independent, though there may be non-linear dependencies still present. The limits for $b_i$ are derived by examining the distributions of the parameter values required to generate the training set. Since the variance of $b_i$ over the training set can be shown to be $\lambda_i$, suitable limits are likely to be of the order of $-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i}$ since we are assuming a multivariate gaussian distribution and most of the population lies within three standard deviations of the mean. The described Point Distribution Model is defined on 2-D images, but it can be extended to deal with volume data, considering a set of labeled 2-D slices of a 3-D object.

### 2.2.2 Contour-based shape descriptors

Existing techniques for 2-D shape representation and description can be generally classified in contour-based methods and region-based methods. Area based methods are not sensitive to small changes on the shape and even large changes on the boundary have a small effect on the shape area. Hence, these methods are not greatly affected by the presence of noise, but this also means that they are unable to perceive small variations on the shape, which implies that shape details are ignored.

Contour shape techniques, on the contrary, only exploit shape boundary information and can perceive small changes in the shape, however this implies that such methods can be affected by the presence of noise in the shape, producing undesirable results: a robust distance measure is strongly advisable. There are two types

of very different approaches for contour shape modeling: continuous approach (global) and discrete approach (structural). Continuous approaches do not divide shape into sub-parts: usually a feature vector derived from the integral boundary is used to describe the shape. A suitable measure of shape similarity for a continuous approach is a metric distance between the acquired feature vectors. A commonly used metric distance is the Hausdorff metric, which measures how far two subsets of a metric space are from each other. Hausdorff distance is a classical correspondence-based shape matching method and it has often been used to locate objects in an image and measure similarity between shapes [36]. Given two shapes represented by two set of points: $A = a_1, a_2, \ldots, a_p$ and $B = b_1, b_2, \ldots, bq$, the Hausdorff distance between $A$ and $B$ is defined as

$$H(A,B) = \max\left(h(A,B), h(B,A)\right) \tag{2.2.4}$$

where $h(A,B) = \max_{a\in A}\min_{b\in B}\|a-b\|$ and $\|\cdot\|$ is the underlying norm on the points of $A$ and $B$, usually Euclidean distance. In other words, given two sets, the Hausdorff distance is the greatest of all the distances from a point in one set to the closest point in the other set. If every point of either set is close to some point of the other set then they are close in the Hausdorff distance.
However, this distance measure is too sensitive to noise or outlier. A single point in $A$ that is far from anything in $B$ will cause $h(A,B)$ to be large. Rucklidge [37] introduce a modified Hausdorff distance, to address the problem of noise sensitivity of the original definition:

$$h^f(A,B) = f^{th}_{a\in A}\min_{b\in B}\|a-b\| \tag{2.2.5}$$

where $f^{th}_{x\in X}g(x)$denotes the $f^th$ quantile value of $g(x)$ over set $X$, for some value of $f$ between 0 and 1. For example, the $1th$ quantile value is the maximum and the $1/2th$ quantile value is the median. In practice, $f$ is usually set to be $1/2$. The advantage of shape matching using Hausdorff distance is that shape can be matched partially. However, the Hausdorff distance is not translation, scale and rotation invariant. In order to match a model shape with a shape in the image, the model shape has to be overlapped on the image in different positions, different

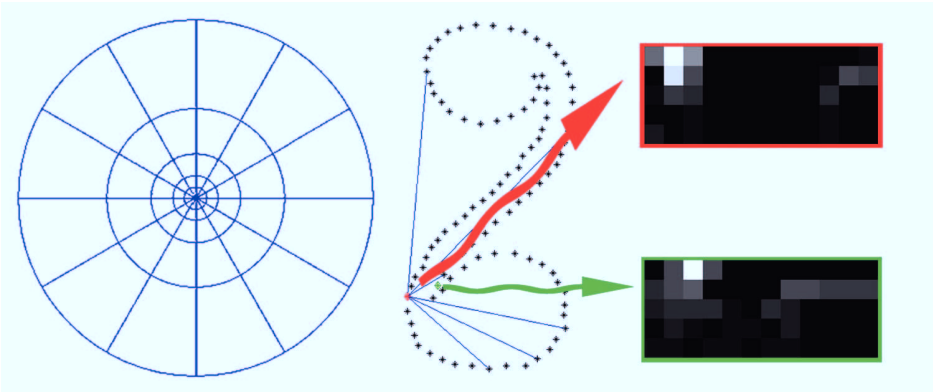


Figure 2.1: The log polar maps of the contest of two different points in the contour of a shape.

orientations and different scales. As the result, the matching is prohibitively expensive.

In our work we focused on contour-based descriptors, which can deal with articulated and flexible shapes, allowing us to better classify 2-D shapes in various applications.

### Shape Context

Belongie et al. propose a correspondence-based shape matching method using shape contexts [38]. Shape matching using shape contexts is an improvement to traditional Hausdorff distance based methods. It extracts a global feature, called shape context, for each corresponding point. The matching between corresponding points is then the matching between the context features. In this approach an object is treated as point set and it is assumed that the shape of the object can be captured by a finite subset of its points, for example sampled uniformly from the external contour. These points do not need to be key-points such as inflection points, cusps, extrema of curvature, etc., and can be derived from edge pixels found by a simple edge detector, whose output is a set $P$ of $n$ points $P = \{p_1, p_2, \ldots, p_n\}$, $p_i \in \mathbb{R}^2$.

The shape context considers the set of $n-1$ vectors originating from a point to all other sample points on a shape, which expresses the configuration of the whole

shape with respect to the reference point. Since shapes may vary from an instance to another in the same category, the full set of vectors is too detailed and does not guarantee robustness. The distribution over relative position is a more robust and compact description, hence for each point $p_i$ on the shape, a coarse histogram $h_i$ of the relative $n-1$ coordinates is computed:

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in bin(k)\} \tag{2.2.6}$$

The length $r$ and orientation $\theta$ of the vectors $(q - p_i)$ are quantized to create a histogram map which is the shape context used to represent the point $p_i$. To make the histogram more sensitive to the location of nearby points than to the location of points farther away, these vectors are mapped into a log-polar space representation (Fig. 2.1). Shape contexts are distribution represented as histograms: it is hence possible to use the $\chi^2$ test statistic to define the cost of matching two points. Consider two shapes $P$ and $Q$. A point $p_i$ on the first shape and a point $q_j$ on the second shape, $C_{ij} = C(p_i, q_j)$ denotes the cost of matching the two points in exam, where $C(p_i, q_j) = 1/2\Sigma_{k=1}^{K} \frac{[h_i(k)-h_j(k)]^2}{h_i(k)+h_j(k)}$ and $h_i(k)$, $h_j(k)$ denote the $K - bin$ normalized histogram respectively at $p_i$ and $q_j$. Given the set of costs $C_{ij}$ between all pairs of points $p_i$ of $P$ and $q_j$ of $Q$, the total cost of matching $H(\pi) = \Sigma_i C(p_i, q_{\pi(i)})$ has to be minimized, with the constraint of one-to-one matching. The matching of two shapes then is done by matching two context maps of the shapes, which is a matrix-based matching. It minimizes the total cost of matching between one context matrix and all the permutations of another context matrix. To reduce the matching overhead, the shortest augmenting path algorithm for the matrix matching is used. When the number of sample points is not equal on the two shapes, the matching is done adding dummy nodes to the smaller point set, with a constant matching cost of $\varepsilon_i$. Since all measures are taken with respect to points on the object, invariance to translation is intrinsic to the definition of the descriptor. The use of histograms makes shape context insensitive to small deformations, non linear transformation occlusions and presence of outliers. Additional robustness to outlier can be added not allowing points labeled as outlier to contribute to any histogram. The above definition of shape context is not invariant to rotation, which can be obtained using a relative frame instead of the absolute frame described, e.g.

treating the tangent vector at each point as the positive x-axis.

### Contour Flexibility

It this Section we briefly report the main ideas about contour flexibility. The reader is urged to find the details of this algorithm in [39]. Vision science has developed as an interdisciplinary research field, frequently involving concepts and tools from computer science, image processing, biology, psychology and cognitive science, for example the importance of high curvature points in shape perception has been already described in [40] and [41]. From this theories the important features for shape recognition come from parts of the object where changes occur (e.g. corners). A recent descriptor, named contour flexibility [39], depicts the deformable potential at each point along a closed curve, extracting both global and local features, with proper trade-off between them. It can be observed that in most cases articulated high flexible parts correspond to high curvature points. Let $\ell$ be a closed simple contour, surrounding a bounded domain $D$. For a point $p$ on a contour $\ell$ and a given radius $r$, the contour flexibility is defined as $\omega(p,r) = min(\omega_+(p,r), \omega_-(p,r))$, where $\omega_+$ is the interior flexibility and $\omega_-$ is the exterior flexibility:

$$\omega_+(p,r) = \frac{\int_{C^+_{p,r}} k_+(x)dx}{\int_{C^+_{p,r}} dx} \tag{2.2.7}$$

$$\omega_-(p,r) = \frac{\int_{C^-_{p,r}} k_-(x)dx}{\int_{C^-_{p,r}} dx} \tag{2.2.8}$$

$C^+_{p,r}$ and $C^-_{p,r}$ respectively are the connected components containing $p$ in the sets $\{x \in D : \| x-p \| \leq r\}$ and $\{x \in \mathbb{R}^2 \backslash D : \| x-p \| \leq r\}$. $k_+$ and $k_-$ are two function of distance transform on

$$k_+(x) = d(x, \mathbb{R}^2 \backslash D),\ x \in \mathbb{R}^2 \tag{2.2.9}$$

$d(\cdot,\cdot)$ is the minimum Euclidean distance between two sets. The radius $r$ is called bendable size and should be tuned taking into account the width of the limb-like parts of an object. The bendable size $r$ and the contour flexibility are proportional

to the scale of the contour but invariant to translation, rotation and the choice of starting point for the parameterization of the contour. Let

$$z(t) = (x(t), y(t)), 0 \leq t < 1 \tag{2.2.10}$$

be the arc-length parameterization of a contour $\ell$. Since it is difficult to match flexible parts of two contours (e.g., using uniform sampling of the contour), a better strategy is to give large weights to inflexible landmarks and smaller weights to more flexible landmarks, using more samples on the segments of a contour which are more inflexible. Let

$$\gamma(t) = \frac{\int_0^t \omega(z(u), r)du}{\int_0^1 \omega(z(u), r)du} \tag{2.2.11}$$

the optimal sampling can be obtained with sampling speed $d\gamma/dt$. Considering each landmark a complex number $z^k = x^k + jy^k$, the contour can be treated as a complex vector $\Phi = (z^1, z^2, \ldots, z^n)^T$ and the Procrustean Distance between two sequences can be used for global matching of $\Phi_1$ and $\Phi_2$:

$$\hat{d}(\bar{\Phi}_1, \bar{\Phi}_2) = cos^{-1}(max_{s \in N} |\langle \bar{\Phi}_1, \sigma^s(\bar{\Phi}_2) \rangle|) \tag{2.2.12}$$

where $\sigma^s(\bar{\Phi}_2)$ is a cyclic permutation of $\bar{\Phi}_2$, to achieve independence from the starting point and " $<>$ " denotes the inner product of two complex vectors. The matching score between the two shapes represented by $\bar{\Phi}_1$ and $\bar{\Phi}_2$ is determined by

$$M_d = \frac{\alpha}{\pi}\hat{d}(\bar{\Phi}_1, \bar{\Phi}_2) + D(\Omega_1, \Omega_2) \tag{2.2.13}$$

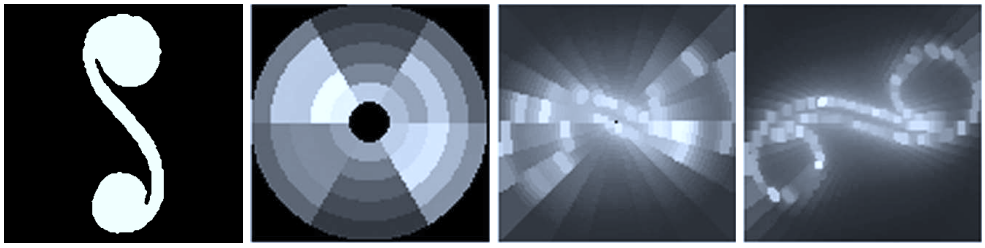
Where $D(\cdot, \cdot)$ is the warping distance between the two sequences of the contour flexibility values extracted from the contours; $\Omega_1$, $\Omega_2$ are the sequences of the values of the contour flexibility at each landmark of uniformly sampled sequences, starting form the leading landmark found $\hat{s}$ by (2.2.12), $\alpha$ weighting factor, $\pi$ normalization factor.

### Circular Blurred Shape Model

Circular Blurred Shape Model is a recent rotationally invariant descriptor, which makes use of a correlogram structure to capture spatial arrangement of object parts, shared among regions defined by circles and sections [42]. The correlogram structure can be defined as follows: given a number of circles $C$, number of sections $S$, and an image region $I$, a correlogram $B = \{b_{\{1,1\}}, \ldots, b_{\{C,S\}}\}$ is a radial distribution of sub-regions of the image. In each region $b$ of the correlogram the centroid $b*$ can be located by its coordinates and used as reference point. The regions around $b$ define the neighborhood of $b$, and the number of neighbors depends on the spatial location $b$ (e.g. inner circle, middle circle or extern circle).

The descriptor makes use of the information obtained from the contour of the object, which can be extracted for example by means of an edge detector. Every point $x$ in the contour map is taken into account, calculating first the distance to the corresponding centroid of the region in which $x$ lies, then the distances to the centroids of neighbor regions. The inverse of these distances are computed and normalized by the sum of total distances and the obtained values are then added to the corresponding positions of the descriptor vector $v$, which as dimension $C \times S$. As in the case of shape context, the use of histograms makes description tolerant to irregular deformations. As for the complexity, for a map of $k$ relevant contour points, the computation of the descriptor requires $O(k)$ simple operations. Parameters $C$ and $S$ defines the degree of spatial information taken into account in the description process (i.e., as the number of regions increase the description becomes more local) and should be tuned for each particular application. See also Fig. 2.2.

To obtain a rotationally invariant descriptor a second step needs to be included. The main diagonal $G_i$ of correlogram $B$, with the highest density, is searched. This diagonal is then used as a reference to rotate the descriptor. The orientation of the rotational process, so that $G_i$ is aligned with the x-axis, is that corresponding to the highest description density at both sides of $G_i$. Once obtained the rotationally invariant CBSM descriptor, it can be used to design a symbol spotting methodology (e.g., using Adaboost to distinguish between foreground and background) or a multi-class classifier by embedding binary classifiers, for instance using Adaboost



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

Figure 2.2: Circular Blurred Shape Models of the same shape with different choices of the parameters $C$ and $S$.

to define a classifier based on the features that best discriminate one class against another, combining then binary classifiers in a Error Correcting Code Framework.

# Chapter 3

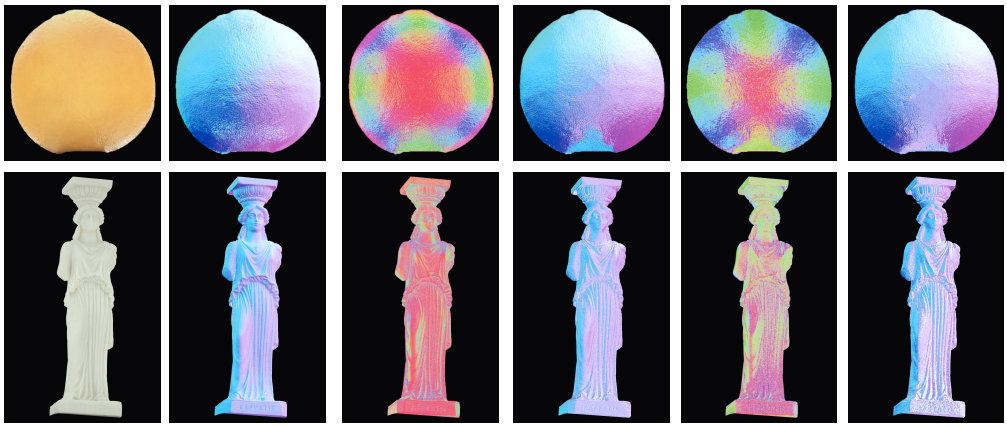
# Surface Normals Estimation from Spherical Incident Lighting

## 3.1 Introduction

Accurate shape and appearance estimation is a crucial component in many computer vision applications. Despite being at the focal point of intense research, shape and appearance estimation remains a challenging problem, especially under uncontrolled real-world conditions. Estimation of both small-scale surface shape as well as surface reflectance greatly benefits from accurate knowledge of surface orientation. In this work, we propose a novel method for estimating surface orientation from the Stokes polarization vector under a *single* spherical incident illumination condition that is either circularly polarized (Fig. 3.1,c-d) or unpolarized (Fig. 3.1, e-f).

Polarization cues have previously been employed to separate diffuse and specular reflectance components (e.g., [43, 44, 45, 30, 46, 47, 48, 49, 50], to classify materials (e.g., [43, 51, 52, 53], to estimate reflectance properties (e.g., [54, 55]), and to estimate surface normals (e.g., [56, 21, 25, 57, 58, 31, 22, 26, 27, 23, 28]. Motivated by the polarization characteristics of natural lighting, which is either unpolarized (overcast sky) or linearly polarized (sunlight), most of these methods, with exception of [30, 50, 31] focus solely on linear polarization cues. In contrast, in this work we leverage observations of the view-independent symmetric Stokes

reflectance field – which encodes the impact of unpolarized, linearly polarized, as well as circularly polarized reflected light – for estimating surface normals under *constant* incident spherical illumination. We show that for many dielectric materials, a non-negligible circularly polarized reflectance is observed, not only under circularly polarized incident lighting, but also for unpolarized lighting. Leveraging these circular polarization cues greatly improves normal estimation for front facing surfaces. We demonstrate that both types of incident lighting can be used to reliably estimate surface normals from observations of the Stokes reflectance field, and show how this theory can be applied to normal estimation under uncontrolled outdoor illumination. The techniques discussed in this Chapter have been published in [59, 60, 61].



(a) Subject (b) Photometric Normals (c) Circular Stokes Param. (d) Circ. Pol. Normals (e) Unpol. Stokes Param. (f) Unpol. Normals

Figure 3.1: Estimating surface normals (encoded as $\frac{1}{2}(\mathbf{x}+1) \rightarrow R$, $\frac{1}{2}(\mathbf{y}+1) \rightarrow G$, and $\frac{1}{2}(\mathbf{z}+1) \rightarrow B$) from Stokes parameters (encoded as $|s_3| \rightarrow R$, $|s_1| \rightarrow G$, and $|s_2| \rightarrow B$) of specularly reflected incident spherical illumination. Surface normals inferred from circularly polarized illumination (c-d), and unpolarized illumination (e-f), compared to surface normals obtained from a photometric stereo variant [30](b). Top-row: Plastic orange - $\phi$-ambiguity resolved by growing normals inward. Bottom-row: Marble statue - $\phi$-ambiguity resolved using an additional measurement.

## 3.2 Surface Normal Estimation from Stokes Vectors

In previous work [54] it has been shown that the Stokes vector may be used to recover reflectance parameters, if the incident lighting is circularly polarized. Here we show how to infer per-pixel surface normal **n** from the observation of the scene under either circularly polarized or unpolarized spherical illumination.

### 3.2.1 Background: Mueller Calculus

Before detailing how surface normals can be estimated from measurements of the Stokes reflectance field under a single spherical illumination condition, we first give a brief overview of the related theory on Stokes parameters and Mueller calculus. We refer the interested reader to [62] for a more in depth overview of polarization and Mueller calculus.

In what follows we assume that the surface consists of a homogeneous dielectric material, and the coordinate frame is that of the camera, i.e., the camera looks down the -Z axis, the X axis points right, and the Y axis points up.

Polarization describes the mutually perpendicular components of the transverse oscillation of light waves (see Appendix **??** for more details). According to Mueller calculus, the polarization state of light can be described by the 4-element Stokes vector $\mathbf{s} = (s_0, s_1, s_2, s_3)$, where: $s_0$ represents the total power, $s_1$ is the power of the $0^\circ$ linear polarization, $s_2$ is the power the $+45^\circ$ linear polarization, and $s_3$ is the power of right circular polarization. The degree of polarization, or ratio of the power of polarized light versus the total power equals:

$$DOP = \frac{\sqrt{s_1^2 + s_2^2 + s_3^2}}{s_0}. \tag{3.2.1}$$

The Stokes parameters are related to the polarization ellipse coordinates by:

$$2\psi = \arctan\frac{s_1}{s_2}, \tag{3.2.2}$$

$$2\chi = \arctan\frac{s_3}{\sqrt{s_1^2 + s_2^2}}. \tag{3.2.3}$$

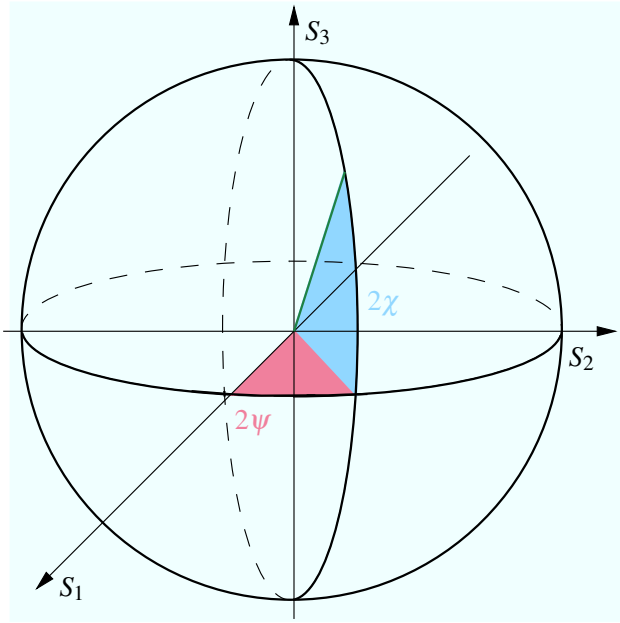



Figure 3.2: The last three Stokes parameters ploted as spherical coordinates on a Poincaré sphere.

Fig. 3.2 illustrates this relation on a Poincaré sphere. Note that the radius of the sphere is the degree of polarization times $s_0$.

When a polarized ray hits a reflective surface (i.e., specular), the resulting change to Stokes vector is predicted by Mueller calculus:

$$\mathbf{s}' = \mathbf{C}(\phi)\mathbf{D}(\delta;\mathbf{n})\mathbf{R}(\theta;\mathbf{n})\mathbf{C}(-\phi)\mathbf{s}. \tag{3.2.4}$$

Each of the linear operators **C**, **D** and **R** can be compactly represented by a matrix. **C** is the Mueller rotation that brings the linear polarization parameters into a canonical frame:

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\phi & -\sin 2\phi & 0 \\ 0 & \sin 2\phi & \cos 2\phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.2.5}$$

where $\phi$ is the angle between the the plane of incidence and the **x** axis in the

observation coordinate system (the camera in our case). The plane of incidence is the plane containing the incident ray, the exitant ray, and the surface normal $\mathbf{n}$, hence $\phi$ is equal to $\arccos(\mathbf{n}\cdot\mathbf{x})$.
$\mathbf{R}$ is the Mueller reflection matrix:

$$\mathbf{R} = \begin{pmatrix} \frac{\mathbf{R}_\parallel+\mathbf{R}_\perp}{2} & \frac{\mathbf{R}_\parallel-\mathbf{R}_\perp}{2} & 0 & 0 \\ \frac{\mathbf{R}_\parallel-\mathbf{R}_\perp}{2} & \frac{\mathbf{R}_\parallel+\mathbf{R}_\perp}{2} & 0 & 0 \\ 0 & 0 & \sqrt{\mathbf{R}_\parallel\mathbf{R}_\perp} & 0 \\ 0 & 0 & 0 & \sqrt{\mathbf{R}_\parallel\mathbf{R}_\perp} \end{pmatrix}, \tag{3.2.6}$$

where $\mathbf{R}_\parallel$ and $\mathbf{R}_\perp$ are the Fresnel equations for the parallel and perpendicular components, respectively, as functions of the incident angle $\theta$ (i.e., $\arccos(\mathbf{n}\cdot\mathbf{z})$). Finally, $\mathbf{D}$ is the Mueller retardation matrix:

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\delta & \sin\delta \\ 0 & 0 & -\sin\delta & \cos\delta \end{pmatrix}, \tag{3.2.7}$$

and $\delta$ is the phase shift. The phase shift differs depending on whether the material is a dielectric or a metal. For dielectric materials it is either $180^\circ$ when $\theta \leq \theta_B$, or $0^\circ$ when $\theta > \theta_B$, where $\theta_B$ is the Brewster angle. For dielectric materials, this essentially flips the signs of $s_2$ and $s_3$ over the Brewster angle.

Finally, the Stokes reflectance field [54] is defined as the description of the Stokes vectors resulting from a single surface interaction, under a user-defined incident light field, computed for every surface normal direction.

### 3.2.2 Measurement of Stokes Vectors

Direct observation of the Stokes vectors is difficult because typical polarizers do not block unpolarized light completely. In order to minimize the impact of polarizer inefficiency, the Stokes parameters are computed indirectly from four different measurements, following the acquisition scheme of [54] to capture the Stokes reflectance field. Specifically, four photographs of a surface are recorded with

four different polarizers in front of the camera: a linear polarizer rotated $0^\circ$ ($P_H$), $45^\circ$ ($P_{45}$), $90^\circ$ ($P_V$), and a (left) circular polarizer ($P_\circ$). Note that each polarized image receives $(50+\varepsilon)\%$ unpolarized radiance. We can then robustly compute the Stokes vector components as: $s_0 = P_H + P_V$, $s_1 = P_H - P_V$, $s_2 = 2P_{45} - s_0$, and $s_3 = s_0 - 2P_\circ$. Assuming similar efficiencies for the linear and circular polarizers, this scheme has the advantage that any unpolarized light that passes through the polarizers will be canceled out in the computation of $s_1$, $s_2$, and $s_3$, because of the subtraction. Since $s_0$ may overestimate the unpolarized power by a factor of $2\varepsilon$, it will not be used in any of our computations to estimate surface orientation.

### 3.2.3 Uniform Circularly Polarized Incident Lighting

We will develop our theory first on the case where the incident illumination has power $\Phi$ over the sphere of incident directions and is circularly polarized. In other words $\mathbf{s}(\omega) = \mathbf{s} \sim (\Phi, 0, 0, \Phi)$. Applying Mueller calculus for diffuse and specular surface interactions yields the Stokes reflectance vector $\mathbf{s}'$:

$$\begin{aligned} \mathbf{s}' &= (s'_0, s'_1, s'_2, s'_3) \\ &= (\rho_s \Phi \frac{\mathbf{R}_\perp + \mathbf{R}_\parallel}{2} + \rho_d \Phi, \\ &\quad \rho_s \Phi \frac{\mathbf{R}_\perp - \mathbf{R}_\parallel}{2} \cos 2\phi, \\ &\quad \rho_s \Phi \frac{\mathbf{R}_\perp - \mathbf{R}_\parallel}{2} \sin 2\phi, \\ &\quad \mp \rho_s \Phi \sqrt{\mathbf{R}_\parallel \mathbf{R}_\perp}), \end{aligned} \tag{3.2.8}$$

where $\rho_d$ and $\rho_s$ are the diffuse and specular albedo of the dielectric material. Note that $s'_0$ is the result of both specular as well as diffuse reflectance. While the impact of diffuse reflectance can be easily removed from $s'_0$ using the degree of polarization [54], we avoid using this component in our computations to minimize the impact of potential diffuse pollution (e.g., due to suboptimal polarizers). Note that $s'_1$, $s'_2$ and $s'_3$ are the result of specular reflections only, and hence unaffected by any diffuse pollution.

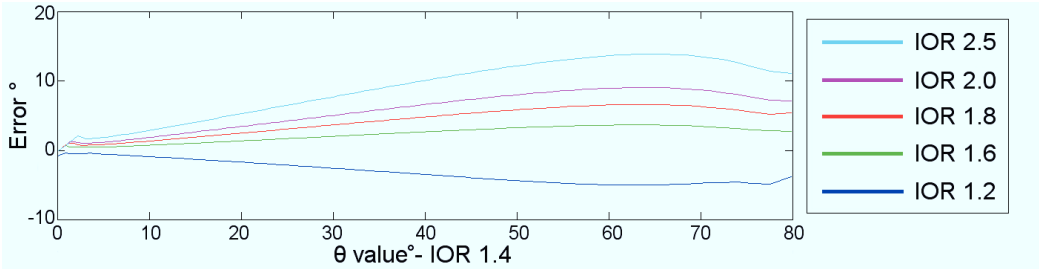



Figure 3.3: Error plots for $\theta$ estimated with a fixed index of refraction of 1.4 for materials with increasing index of refraction.

To compute $\theta$, we establish a relation $\xi$ between the Stokes components $s_1'$, $s_2'$ and $s_3'$ that is independent of specular albedo $\rho_s$ and the power of the incident light source $\Phi$:

$$\xi = \arctan\left(\frac{s_3'}{\sqrt{s_1'^2 + s_2'^2}}\right) = \arctan\left(\mp 2\frac{\sqrt{\mathbf{R}_\parallel \mathbf{R}_\perp}}{\mathbf{R}_\perp - \mathbf{R}_\parallel}\right). \tag{3.2.9}$$

Note that $\xi$ is implicitly related by a non-linear one-to-one mapping to $\theta = arccos(\mathbf{z}\cdot\mathbf{n})$ via the Fresnel equations $\mathbf{R}_\perp$ and $\mathbf{R}_\parallel$. $\theta$ can be directly computed from the Fresnel equations; in our implementation we invert this non-linear mapping by precomputing a lookup table that maps $\xi$ to $\theta$, obtained by evaluating Eq. (3.2.9) for a dense sample of $\theta$, and assuming a fixed index of refraction of 1.4.

Fig. 3.3 illustrates the impact of on the accuracy of the recovered $\theta$ when the true index of refraction differs from the fixed index. As can be seen, the error remains below 5° for moderate indexes of refraction and for $\theta < 60$. Alternatively, the index of refraction can be accurately computed, for each pixel, taking into account the exact amount of total reflected radiance, which is governed by the well-known Fresnel equations (see A.0.2 and A.0.3). From this we can write $\eta$ as a function of $\mathbf{R}_\perp$ and $\mathbf{R}_\parallel$, which can be inferred from the Stokes vector; see [54] for a thorough description of this method.

The remaining azimuthal angle $\phi$ can be directly computed from the linear

components $s'_1$ and $s'_2$:

$$\arctan\left(\frac{s'_2}{s'_1}\right) = \arctan\left(\frac{\sin 2\phi}{\cos 2\phi}\right) = 2\phi. \tag{3.2.10}$$

Note that this is exactly the same as the *angle of polarization* used in prior work. Similarly as in prior work, this relation is ambiguous: $\phi$ and $\phi + \pi$ both satisfy the above equation. Resolving this ambiguity as been the focus of many prior work. In this work, we employ two alternative strategies:

1. For convex objects, we can grow the normals in from the silhouette, assuming that the normals at the silhouette are orthogonal to silhouette edge and the view direction. This strategy was also employed in prior work such as [25].
2. Alternatively, we can capture an additional photograph of the surface while lit by another known spherical illumination condition $I(\cdot,\cdot)$ such that $I(\phi,\theta) \neq I(\phi+\pi,\theta)$. This strategy is useful when incident lighting can be precisely controlled (e.g., laboratory setting). We will also employ a variant of this for normal estimation in uncontrolled outdoor lighting.

It can be shown that Equations (3.2.9) and (3.2.10) are closely related to the spherical coordinates of Poincaré sphere (respectively $2\chi$ and $2\psi$).

### 3.2.4 Uniform Unpolarized Incident Lighting

In the case that the incident lighting is unpolarized (i.e., the Stokes vector is $(\Phi,0,0,0)$), Mueller calculus for diffuse and specular surface interactions predicts the following resulting reflected Stokes vector:

$$\mathbf{s}' = \left(\rho_s\Phi\frac{\mathbf{R}_\perp + \mathbf{R}_\parallel}{2} + \rho_d\Phi, \rho_s\Phi\frac{\mathbf{R}_\perp - \mathbf{R}_\parallel}{2}\cos 2\phi, \rho_s\Phi\frac{\mathbf{R}_\perp - \mathbf{R}_\parallel}{2}\sin 2\phi, 0\right). \tag{3.2.11}$$

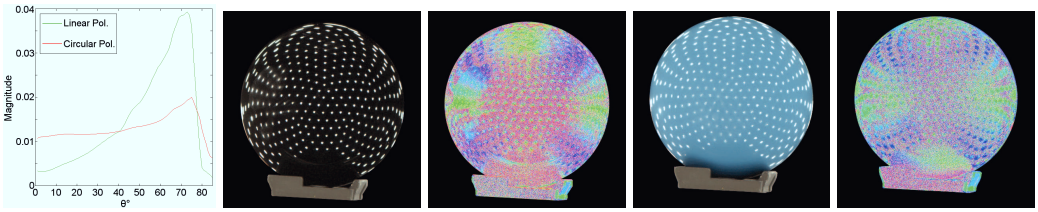
Observe that $s'_1$ and $s'_2$ are solely due to specular reflections. $\phi$ can be computed similarly as before using Eq. (3.2.10). However, Eq. (3.2.9) cannot be employed for estimating $\theta$, because the circular Stokes component $s'_3$ differs. While a relation can be expressed in terms of $s'_1$ and $s'_2$ (e.g., $\sim s'^2_1 + s'^2_2$), such a relation will suffer from a low SNR when $\mathbf{R}_\perp \approx \mathbf{R}_\parallel$. While no circular polarization is

predicted by Eq. (3.2.11), we experimentally detected a small quantity of left circularly polarized reflectance under unpolarized incident illumination. We observe that, while fairly constant for different angles of $\theta$, it is stronger than the observed amount of linearly polarized reflectance near normal incidence (Fig. 3.4, (a)) . This provides a means to improve the accuracy of surface normals for front facing surfaces. We believe that this observed circularly polarized reflectance is due to polarization preserving (subsurface) scattering, which was not taken into account when computing Eq. (3.2.11). Fig. 3.4, (b-e) shows an experimental validation that indicates that the observed circularly polarized reflectance is not due to specular reflections (i.e., non-zero $s_3'$ is observed between the sharp highlights). Furthermore, it has been shown that scattering from randomly oriented particles can give rise to circular polarization (see [63], p.451), and that polarization is preserved for an average of 2.5 scattering events [64]. Exact modeling of the corresponding Mueller matrix, and thus the resulting $s_3'$, for polarization preserving scattering is difficult due to the many unknown factors. Instead of relying on an exact formulation of $s_3'$ for computing $\theta$, an example-based strategy is employed. The Stokes reflectance field of a dielectric object with known shape (e.g., a sphere) is recorded under unpolarized incident lighting. To account for differences in specular albedo and scattering properties, a first-order correction is performed by scaling the maximum of $\sqrt{s_1'^2 + s_2'^2}$ and of $s_3'$ in the target dataset to match those of the exemplar dataset. This first-order correction works well for objects that exhibit a rich variation in surface normals. Finally, the normal of a surface point is computed by finding the best matching $\sqrt{s_1'^2 + s_2'^2}$ and $s_3'$ pair on the exemplar.

### 3.2.5 Uncontrolled Outdoor Illumination

The extension of the proposed method to uncontrolled outdoor environments builds on two observations:

- We observe that overcast sky is unpolarized, and the content varies approximately as: $I(\phi, \theta) \sim sin(\phi)$. Such an illumination condition is slowly varying, and fulfills the condition that $I(\phi, \theta) \neq I(\phi + \pi, \theta)$, and thus it is suitable for resolving the $\phi$-ambiguity.



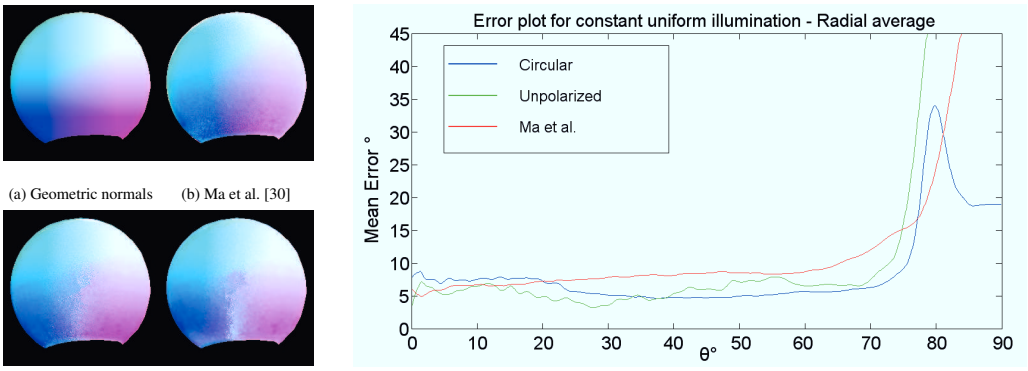

(a) Circular vs Linear mag. (b) Black ball (c) Black ball Unpol. Stokes (d) Blue ball (e) Blue ball Unpol. Stokes

Figure 3.4: (a) Plot of reflected circular vs linear polarization under uniform spherical illumination as function of $\theta$. (b - e) Stokes parameters of two sharp specular balls under uniform spherical illumination (emitted from a LED sphere with 346 lights) showing circular polarization (red) between the observed specular highlights which are linearly polarized.

- Furthermore, if the content of the environment lighting varies slowly in comparison to the sharpness of the specular reflection (which is the case for overcast sky), then we can approximate the intensity of incident lighting over the solid angle of significant specular response as a constant scale factor $s_\omega$. Hence, for every surface point, the specular response is similar (up to a scale factor $s_\omega$) as if it was captured under a constant lighting condition. However, care has to be taken when performing the first-order correction of $\sqrt{s_1'^2 + s_2'^2}$ and $s_3'$ since different surface points' specular reflections are possibly scaled by a different scale factor $s_\omega$. We propose to either capture the exemplar in the same environment (and hence it includes $s_\omega$), or alternatively perform a first-order correction on the ratio $s_3'/\sqrt{s_1'^2 + s_2'^2}$ instead of the individual components (effectively dividing out $s_\omega$). The latter is similar to the ratio used in Eq. (3.2.9).

Again, we can readily apply the theory outlined for the uniform unpolarized incident lighting case to compute surface normals under uncontrolled overcast illumination.

The inclusion of circular polarization yields a more robust estimation of surface normals compared to prior work such as [27, 28, 25, 26, 29] which rely solely on linear polarization cues to estimate the in-plane incident angle from the *reduced* degree of polarization: $DOP' = \sqrt{s_1^2 + s_2^2}/s_0$ (i.e., $s_3$ is implicitly assumed to be zero). We found the estimation error to reduce by $5^\circ$ close to normal incidence with the inclusion of circular polarization.




(a) Geometric normals (b) Ma et al. [30]

(c) Circ. Pol. Normals (d) Unpol. Normals

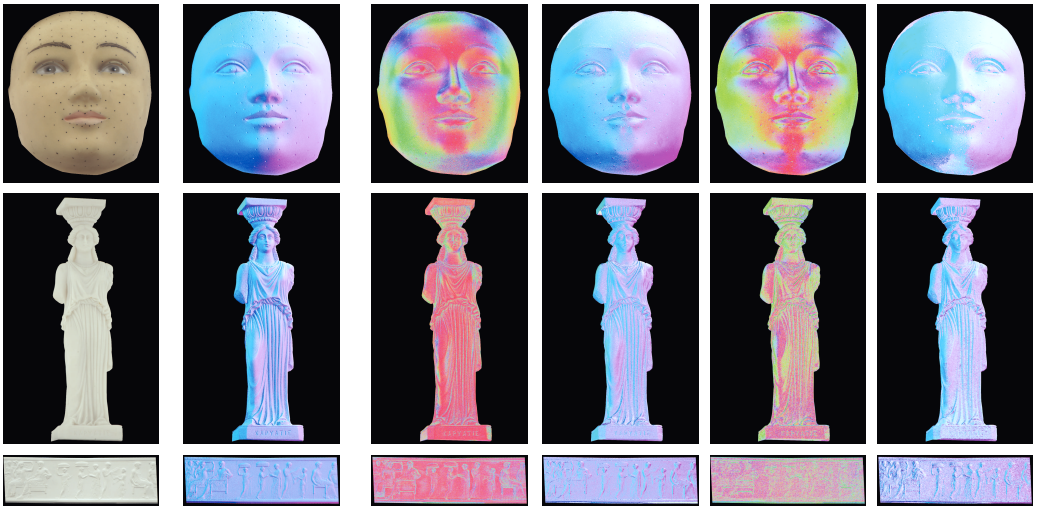
(e) 1D error plot

Figure 3.5: Surface normals of a spherical ball estimated from Stokes parameters of incident spherical illumination. Surface normals inferred from circularly polarized illumination (c), and unpolarized illumination (d), compared to known ground truth geometric normals (a), and compared to normals obtained from linearly polarized incident lighting using the method of [30] (b).

## 3.3 Results

We validate our theory in both controlled and uncontrolled lighting conditions. Fig. 3.1 shows surface normal estimation results for a convex plastic orange (top row) under controlled incident lighting generated by an LED sphere with 346 lights. We compare the estimated surface normals under circularly polarized and unpolarized spherically uniform incident lighting to surface normals obtained by the photometric method of Ma et al. [30]. The $\phi$ ambiguity is handled by growing the normals in from the silhouette. Fig. 3.1 (bottom row) presents the results under similar conditions of a more complex marble statue with concavities. The $\phi$-ambiguity is solved by capturing an additional lighting conditions: $I(\phi,\theta) = (\sin(\phi)+1)/2$.

Fig. 3.5 gives a quantitative error analysis of the surface normal estimation for an object with known shape (i.e., sphere). As can be seen the quality of the estimated normals is good, except close towards extreme angles due to reflection occlusion. Furthermore, the surface normals estimated under circular incident lighting exhibit a better SNR for front facing surfaces compared to those acquired under unpolarized incident lighting. The mean angular error is around $7^\circ$ for incident angles less than $75^\circ$. Note that normals towards the bottom of the sphere



(a) Subject (b) Photometric Normals [30] (c) Circular Stokes Param. (d) Circ. Pol. Normals (e) Unpol. Stokes Param. (f) Unpol. Normals

Figure 3.6: Estimated surface normals from Stokes parameters under idealized simulated outdoor lighting conditions. Surface normals inferred from circularly polarized illumination (c-d), and unpolarized illumination (e-f), compared to photometric normals [30] (b). Top-row: Plastic maquette - surface normal map of face grown inward from the silhouette. Center-row: Marble statue - $\phi$ ambiguity resolved using directional cues from the incident illumination. Bottow-row: Plaster bas relief.

exhibit a larger error due to reflection occlusion from the stand supporting the ball.

Fig. 3.6 shows normal estimation results under a spherical linear intensity gradient in the top-down direction: $I(\phi,\theta) = (\sin(\phi)+1)/2$, which simulates an idealized outdoor overcast condition. Fig. 3.6 (top row) show the estimated normals of a plastic maquette under this simulated lighting condition. The $\phi$-ambiguity was handled by growing in the normals from the silhouette. Fig. 3.6 center and bottom row, show a marble statue and a plaster bas relief respectively captured under the same single simulated lighting condition. For these two cases, the $\phi$-ambiguity was solved using the intensity information of the incident lighting condition.

Fig. 3.7 shows results of surface normal estimation from outdoor illumination on a cloudy day for the convex plastic orange as well as a jade Confucius statue with several concavities. The exemplar sphere was captured under the same lighting condition.

(a) Subject (b) Unpol. Stokes Param (c) Unpol. Normals

Figure 3.7: Estimated surface normals from Stokes parameters of diffuse outdoor illumination. Top-row: Plastic orange. Bottom-row: Confucius statue.

## 3.4 Discussion

In this chapter we presented a novel technique for estimating surface normals from polarization cues obtained from the Stokes reflectance field captured in just four photographs from a single viewpoint and under a single (unpolarized or circu-

larly polarized) constant spherical incident lighting. While the theory is based on Meuller calculus and is strictly for homogeneous dielectric materials with a constant specular roughness, we found it to work well in practice for both dielectrics as well as dielectric-metal composites. We validated the proposed technique in both controlled as well as uncontrolled outdoor lighting conditions.
A key empirical observation is that unpolarized light becomes slightly circularly polarized. Additionally, because the proposed technique only requires four photographs of a scene from a fixed viewpoint and under a single lighting condition, the proposed technique could potentially enable a single shot normal estimation device based on the assorted pixel [65] concept.

At the current state of our research we are able to capture also live subjects (e.g. a human face), although a typical acquisition session lasts a few seconds, because fo the need to flip the polarizer in front of the camera, plus the time required to capture the pictures required by [30] to infer accurate surface normals, in case we need a ground truth for comparison. Even if the subjects are generally able to stay still during the acquisition session in a laboratory setting, we detected a small amount of motion for our subjects, which can bias the normals estimate. We found safer to include a few extra frames, lit by a uniform spherical incident lighting, used as reference to run the joint optical flow technique described in [66]. In Figure 3.8 the estimated normals for a human face are shown; in particular the subject changed slightly her pose during the session, introducing motion around all three axis. As a consequence blurring and other artifacts are introduced, in particular on the unpolarized normals because of their lower signal to noise ratio.

Our method is closely related to [54] and they can be effectively combined together in case of circularly polarized incident lighting, as mentioned in Section 3.2.3. In this case it is possible to obtain a full description of the BRDF of the object, for each pixel, in just four observations: specular albedo, diffuse albedo, index of refraction, specular roughness and surface normals.

The surface normal estimation can also suffer from bias due to any reflection occlusion. We also restrict the analysis to spherically symmetric lighting conditions with respect to polarization, i.e., circularly polarized and unpolarized incident lighting. Linearly polarized incident lighting typically results in view-dependent Stokes reflectance fields as well as very little variation in the measured

(a) Female subject (b) Photometric Normals (c) Circ. Polarized Normals (d) Unpolarized Normals

Figure 3.8: Estimating surface normals (encoded as $|\mathbf{x}| \rightarrow R$, $|\mathbf{x}| \rightarrow G$, and $|\mathbf{x}| \rightarrow G$ to better focus on motion artifacts introduced by a slight change in the pose of the subject during the acquisition) . Surface normals inferred from circularly polarized illumination (b), and unpolarized illumination (c), compared to surface normals obtained from a photometric stereo variant [30](a).

Stokes parameters for robust surface normal estimation. The method inherently relies on the incident polarization state of light to be uniform from all directions. This assumption is violated under linearly polarized illumination or strongly reflected bounce light from the floor or walls for uncontrolled indoor or outdoor measurements.

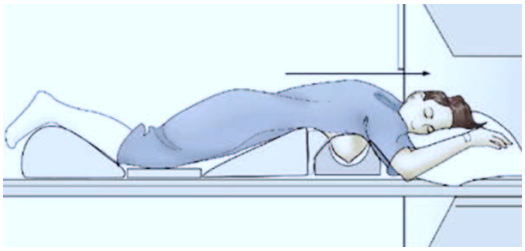
# Chapter 4

# Statistical 3D Shape Modeling

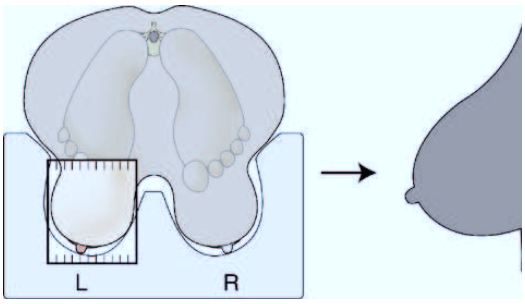
## 4.1 Introduction

The female breast is a complex three-dimensional shape, with fuzzily defined boundaries and only a few easily identiable anatomical landmarks. Evaluating the post-operative outcome of breast plastic surgery is currently a little more than craftsmanship and intuition, strongly relying on the professional skills and personal experience of plastic surgeons, which cannot be embedded into a standard clinical procedure. Visual assessment is the common practice among surgeons, since no universally accepted breast shape analysis technique is currently available, and clearly is barely reproducible. Devising reliable and objective methodologies to objectively analyse natural and reconstructed breast shape, to complement the direct inspection of the surgeon both in the pre-operatory and postoperatory phase, is an important research issue in plastic and reconstructive surgery. The techniques should be the least invasive and at reasonable cost.

Our proposal makes use of Nuclear Magnetic Resonances Imaging (MRI). This is nowadays a standard radiological test that is performed in most of the hospitals, which provides good information about internal breast structures. It also provides, at practically no cost, precise numerical measurement of the 3D external surface of a womans breast. These data are especially valuable if the patient is carefully put in a standard prone position during the MRI acquisition (Fig. 4.1).



(a)




(b)

Figure 4.1: The patient is lying prone and she must have comfortable pillow for head and arms. Care as to be taken in centering the breast in the coil. Patient positioning as seen from the side (a) and from behind (b), together with the idealized shape of a sagittal section of the breast. Courtesy of: Cornell MRI Laboratory Group (http://weill.cornell.edu/mri/).

For this purpose a collection of MRI data relative to about 40 patients have been acquired with a homogeneous clinical protocol. Using this data set we have the opportunity to explore the variability space of the human breasts and to try to isolate few numerical parameters able to describe the shape variations observed in the women population.

We have pre-processed the collected MRI data to obtain geometrical models of the breasts eliminating from them most of the experimental noise. These standardized surfaces have been in turn processed with the Principal Component Analysis to obtain an average shape and a small set of principal orthogonal modes that are able to explain and model most of the observed variation in the data.
At this stage we are already able to present an interesting way to objectively evaluate a patients breast within a span between some extremal cases.
The description of the human shape by mean of a set of numerical parameters

has a long history. Perhaps the first to propose such a way to describe the human body are the Renaissance artists (Leonardo da Vinci, Albrecht Durer). Efforts to systematically measure body shape parameters have been carried out in military environments [67]. These efforts have been perfected by the availability of recent laser body scanning techniques. CAESAR is a joint European and American funded effort that has brought to the constitution of publicly available, massive database of the whole human body shapes [68]. Since the conclusion of this project, studies about anthropometric measurements based on these public data have become abundant in the published literature. Principal Component Analysis has been used in order to embed the space of human body shapes in some low dimensional geometric variety [69].

As for specific organs of the human body considerably less work have been done and published. Regarding human breasts few experimental approaches have been published both from the point of view of industrial and clothing applications [70] and of medical studies [71]. For breast evaluation, laser scanning techniques are not yet sufficiently robust: typically scanning time is too long and patients breathing interferes too much with the quality of the final data. Moreover areas of the female torso may remain occluded to the optical laser ray [72]. As for the use of a parametric model to describe the shape of the human breast a seminal paper using a super quadric approach is [73]. In this chapter we follow the approach of [69] applying the principal modes obtained with PCA to the problem of describing the breast. Our work has been published in [74, 75].

## 4.2 Proposed Methodology

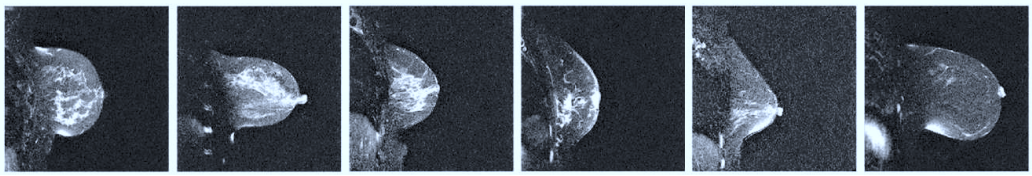
### 4.2.1 The Dataset

To collect our dataset we acquired 46 MRI of womens breasts, with a homogeneous clinical protocol. The volunteers varied in age from 21 to 76 years, in order to take into account the effect of natural aging on the breast shape. In all the resonances the patient was lying prone and left the breasts free to hung down influenced only by gravity within the instrument (Fig. 4.1).

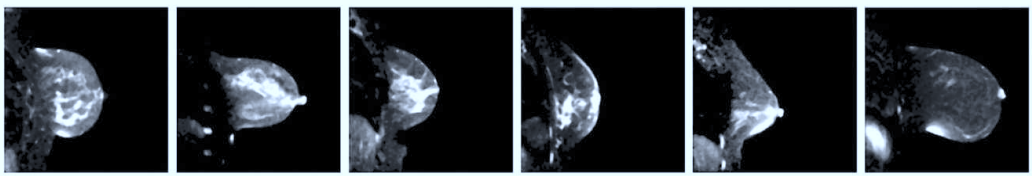
Both right and left breasts images were acquired in this way. The volunteers

varied in age from 21 to 76 years. The majority of the cases are relative to healthy women, but some pathological typical cases have been also included in the study. Care has been taken not to include extremely aberrated or incomplete shapes in our computations, to avoid the introduction of bias in our model. The whole volume of the resonance for each patient is made of 100 slices (50 slices for each breast).

The rough data present heavy noise, due to the MRI scanner and the imaging process, and need to be cleaned and registered in a unique reference frame for further processing. Figure 4.2 shows some example of noisy MRI slices and the corresponding filtered output. To obtain acceptably smooth surfaces we apply the processing pipeline described in the following section.



(a)



(b)

Figure 4.2: Some example of noisy MRI slices (a) and the corresponding filtered output (b).

### 4.2.2 Surface Smoothing with Polynomial Fitting

Noise reduction at each MRI slice is the very first and preliminary processing that has to be performed on the data. The hypothesis of additive white Gaussian noise, with zero mean and variance $\sigma_L^2$ at each slice $L$ is assumed. This allows to separately process each slice as an independent picture. To statistically evaluate the noise variance $\underline{\sigma_L^2}$ in a slice we sample a reasonably large region $R$ where, with high probability, there is no tissue. A natural candidate for $R$ is the corner of the

slice opposite to the breast. The knowledge of $\underline{\sigma_L}$ allows to precisely tune a rotationally symmetric Gaussian lowpass filter of size $L$. This first smoothing still leaves some amount of salt and pepper disturbances. A median filtering is used to reduce this kind of noise without affecting edges and hence without perturbing the profile of the breast/air interface (Figure 4.3(c)).

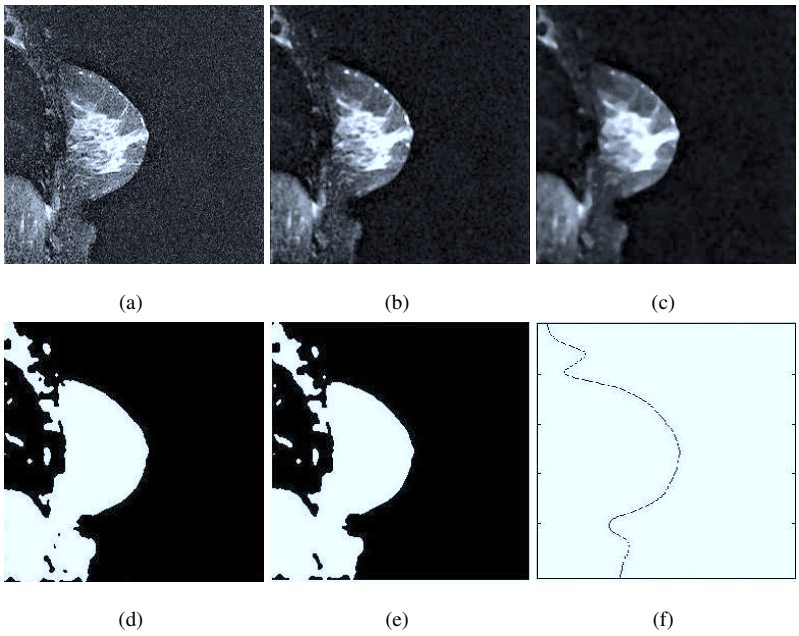



Figure 4.3: Steps of surface fitting. a) input slice; b) output of gaussian filter; c) output of median filter; d) thresholding; e) morphological closing; f)robust local regression.

The precise identification of the breast/air interface is subsequently performed with a binarization procedure. An adaptive threshold for the binarization is found on each slice separately: the threshold value is determined looking at the histogram of the pixel values in the region $R$ considered above. $R$ is relative only to the air and ideally it should appear totally black. For this reason the natural choice for the threshold value is the maximum observed non zero value in $R$. The resulting binarized images still may present isolated dark areas within the tissue region and isolate bright spikes in the air region. These artifacts are appropriately removed with standard filtering (Figure 4.3(d)).

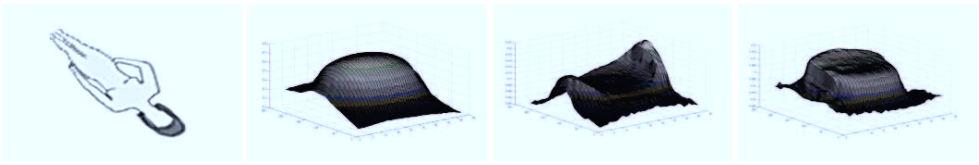
Figure 4.4: Three views of the 3-D reconstruction using Marching Cubes algorithm of a breast in our dataset. In all views it is evident the degree of ptosis which affects the patient.

To naively follow the border between black and white areas in the slice at this stage would produce a very jagged contour while a more regular curve is desirable. Regularization is achieved first applying some morphological operators and hence fitting a polynomial curve. More precisely a morphological binary dilation, followed by a morphological erosion (morphological closing) [76], with a $3 \times 3$ pixels square as a structuring element is performed. Eventually a local robust regression using weighted linear least squares and a second degree polynomial model is used to further regularize the curve. To ensure robustness the regression weights are assigned in such a way that probable outliers gets a lower weight and zero weight is assigned to data outside six mean absolute deviations.

The curves resulting from the application of this procedure on each slice are finally assembled together in a surface model. Figure 4.4 shows a 3-D reconstruction of the breast surface using Marching Cubes [77]. It is possible to obtain a smoother 3-D surface by means of bicubic interpolation.

### 4.2.3 Principal Component Analysis of Breast Shapes

Out of the complete data set we have set apart, for further testing, ten randomly chosen breast shapes. All the other smooth surface data (36 breast pairs) have been processed using a standard implementation of Principal Component Analysis in



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

Figure 4.5: The average shape (a) and the first three components (b)-(d) in our dataset.

MATLAB. The average shape in our dataset and the first three components are reported in Fig. 4.5

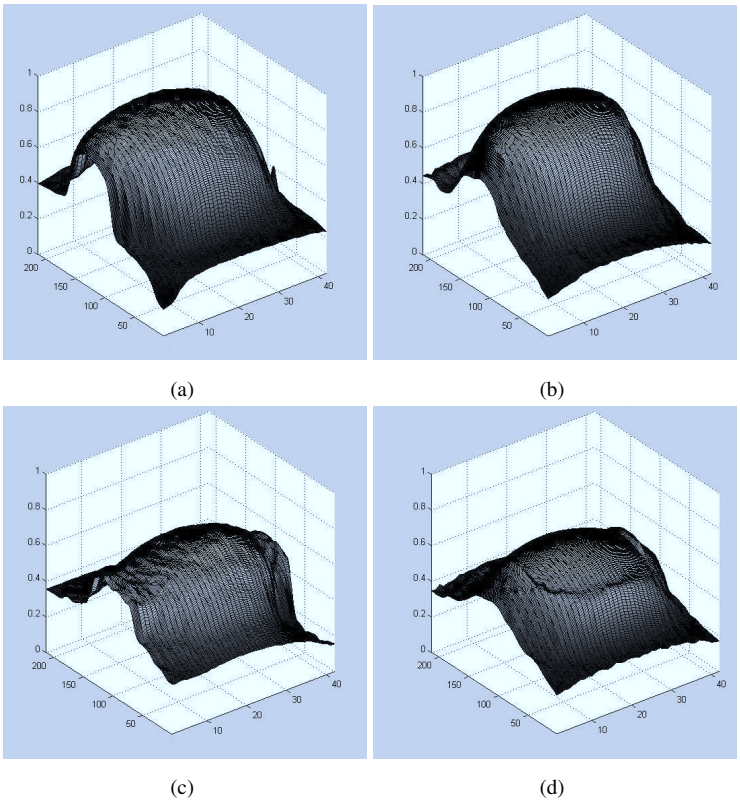
Table 4.1 reports the percentage relevance of the six eigenvalues associated with the first six components. In particular the first three components explain about 82% of the total data variation. PCA theory guarantees that the approximation error introduced leaving the less significative modes out of the reconstruction is mathematically bounded.

| Eigenvalue $\lambda_i$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ | $\lambda_6$ |
|---|---|---|---|---|---|---|
| Contribution % | 61.6 | 15 | 5.27 | 3.77 | 2.53 | 2.39 |

Table 4.1: Eigenvalues of the covariance matrix.

To experimentally confirm that the first six components are sufficient to produce a fairly good approximation of the real data even in the case of extremely deformed shapes, we report the results of the reconstruction of some real data using only the first six modes both in the case of normal breast than in the case of a severely deformed one. The reconstructed models are shown in Figure 4.6. The reader may visually appreciate the faithfulness of both the reconstructions.

The knowledge of the average shape of the breast together with the principal modes allows the synthesis of new models. To enable the physician to interactively explore the patient data and the potentiality of the proposed modelling technique we have implemented a graphical JAVA application. The program makes use of the *JAVA3D* library to interactively show a breast model. The model can be moved in $3-D$ with mouse gestures. A set of six sliders provides the user with the possibility to change the contribution to the final shape of each of the main six modes, and the effect of changes in weights is immediately visible. A screenshot



(a) (b)

(c) (d)

Figure 4.6: Results of the reconstruction of some real data. a) a normal breast and b) its reconstruction using the first six modes; c) a deformed breast and d) its reconstruction.

of the application is shown in Figure 4.7.

## 4.3 Results

We have tested the proposed methodology in two ways: reconstruction and guidance to qualitative analysis. As we mentioned above to test the expressive power of the proposed shape descriptors we set apart as control set a small number of randomly chosen MRIs (10) for the modes extraction phase and we reconstructed these left out breasts from the computed orthogonal modes. The control set included cases of fairly standard breasts as well as cases of great deviance from the norm. In all cases the reconstruction appeared to be visually satisfactory although a degree of approximation has been introduced. The mean error distribution is


Figure 4.7: The graphical user interface of our JAVA application, to interactively explore patient data.

shown in Figure 4.8. Observe that most of the error is localized only in peripheral areas of the breast and in the inframammary fold.

A relevant issue is if each mode may be interpreted as a morphological macroscopic feature with some clinical meaning. Although a precise mapping of the proposed modes to common properties like volume, roundness, concavity, etc is unlikely, at least for the first two modes some correlation can be made. In particular the first mode (whose associate eigenvalue weights 62%) induces mostly volume variations (Figure 4.9a). As for the second mode (whose associate eigenvalue weights 15%) it induces mostly variations in the degree of protrusion of the breast in front of the sternum (Figure 4.9b). These observation suggested the construction of a linear scale to qualitatively evaluate breast shapes. Please note that the proposed scale is just exemplary, but we believe that our approach may be helpful in the direction of the construction of a well balanced and universal scale.

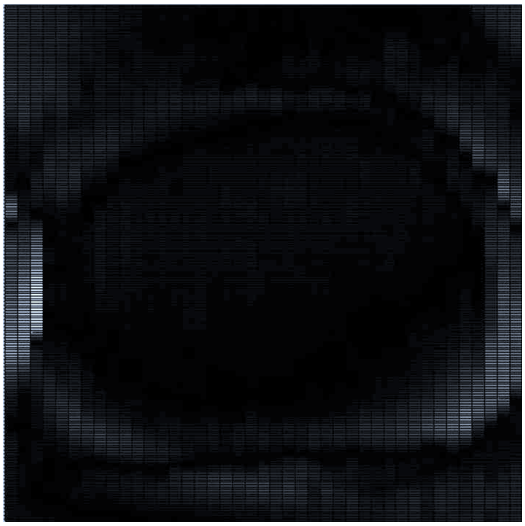
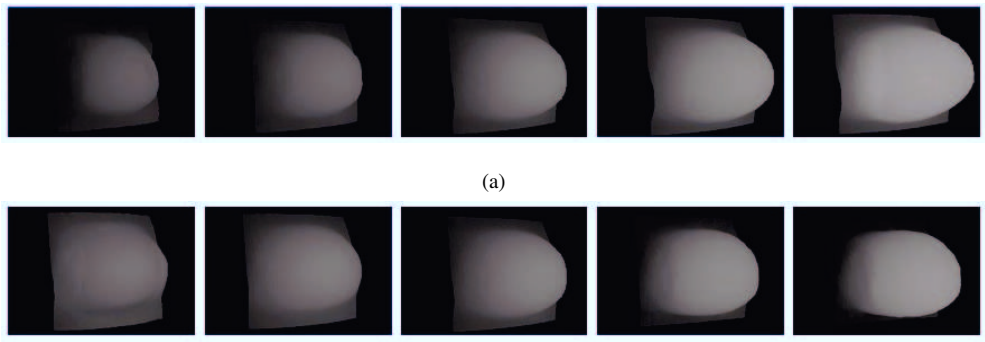


Figure 4.8: Mean reconstruction error distribution.



(a)

(b)

Figure 4.9: (a) Variations induced by the first parameter; (b) Variations induced by the second parameter.

Keeping in mind the previous caveat, we selected two extremal breast shapes in our database. In particular the doctors chose as a reference the case of a patient whose photograph is shown in Figure 4.10. The patient is a 48 aged woman whose left breast has been surgically reconstructed after the resection of a tumor, while the right breast has been only subject to normal aging. These left and right breasts constitute in a way two extremal cases and it is, in a first hypothesis, reasonable to believe that all the breast shapes may span between these two poles. To check

this idea we computed the "distance" of the other breast shapes from the two extremal cases. More precisely, the "distance" from shape $A$ to shape $B$ is defined, in this context, as the Euclidean distance in $\mathbb{R}^d$ between the normalized first $d$ coefficients in the PCA expansions of $A$ and $B$. We experimentally found visually clearer results for $d = 2$. Figure 4.11 shows a plot of the breast shapes in a $X - Y$ plane. The $X$ coordinate of each data point represents the distance of the breast from the left breast of Figure 4.10; the $Y$ coordinate of each data point represents the distance of the breast from the right breast of Figure 4.10.

Figure 4.10: Reference case: woman whose left breast has been surgically reconstructed after the resection of a tumor, while the right breast has been only subject to normal aging.

We partitioned the span between the two extremal cases into 5 parallel strips ($S1 - S5$). For each not empty strip we choose a central data point. Figure 4.12 shows the central sections of the 4 selected shapes between the central sections of the two reference breasts. As it is evident the progression between the two extremal cases appear to be gradual and of immediate clinical meaning, demonstrating that the proposed parametrization may be used to develop a reliable, objective technique to evaluate the shape of a breast, whether it has been surgically reconstructed or not.

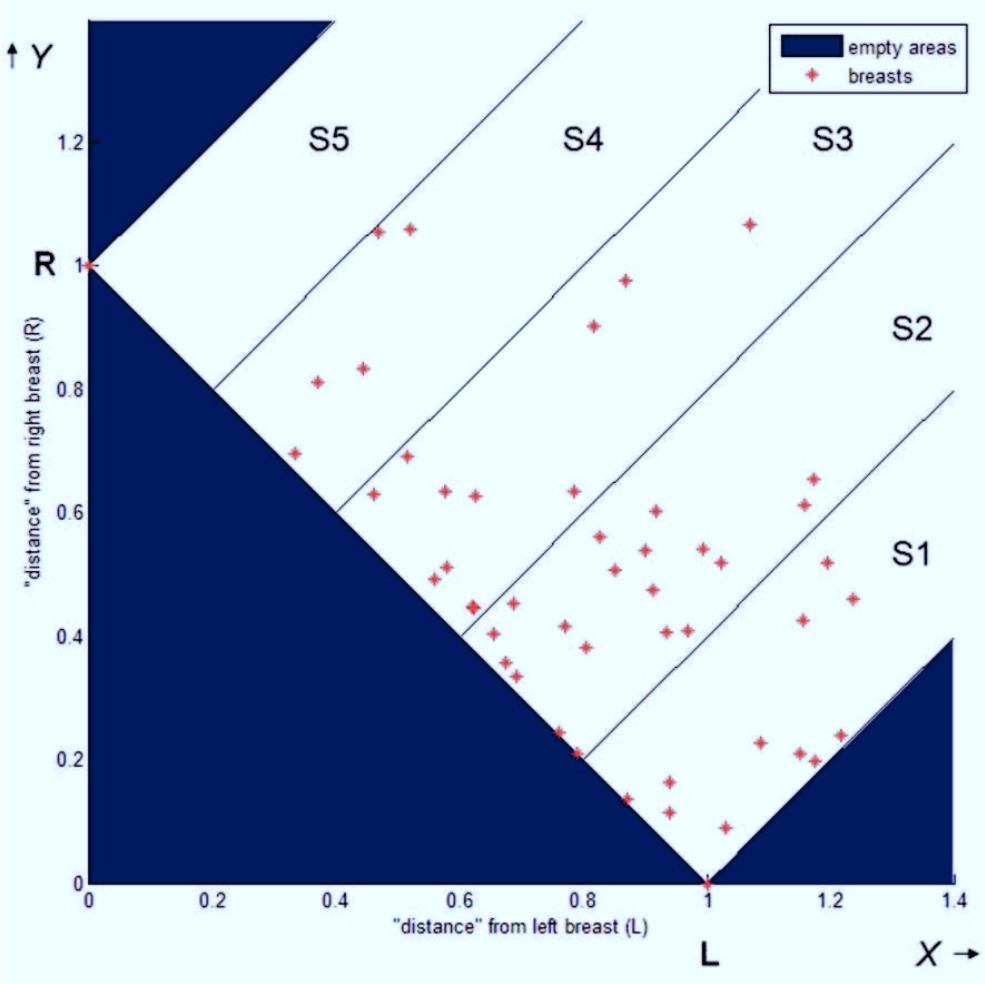
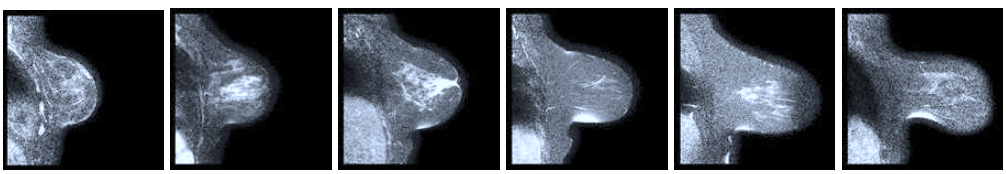



Figure 4.11: The $X$ coordinate of each data point represents the distance of the breast from the left breast ($L$) of Fig. 4.10; the $Y$ coordinate of each data point represents the distance of the breast from the right breast ($R$) of Fig.4.10.



$Left breast(L)$ $S1$ $S2$ $S3$ $S4$ $Right breast(R)$

Figure 4.12: Central sections of the 4 selected shapes between the central sections of the two reference breasts; strip S5 is empty.

## 4.4 Discussion

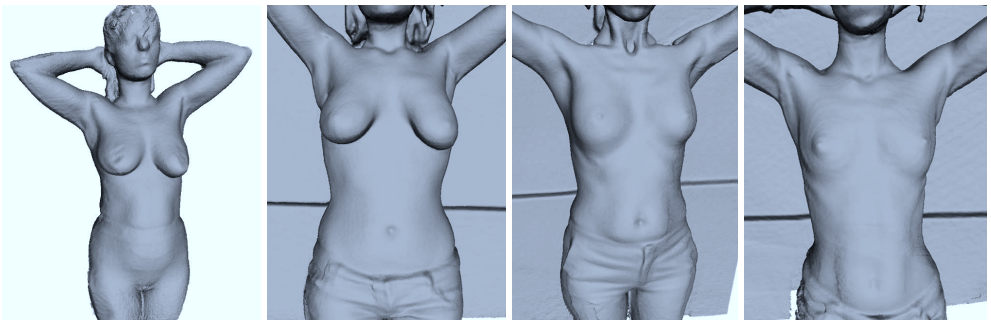
In this paper we have presented an embedding of the shape space of the human female breast into a low dimensional linear parameter space. The proposed parametrization has been experimentally obtained from a set of purposely collected and properly processed MRI data. The data have been processed for noise removal and analyzed with the PCA technique. A first medical assessment of the model, done using a 3-D software especially developed for this application, proved that this technique may be of clinical relevance and a tentative qualitative scale for breast evaluation has been proposed. Although the proposed technique has proven being useful, easy to apply and at a reasonable cost, recent advance in technology made available a number of low cost devices which might be taken into account to complement or even replace MRI technique. The reasons which encourage further research are mainly three:

- MRI requires well trained personnel, not only for using the scanner itself but even for a careful patient positioning;
- the exam takes time and the patient has to stay as still as possible, otherwise breathing and motion artifact will be introduced;
- the resolution of each slice is adequate along the sagittal and coronal axis. However the number of slices is limited to fifty for each breast, and each slice is relative to three millimeter of thickness along the transverse axis, hence the resolution is quite limited. Figure 4.4 makes clear this limitation.

Kinect [78] is a new and widely-available commodity sensor platform that incorporates a structured light based depth sensor, that generates a eleven bit depth map at $30Hz$, with a resolution of $640 \times 480$. Taking into account the cost of the device, the quality of this depth map is generally remarkable, but the depth images contain numerous holes" where no structured light depth reading was possible. This phenomena is due to the properties of certain materials or scene structures which do not reect infra-red (IR) light, very thin structures or surfaces at glancing incidence angles; motion blurring can also lead to missing data. Newcombe et al. [79] recently developed a system for accurate real-time mapping of complex

and arbitrary indoor scenes in variable lighting conditions, using a Kinect sensor and commodity graphics hardware. All of the depth data streamed from a Kinect sensor are fused into a single global implicit surface model of the observed scene in real-time. The current sensor pose is simultaneously obtained by tracking the live depth frame relative to the global model using a coarse-to-ne iterative closest point algorithm, which uses all of the observed depth data available. This method allows dense surfaces to be reconstructed in real-time, with a good level of detail and robustness. A similar approach is implemented in [80].

We used used the Kinect along with the *ReconstructMe* 3-D scanning software to capture a new dataset and to build a prototypal system for breast shape analysis. The acquisition protocol is way less constrained than the MRI protocol and the patient is required to stand still for only a few seconds, depending on the GPU hardware capabilities. Preliminary tests shows that the best tradeoff between acquisition time, patient comfort, quality and reproducibility of the acquired shapes can be obtained performing the pose depicted in figure 4.13(a). The patient is standing still, with her hands crossed behind the head and the elbows at the same height of her nose. The described pose allows to accurately capture enough depth information about the inframammary crease even in presence of a sever degree of ptosis.



(a) Patient pose (b) (c) (d)

Figure 4.13: 3-D female breast shapes acquired using a low cost depth-camera (Kinect); a) the patient is standing still, with her hands crossed behind the head and the elbows at the same height of her nose. (b-d) three patient acquired in the same pose.

At this stage we are able to show the quality of the acquired breast shapes, and some examples are reported in figure 4.13(b-d).
It is evident that the quality of the 3-D models is quite high, potentially enabling a more accurate shape analysis. Acquisition time is around 30 seconds during which our patients are generally able to stay still, but it can be reduced using high-performance graphics hardware. Future research to be done in this area will include refinement and clinical evaluation of a qualitative shape space and a thorough investigation of any issue which can arise from the absence of a rigid positioning protocol for the patients.

# Chapter 5

# Statistical 2D Shape Modeling

## 5.1 Introduction

In this chapter we demonstrate a pratical use of statistical 2-D shape models, applying a technique derived by [33] to the problem of evaluating overall health condition of animals. In particular, semi-automatic and automatic assessment of dairy cattle health is addressed.

The energy reserves in cows in terms of body fat stores and mobilization during the different lactation stages have important implications for milk production, animal well-being, reproductive performance, and, more generally, farm productivity. Body Condition Score (BCS) is widely considered an important tool for management of dairy cattle because it is a simple and repeatable system used to evaluate body fat stores and estimate cumulative energy balance through visual or tactile inspection [81]. The score range used by most dairy management advisors applies a scale from 1 to 5, with 1 representing emaciated cows and 5 representing obese cows. Despite the general consensus on the benefits of the BCS evaluation in farms, only a small percentage of US dairy farms have adopted this practice in the production chain. The main reasons that discourage the use of the traditional BCS evaluation techniques are the lack of computerized reports [82], the subjectivity in the judgment that can lead to different scores for the same cow under consideration, and the complex, not immediate, and time consuming on-farm training of technicians. Furthermore, the measurements must be revised frequently on each

cow augmenting hence the costs for the farms.

The feasibility of estimating the BCS from digital images has been demonstrated in recent works. Ferguson et al. [83] assess the ability to assign a BCS to a dairy cow, by a human observer, directly from digital photographs or a video taken from the rear of a cow at a 0 to 20 degree angle relative to the tail head. In [84] a semi-automatic BCS estimation technique, from a single digital image, is presented. A dorsal view is captured automatically as cows pass through a weigh station and then 23 anatomical points are considered, to define the shape of the body of the cow. These points, selected in a manual way, are used to compute 15 angles around the hooks, pins, and tailhead, in order to describe the cow's contour. A regression machine is employed to infer the BCS from the computed angles. Halachmi et al. [85] test the hypothesis that the body shape of a fat cow is rounder than the shape of a thin cow and, therefore, may better fit a parabolic shape. The posterior part of the cow is considered in performing the parabolic fitting. Images are acquired by means of a thermal camera that allows a very simple and straightforward shape extraction. The BCS estimation is achieved by considering the absolute differences between the real body shape and the fitted parabola.

Despite the progress in this research area, such studies have not addressed the problem of modeling the shape of body cows to build a robust descriptor for automatic BCS estimation. Among the visual cues used by human visual system, the shape provides important information that allows humans to distinguish between objects of different categories [38] as well as information that are relevant to understand the differences in the appearance of an object within a specific class [33]. In computer vision literature, several shape descriptors have been proposed [38, 33, 86, 87]. More specifically, shape descriptors based on Principal Component Analysis (PCA) [33, 86] are used to consider the different variability of anatomical landmarks with respect to the average shape.

The aim of our study is to model the shape cows, as seen from above, by capturing variability with respect to a "prototype" shape properly derived by a set of examples and then exploiting these variabilities to describe the involved shapes in a reconstructive way. BCS estimation is performed after learning a

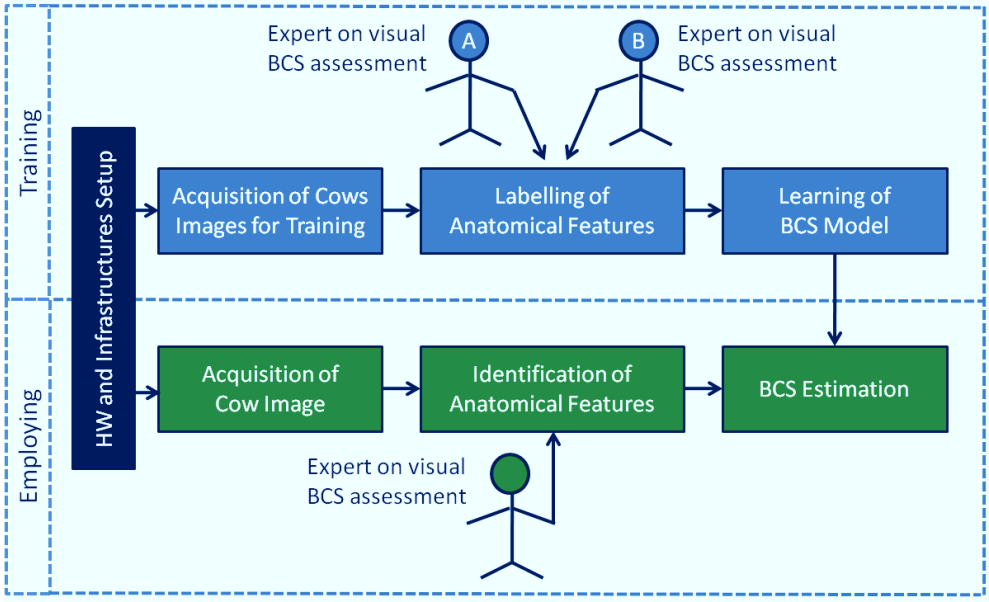



Figure 5.1: The scheme of a system for semiautomatic estimation of BCS.

regressor on a kernel PCA space of cows shapes. A further objective was to build a benchmark dataset useful for dairy cattle research purposes, available through the Internet[1]. Experimental results confirm the effectiveness of the proposed approach that outperforms all previous approaches in terms of BCS evaluation accuracy.

## 5.2 Proposed methodology

### 5.2.1 System overview

The first step of our research is semi-automatic evaluation of the BCS from digital images and it has been published in [88, 89, 90]; a general scheme such a system is shown in Figure 5.1. The system consists of two different blocks: *Training Block* and *Employing Block*. The *Training Block* is used to learn the parameters of the model exploited to infer the BCS from features extracted on digital images. The parameters are learned by using a set of labeled examples. Once the training is completed, learned parameters are used in the model to infer the BCS of new samples during the employing phase. Both *Training Block* and *Employing Block*

[1]The BCS Database is available at: http://iplab.dmi.unict.it/bcs/

make use of the same hardware infrastructure (e.g., digital camera parameters and position).
Each block consists in different modules organized in a sequential pipeline.
The *Training Block* is composed of three modules as follows: acquisition of training examples, labeling of anatomical features, and learning the BCS model parameters.

- The *Acquisition Module* is used to acquire images to be used as samples in learning the model parameters. The sample cow images to use in the *Acquisition Module* should include the range of the number and variety of samples to be acquired, since this set of acquired images should be representative of all possible BCS values. During the acquisition phase, technicians should evaluate BCS of the involved cows on site, building a consistent labeled dataset which contains images together with the corresponding BCS. In order to build a system capable to infer the BCS from anatomical features of cows (e.g., hook angles), the features must be first labeled by experts and then used to learn the parameters of the chosen BCS model.

- In the *Labeling Module* an expert uses a software tool (which will be described in Section 5.2.4) to mark the anatomical features of interest on the acquired digital images. Labeled points are related to the features that should be taken into account to assess BCS. Several experts and technicians (at least two) should be used independently in the labeling phase to guarantee that the final labeled dataset is not biased by subjectivity of just one technician.

- The *Learning module* is devoted to establishing the set of parameters involved in the BCS model (e.g., regressor on anatomical points) from the labeled dataset. The *Training Block* uses the *Learning and Acquisition Modules* to construct the parameters of a model to estimate the BCS of a cow from the images labeled by the expert(s).

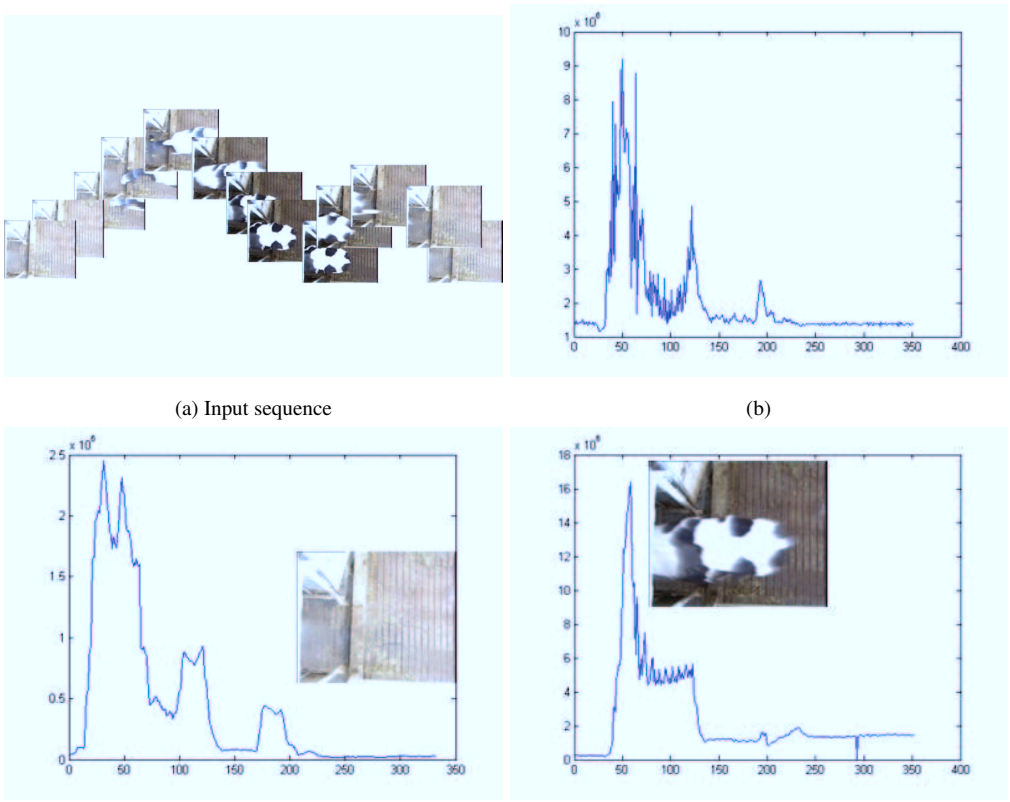
The *Employing Block* is composed of three sequential modules: acquisition of a new unlabeled image of a cow, identification and labeling of anatomical features

of the image, and application of the model generated in the learning phase to assign a BCS to the sample.

**Barn description and image acquisition.** Images are acquired by means of a standard network digital camera, with pan and tilt functions. This camera can be monitored from the local area network or from the Internet and can switch between color images during daytime, and black/white images in low light conditions (or nighttime) using the built-in infrared lighting. The camera also offers an *Application Programming Interface (API)* for software integration and its resolution is $704 \times 480$ pixels at up to 30 frames per second for $NTSC$ system and $768 \times 576$ at up to 25 frames per second for *PAL* system. The camera is positioned 3 meters above the floor, to capture images dorsally as cows pass through an exit gate. The choice of this specific gate, accessible from a pair of milking robots, guarantee that all the lactating cows present in that group are examined at least once a day. Cows are constrained briefly at the exit gate for a few seconds which allowed image capture from a relatively stationary cow. Images are acquired at an average frame-rate of $12 fps$ from *10 a.m.* to *2 p.m.* and from *8 p.m.* to *00 a.m.*, in order to take into account two different illumination conditions. All images are saved locally and then processed offline. Cows are not restricted based on coat color (cows had variable combinations of white to black markings), coat condition (sleekness of hair coat), or age and size for sample collection.

### 5.2.2 Image selection

During each sample collection period, the image acquisition system gathers a huge amount of data (approximately 172800 images for each acquisition interval). For each acquisition period only a few images actually contains a cow in the frame (about 40 images). The selection of the images containing cows is a critical issue that cannot be addressed manually. To overcome this problem, a three step algorithm was developed to select only those images that contains a cow image for analysis. First, a filtering procedure is used, employing absolute interframe error or frame difference, which is the analysis of the difference between corresponding pixels in adjacent frames. When a cow passes through the gate the absolute interframe error typically has higher values compared to the absolute interframe error



(a) Input sequence (b)

(c) (d)

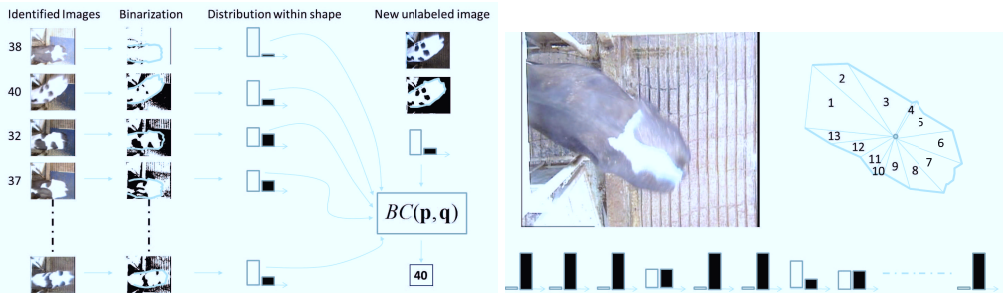
Figure 5.2: Absolute interframe error (b) relative to the series of selected frames (a), local variance analysis (c), and differences between the identified background frame and the other selected frames (d). The peak indicates the image that differs more with respect to the background (i.e., the image that contains the whole cow). The plateau in (c) is strongly related to consecutive background frames.

comparing frames without a cow (background images). For each acquisition period of four hours, a peak and valley plot is obtained. Each peak corresponds to an image containing a cow whereas valleys correspond to background images. After identifying the highest peak images, an automatic process selects a fixed number of images around the peak (200 in our implementation). The image corresponding to the peak might not be the best image to be used for BCS estimation (e.g., only part of the cow could be visible). Optimal images to use for BCS has to be chosen from the 200 selected images. The mean absolute interframe error (Figure 5.2 (b)) is employed as starting point for local variance analysis from the 200 im-

ages (i.e., the variance is computed considering a sliding window of 20 elements). A plateau is strongly related to consecutive background frames (Figure 5.2 (c)). A background frame, $Bg$, is selected from this uniform region and the difference between all the selected frames and $Bg$ is computed (Figure 5.2 (d)). The peak indicates the frame that differs the most with respect to $Bg$: the corresponding frame hence probably contains the whole cow. In order to cope with motion blur, out of focus, and other acquisition problems, five frames around the identified frame are selected. Finally the best frame is manually identified among the five frames. Human interaction is therefore required only at the end of the process and is related to a small subset of frames ($5 \times 40$ per 4 hours of acquisition). In our work, the aforementioned filtering process led to a final dataset with 286 images, relatives to 29 cows.

### 5.2.3 Cow identification and manual body condition scoring

At the beginning of acquisition step, two technicians are employed to identify cows at the exit alley of the milking robot. The clocks of both the image acquisition system and the technicians were synchronized. Technicians fill out a report with one record for each cow involved in the experiment. The report contains the cow identification (ID) from neck collar, BCS assessed by the observer (estimated according to [81]), and a timestamp. Once the report is completed, the assigned BCSs are properly associated to the acquired cows' images by using the timestamp. A semi-automatic procedure is adopted to assign ID label and BCS score (i.e., the ground truth) to all the other acquired images selected by the filtering pipeline. To establish similarity between images labeled by technicians and unlabeled images obtained with the filtering process, each image is first binarized by considering the average color within the shape region, and then represented as a binary distribution taking into account black and white pixels within the shape region (Figure 5.3 (a)). Similarity between the distributions related to different images is measured by using the Bhattacharyya coefficient [91]. Since the binarized texture of cows can be similar in their distribution if the shape is considered as a whole, we adopt a more robust representation of the content of images based on a binary distribution for each subregion of the shape obtained considering anatomi-




(a) Similarity between cows' images.

(b) Binary distribution in each subregion.

Figure 5.3: A semi-automatic procedure was adopted to assign ID label and BCS score (i.e., the ground truth) to all the other acquired images selected by the filtering pipeline.

cal landmarks and the center of mass of the anatomical points (Figure 5.3 (b)). The similarity between two images, represented by using binary distributions of subregions within the shape is obtained by averaging the similarity of distributions of corresponding subregions measured with the Bhattacharyya coefficient. Our software takes in input an image from the labeled dataset and retrieves the first $K$ similar unlabeled images ($K = 10$ in our experiment). A technician is employed to associate the retrieved unlabeled images to the target labeled image through visual inspection. The ID and BCS of the target image are automatically associated by the software to the images selected by the technician. This software is useful to speed up the labeling phase. This procedure led to a dataset of 29 images (one for each cow involved in the experiment, as mentioned at the end of 5.2.2) labeled with the mean BCS estimated by the two technicians.

### 5.2.4 Anatomical points labeling

An ad-hoc JAVA application was implemented to label the 23 anatomical points useful for together with the BCS dataset, subject to the GPL license. The graphical user interface is shown in Figure 5.4. The main window is subdivided into four rectangular areas; lower areas report a reference figure and a legend of the key points to locate; the upper regions present the image of a cow to mark and records the regions as they are identified. An image is automatically loaded from a user specified directory and presented in the box in the upper left corner (Figure

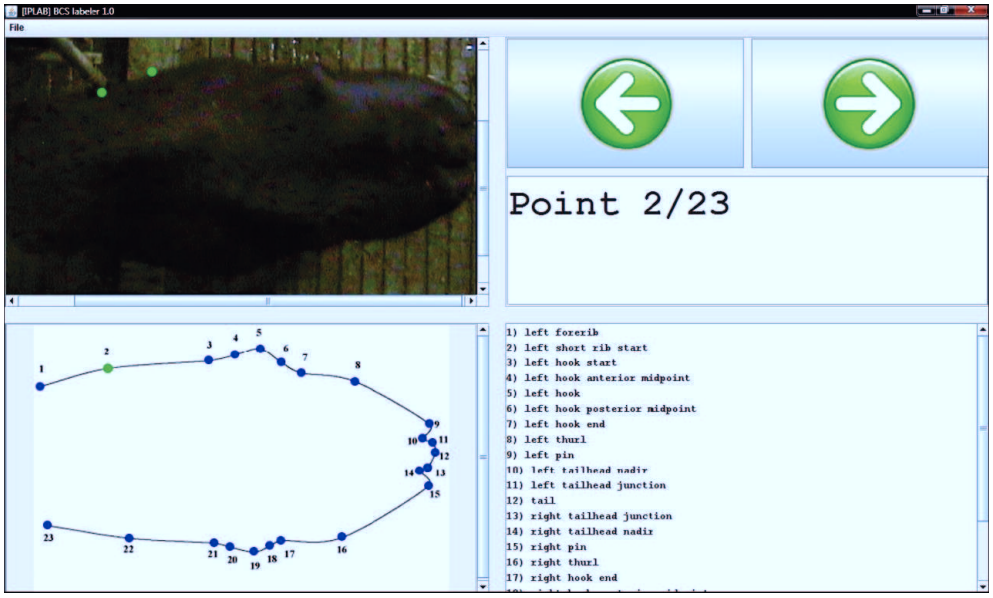


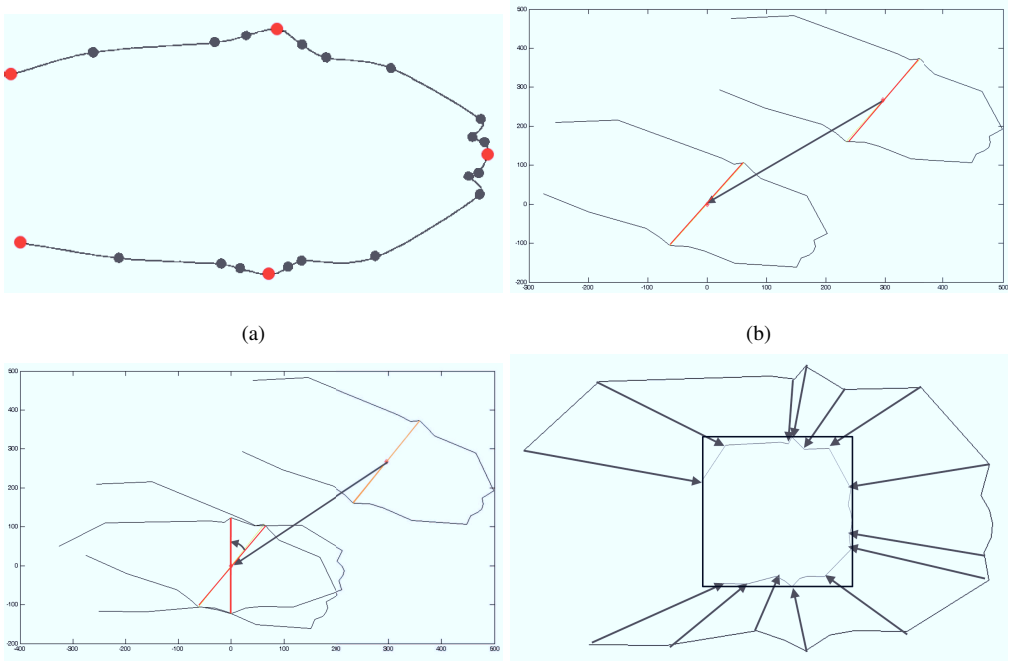

Figure 5.4: The JAVA interface of the labeling software used to mark the 23 anatomical points used for BCS estimation.

5.4). All 23 anatomical landmarks to be marked are listed in the legend in the order in which they have to be identified. To select a point the user right clicks on it with a mouse, to record its location. The selection of a point activates two buttons (upper right quadrant in figure 5.4): the first one deletes the point previously identified whereas the second one permits acceptance of the new point identified and enables the user to move forward (labeling errors can be immediately corrected). A counter in the top right box records the anatomical landmarks as they are identified by the user (Figure 5.4). Once the $23^{rd}$ anatomical point is confirmed, the software performs a consistency check of the selected points, identifying the most likely errors in their placement order (e.g. points from 1 to 9 should have an increasing abscissa).

## 5.3 Proposed methodology

### 5.3.1 Shape alignment

To obtain a consistent shape representation, location, scale and rotational effects need to be filtered out by aligning the corresponding anatomical landmarks of the different involved shapes. Shapes alignment is carried out by establishing a coordinate reference (position, scale, and rotation, commonly known as pose) to which all shapes are referred. The reference anatomical landmarks are the foreribs, the tail, and the right and left hook, as highlighted in Figure 5.5 (a).



(a) (b)

(c) (d)

Figure 5.5: Among the 23 labeled points, only a few anatomical landmarks ((a), highlighted in red) are used as reference for shape translation (b), shape rotation (c), and shape scaling (d).

Since the anatomical landmarks that define the shapes (Figure 5.5 (a)) refer to the image coordinate system, as first, shapes were translated to the origin (Figure 5.5 (b)). Shapes were then rotated such that the left ileal tuberosity (hook

bone) and the right ileal tuberosity (hook bone) had the same horizontal coordinate (Figure 5.5 (c)). To perform translation and rotation of shapes, the middle point between the left hook bone and the right hook bone is taken into account. The choice of the mid-point between the hook bones provides a fixed point of reference for performing translation and rotation of the body image with respect to this point so that the shapes differs only on scale; shapes are finally scaled to fit in a unit square (Figure 5.5 (d)). The landmarks with minimum and maximum values of $x$ coordinates were used to scale the shape with respect to the $X$ axis, whereas the landmarks with minimum and maximum values of $y$ coordinate were used to scale the shape with respect to the $Y$ axis. After alignment, all shapes refer to the same coordinate system (centered into the origin) and can be modeled using the statistics on the 23 anatomical landmarks.

### 5.3.2 PCA based shape descriptor

We started by considering the following shape definition: "Shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object" [92]. Shapes are typically represented by locating a finite number of landmarks on the outline of an object. The mathematical representation for $n$ landmarks located into the shape of an object is a $2n$-dimensional column vector:

$$\mathbf{s} = [x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n]' = [s_1, s_2, \ldots, s_n, s_{n+1}, s_{n+2}, \ldots, s_{2n}]' \qquad (5.3.1)$$

Let $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ be a set of shapes and $\hat{\mathbf{S}} = \{\hat{\mathbf{s}}_1, \ldots, \hat{\mathbf{s}}_m\}$ the set of shapes obtained through the alignment procedure. The mean shape of $\hat{\mathbf{S}}$ corresponds to the vector that minimizes the sum of the squared-error criterion function with respect to the shapes in $\hat{\mathbf{S}}$. Hence, the sample mean, calculated according to formula 2.2.2, is the zero-dimensional descriptor of the dataset $\hat{\mathbf{S}}$ and can be considered as a "prototype" of the data, in the sense that it is the most similar to all the samples in the dataset, but it does not reveal any of the variability in the data. The modes of variations, i.e. the ways in which the points of the shape tend to move with respect to the average shape, can be found by applying principal component analysis to

the deviations from the mean [33, 86]. More specifically, taking into account the $2n \times 2n$ covariance matrix $\mathbf{C}$ built taking into account the samples in $\hat{\mathbf{S}}$, the modes of variation of the points of the shapes are described by the unit eigenvectors of $\mathbf{C}$, such that:

$$\begin{aligned} \mathbf{C}\mathbf{e}_i &= \lambda_i \mathbf{e}_i \qquad i = 1, \ldots, 2n \\ \mathbf{e}_i' \mathbf{e}_i &= 1 \qquad i = 1, \ldots, 2n \end{aligned}$$

where $\lambda_i$ is the $i^{th}$ eigenvalue of $\mathbf{C}$. The eigenvectors $\mathbf{e}_i$ of the covariance matrix corresponding to the largest eigenvalues describe the most significant modes of variations in the variables used to derive the covariance matrix, as detailed in Section 2.2.1.

### 5.3.3 Kernel PCA based shape descriptor

Shape descriptors based on Kernel Principal Component Analysis (KPCA) have been successfully used for statistical shape analysis and recognition [93, 94]. Kernel PCA is an extension of classic PCA to deal with non-linear cases using the technique of kernel methods. The basic idea beyond kernel methods is to map the data in the input space ($\hat{\mathbf{S}}$ in our case) into a high dimensional feature space via some non-linear function $\Phi$ and then apply a linear method in the augmented space to do further analysis. Note that PCA is a particular instance of the kernelized method, in which the mapping function is $\Phi(\mathbf{x}) = \mathbf{x}$.

Let $\Phi : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{n_\Phi}$ be a mapping function acting on the input space $\hat{\mathbf{S}}$. The mean shape of $\hat{\mathbf{S}}_\Phi$ can be simply computed as follows:

$$\bar{\hat{\mathbf{s}}}_\Phi = \frac{1}{m} \sum_{i=1}^{m} \Phi(\hat{\mathbf{s}}_i) \tag{5.3.2}$$

The sample mean $\bar{\hat{\mathbf{s}}}_\Phi$ is the vector $\mathbf{v}_\Phi$ that minimizes the sum of the squared error criterion function:

$$\bar{\hat{\mathbf{s}}}_\Phi = \underset{\mathbf{v}_\Phi}{\operatorname{argmin}} \sum_{i=1}^{m} \| \mathbf{v}_\Phi - \Phi(\hat{\mathbf{s}}_i) \| \tag{5.3.3}$$

The sample mean $\bar{\hat{\mathbf{s}}}_\Phi$ is the zero-dimensional descriptor of the dataset $\hat{\mathbf{S}}_\Phi$ and

the modes of variation, the ways in which the points of the shape into the Kernel space tend to move with respect to the average shape, can be found by applying principal component analysis (PCA) to the deviations from the mean $\bar{\hat{\mathbf{s}}}_\Phi$. In this way a shape into the kernel space can be considered as a linear combination of basis shape into the kernel space, and the basis components can be used as descriptor of the shape. Kernel PCA finds the principal axes by diagonalizing the following matrix:

$$\mathbf{C}_\Phi = \frac{1}{m}\sum_{i=1}^{m}\left[\left(\Phi(\hat{\mathbf{s}}_i) - \bar{\hat{\mathbf{s}}}_\Phi\right)\left(\Phi(\hat{\mathbf{s}}_i) - \bar{\hat{\mathbf{s}}}_\Phi\right)'\right] \tag{5.3.4}$$

Specifically, taking into account the $n_\Phi \times n_\Phi$ covariance matrix above, the modes of variation are described by the unit eigenvectors of $\mathbf{C}_\Phi$ such that:

$$\begin{aligned} \mathbf{C}_\Phi \mathbf{e}_j^\Phi &= \lambda_j^\Phi \mathbf{e}_j^\Phi \qquad j = 1,\ldots,n_\Phi \\ \mathbf{e}_j^{\Phi\prime}\mathbf{e}_j^\Phi &= 1 \qquad\qquad j = 1,\ldots,n_\Phi \end{aligned}$$

where $\lambda_j^\Phi$ is the $j^{th}$ eigenvalue of $\mathbf{C}_\Phi$. The eigenvectors $\mathbf{e}_j^\Phi$ of the covariance matrix corresponding to the largest eigenvalues describe the most significant modes of variations in the variables used to derive the covariance matrix $\mathbf{C}_\Phi$. Taking into account the considerations made by [85], where BCS is estimated using a parabolic fitting of the cows' body shape, in our experiments we have tested a polynomial mapping function to model the shape of cows.

### 5.3.4 Shape Descriptor and BCS estimation

The eigenvectors $\left\{\mathbf{e}_j^\Phi\right\}_{j=1}^{n_\Phi}$ useful to describe the shapes are computed using Kernel PCA (see Section 5.3.3). Any shape in the training set, mapped into the kernel space through $\Phi$, can therefore be generated by using the following equation:

$$\Phi(\hat{\mathbf{s}}_i) = \bar{\hat{\mathbf{s}}}_\Phi + \sum_{j=1}^{n_\Phi} a_{i,j}^\Phi \mathbf{e}_j^\Phi \tag{5.3.5}$$

where

$$a_{i,j}^\Phi = \mathbf{e}_j^{\Phi\prime}\left(\Phi(\hat{\mathbf{s}}_i) - \bar{\hat{\mathbf{s}}}_\Phi\right) \tag{5.3.6}$$

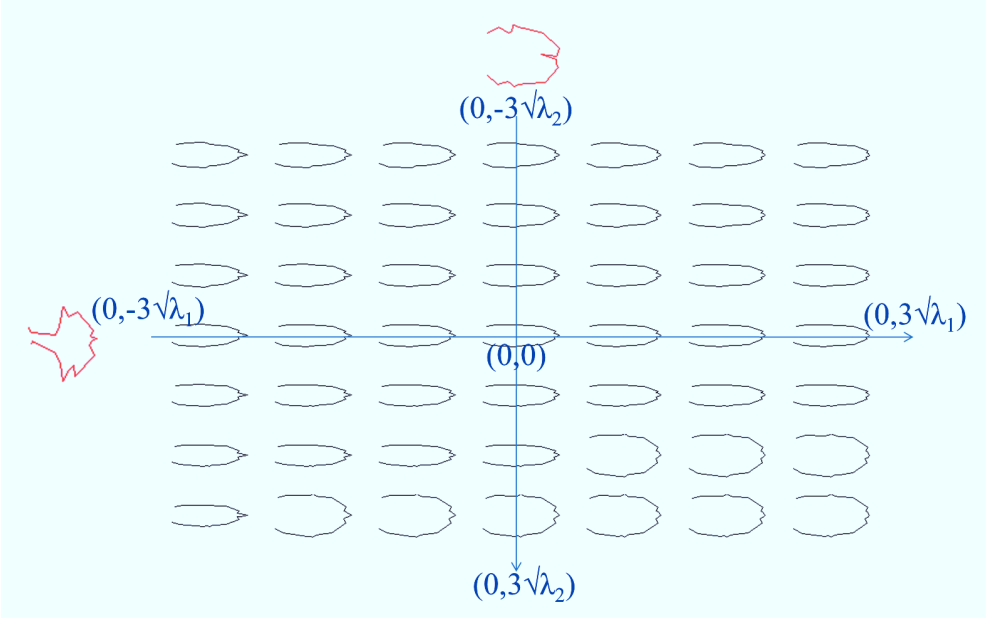



Figure 5.6: Shapes generated by using the third and fourth shape basis. In our case, these two basis are related to the information around the tail and the hooks of the cows.

The eigenvectors $\left\{\mathbf{e}_j^{\Phi}\right\}_{j=1}^{n_{\Phi}}$ are the set of basis of the shapes into the kernel space $\Phi\left(\hat{\mathbf{S}}\right)$ useful to generate new samples. Unseen shapes in the kernel space can be generated by changing the values of each $a_{j,k}^{\Phi}$ taking into account that its variance is represented by $\lambda_k^{\Phi}$. Since most of the sample of the training set lies within three standard deviations of the mean, the suitable range for $a_{j,k}^{\Phi}$ is $\left[-3\sqrt{\lambda_k^{\Phi}}, 3\sqrt{\lambda_k^{\Phi}}\right]$. The range of each $a_{j,k}^{\Phi}$ can be used to detect outlier that in our case are probably due to errors in manual labeling. Figure 5.6 shows an example of shape reconstruction taking into account only two vectors of the shape basis, computed with kernel principal component analysis and a linear mapping function. These two vectors appear to visually correspond to the modes of variation around the hooks and the rear of the cows.

Given a training set of shapes, kernel principal component analysis can be applied after alignment and hence each shape $\hat{s}_j$ can be described by using the vector $\mathbf{a}_j^{\Phi} = [a_{j,1}^{\Phi}, \ldots, a_{j,n_{\Phi}}^{\Phi}]$. The shape descriptors of the training set can be used

together with a linear regressor to build a system for BCS estimation:

$$BCS_j = a^{\Phi}_{j,n_\Phi} w_{n_\Phi} + a^{\Phi}_{j,n_\Phi-1} w_{n_\Phi-1} + \ldots + a^{\Phi}_{j,1} w_1 + w_0 \tag{5.3.7}$$

Given the descriptors of the shape in the training set, the regression model can be fitted by using least squares method. The learned parameters $[w_0, w_1, \ldots, w_{n_\Phi}]$ are then used to infer the BCS of new shape samples describing them by using the basis $\left[\mathbf{e}^{\Phi}_1, \ldots, \mathbf{e}^{\Phi}_{n_\Phi}\right]$ learned on the training set.

## 5.4 Experiments and Results

### 5.4.1 Experiments description

The PCA and kernel PCA based shape descriptor methods were evaluated by comparing them to other existing methods of BCS scoring using digital images [84, 85] from the selected database. In our experiments, we employed kernel PCA to model the shape of cows. Specifically, we used a linear and a polynomial mapping function, respectively $\Phi(\mathbf{x}) = \mathbf{x}$ and $\left(\Phi_{Polynomial}(\mathbf{x}) = \left[x_1^2, x_2^2, \sqrt{2x_1x_2}\right]\right)$, for $\mathbf{x} = [x_1, x_2]$).
In [84] the authors use a single image of the dorsal view of the cow and manually select 23 anatomical points to define the shape of the body of the cow. These points are used to compute 15 angles around the hooks, pins and tailhead. The authors exploit the symmetry of cow shape to obtain 7 composite angles averaging the left and right angles of the shape. The related proposed models are reported below:

$$y_{ij} = \mu + Cow_i + \beta_1 HA_{ij} + \beta_2 PHA_{ij} + \beta_3 \left(HA \times PHA\right)_{ij} + e_{ij} \quad \text{(Bewley 1)} \tag{5.4.1}$$

$$\begin{aligned} y_{ij} =& \mu + Cow_i + \beta_1 HA_{ij} + \beta_2 PHA_{ij} + \beta_3 \left(HA \times PHA\right)_{ij} + \\ &+ \beta_4 TD_{ij} + \beta_5 \left(PHA \times TD\right)_{ij} + e_{ij} \quad \text{(Bewley 2)} \end{aligned} \tag{5.4.2}$$

where:

- $y_{ij}$ is the $j^{th}$ BCS of the $i^{th}$ cow estimated by technicians;
- $\mu$, $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$, and $\beta_5$ are the regression parameters;
- $Cow_i$ is the identifier of the $i_{th}$ cow;
- $e_{ij}$ is the residual error;
- $HA_{ij}$ is the average hook angle,
- $PHA_{ij}$ the average posterior hook angle
- $TD_{ij}$ is the average tailhead depression.

In our experiments, the term $Cow_i$, in the Bewley 1 and Bewley 2 formulas, was not taken into account to test the models since this value does not provide anatomical information useful for BCS estimation: technicians do not use the ID of the cow to estimate the BCS and if in the training phase the ID was included in the learning process, the BCSs of unseen cows (never captured by the system) could not be estimated. Moreover, we think that by knowing the ID of the cow, particularly when BCS is assessed over time, the BCS estimation could be biased from two sources:

- technical observers, influenced by prior BCS of a particular cow;
- the machine algorithm. Since the machine algorithm learns parameters of the model from BCSs, when cow IDs are associated a priori with a BCS, changes over time within a cow may be masked by linking BCS with cow ID. Cow BCS must be assigned prior to linking with cow ID to minimize this bias.

In [85] the posterior part of the cow is considered and a parabolic fitting is performed. The absolute differences between the real body shape and the fitted parabola are used in the BCS estimation as follows:

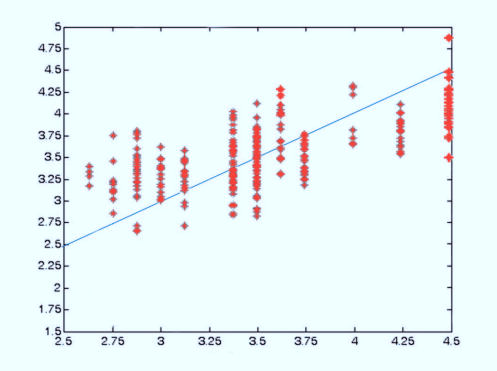
$$BCS_{Thermal} = 5 \times B_{fit} \times \frac{1}{MAE} \quad \text{(modified Halachmi)} \tag{5.4.3}$$

where $B_{fit}$ is the best fit reached in the herd, $MAE$ is the mean absolute error, and 5 is a normalization factor. In our evaluation phase, we employ a modified version of the Halachmi method. Parabolic fitting and BCS estimation are performed considering only the labeled points because we did not use a thermal camera to acquire images and therefore shape extraction from those images is a complex task not addressed in this first step of the study. The mean BCS estimation error is defined as:

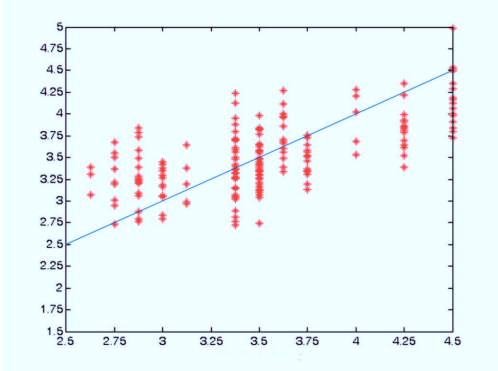
$$\overline{BCS_{ERROR}} = \frac{1}{N} \sum_{i=1}^{N} \left| BCS_{est}(i) - BCS_{tec}(i) \right|, \tag{5.4.4}$$

where $N$ is the number of images (i.e., 286 in our dataset), $BCS_{est}$ and $BCS_{tec}$ are the BCS values estimated from digital images and the mean BCS manually evaluated by technicians, respectively.

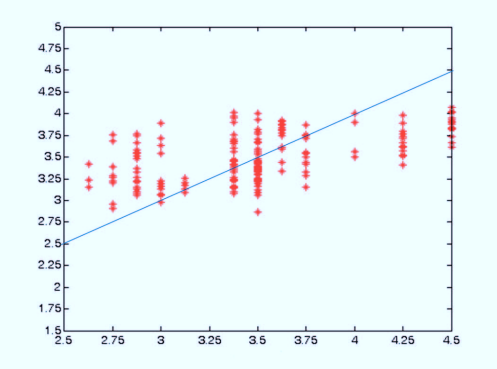
In order to assess the effectiveness of the methods, the Leave One Out Cross Validation (LOOCV) procedure and the Regression Error Characteristic Curves (REC) [95] are used. Each run of LOOCV involves a single observation of the dataset as test, and the remaining samples as training data. This is repeated to guarantee that each sample is used once as the test data. The average error rate is computed taking into account all runs. The REC curve is essentially the cumulative distribution function of the error. It was derived by plotting the errors tolerance versus the percentage of samples predicted within the tolerance. The area over the curve is a biased estimation of the expected error of an employed regression model. This technique enables simple assessment of different regression models by examining the relative position of their REC curves.



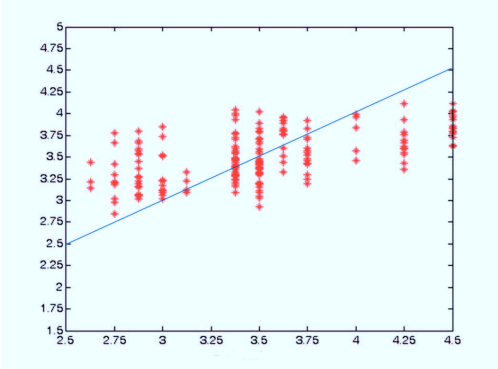
(a) Predicted (Linear Kernel PCA) versus actual BCS.



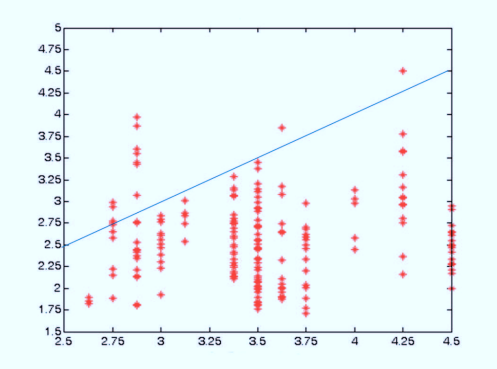
(b) Predicted (Polynomial Kernel PCA) versus actual BCS.



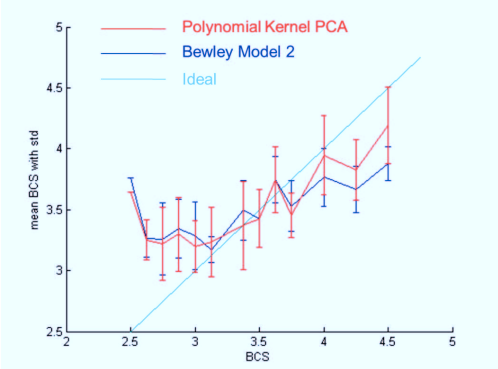
(c) Predicted (Bewley model 1) versus actual BCS.



(d) Predicted (Bewley model 2) versus actual BCS.



(e) Predicted (Halachmi model) versus actual BCS.




(f) Polynomial Kernel PCA approach compared with Bewley model 2.

Figure 5.7: (a)-(e) BCS predicted by the compared approaches versus actual BCS (estimated by technicians). (a) Linear kernel PCA; (b) Polynomial kernel PCA; (c) Bewley model 1; (d) Bewley model 2; (e) Halachmi model. (f) Polynomial Kernel PCA approach (red line) versus Bewley model 2 (blue line): mean values and standard deviations are reported.

| Method | Mean BCS error |
|---|---|
| Modified Halachmi | 0.9837 |
| Bewley (model 1) | 0.3295 |
| Bewley (model 2) | 0.3289 |
| KPCA - Linear Kernel | 0.3218 |
| KPCA - Polynomial Kernel | 0.3059 |

Table 5.1: Mean BCS error comparison

### 5.4.2 Results

BCS is estimated using the five different models over the images of the training set and errors were averaged for all runs of LOOCV; results of errors obtained from BCS estimation are reported in Table 5.1.

The modified Halachmi approach is not able to provide satisfactory results (Figure 5.7 (e)): the parabolic fitting might be not accurate enough since it is performed considering only the labeled points. Bewleys models have similar results (model 2 performs slightly better than model 1) (Figure 5.7 (c-d). Their performances are better for central BCS values (around 3.5) and worst in the extreme cases (2.5 and 4.5). Our approach, in particular the polynomial one, performs better than the other techniques, obtaining satisfactory results even in the extreme cases. Results of our approach are reported in Figure 5.7 (a-b). As shown in Figure Figure 5.7 (f), our approach using a polynomial kernel is able to follow the ideal line better than Bewleys approach. In Figure 5.8 the comparison through REC curve confirms that the proposed approach performs better than state-of-the-art methods in estimating BCS.

## 5.5 Discussion: toward a fully automatic pipeline

The cows body shape is described considering the deviation from the average shape in the kernel space. Our method produces a description of the shape to be used for automatic estimation of BCS. Experimental results confirm the effectiveness of the proposed approach. A second contribution of this work is the new benchmark dataset useful for research purpose, publicly available through

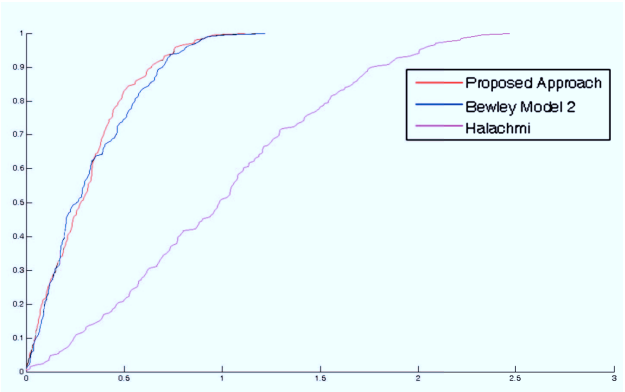



Figure 5.8: Regression error characteristic curves (RES) of some models involved in the comparison: Bewley's model 2 (blue line), Polynomial kernel PCA approach (red line) and Halachmi model (magenta line).

the Internet. BCS estimation systems that work fully automatically (with no user intervention) or at least semi-automatically (with minimal user intervention) are desired to cut down time and costs of the traditional BCS estimation techniques. Moreover, these systems can produce an objective evaluation of BCS in a way that is less invasive for the cows. The main drawback of a semi-automatic system for BCS estimation is due to the manual labeling of anatomical points during both training and employing phases. A fully automatic system for BCS evaluation, in which the shape of a cow is automatically extracted from digital images through segmentation procedure, would be desirable. The automatic extraction of the whole shape could be also useful to build a more robust model for BCS estimation by using more anatomical points, automatically extracted. Hence we decided to extend our research, toward the direction of a fully automatic system for BCS estimation. The proposed technique is described in the following sections.

## 5.6 Automatic BCS estimation: proposed pipeline

The basic approach described in previous sections of this chapter is here extended to grant a fully automatic scoring method that requires no human intervention. The infrastructure for image acquisition is the same as described before; our intervention is mainly devoted to avoid manual labelling by experts in the *Employing*

*Block* (indicated by *Identification of Anatomical Features* in Figure 5.1), which needs to be replaced by automatic shape extraction. Once the whole shape of the cow has been extracted (similarly to previous sections "shape" here is intended as the cow's silhouette as seen vertically from above) we can sub-sample the contour obtaining an arbitrary number of points; for instance we could make use of more points than in our semi-automatic approach, and hence use more information about the contour, to derive our shape descriptor for BCS estimation, avoiding labelling errors and potentially enabling higher accuracy.

### 5.6.1 Image Segmentation and Shape extration

The network camera used in our system records standard RGB images. In RGB space, each color is represented as a triple ($R$, $G$, $B$), where $R$, $G$, and $B$ represent red, green, and blue signals corresponding to different wavelengths of the visible spectrum. An approximately perceptually uniform color space is defined in terms of Hue, Saturation and Value ($HSV$), a phenomenal color space. Phenomenal color spaces attempt to classify colors in relation to how they are perceived and interpreted by the human brain and they are more "intuitive" in manipulating color.
In our preliminary tests $HSV$ led to better segmentation performance than $RGB$ in both noisy and noise-free conditions. According to experimental evidence, a preliminary conversion from $RGB$ to $HSV$ color space is hence performed. The $H$ channel obtained from the color space conversion is processed by means of the Statistical Region Merging (SRM) algorithm [96], a segmentation technique able to capture the main structural components of a digital image using a simple but effective statistical analysis. SRM is based on two parametric choices:

- *Merging Predicate*, used to establish if two regions have to be merged;
- *Merging Order*, used to establish the order used to test the Merging Predicate.

The *Merging Order* is computed using a function $f$ which approximates the following invariant: when any test between two true regions occurs, that means

that all tests inside each of the two true regions have previously computed. Different choices can be done both for *Merging Factor Q* used by the *Merging Predicate* and the $f$ function. In our experiments we heuristically choose $Q = 4$, since it produces better shape estimates. In Figure 5.9 some segmentation results are reported; note that after segmentation, locating the shape of the cow is a straightforward procedure (e.g. since the gate exit gate, described in Section 5.2.1, appears always in the same position in all images (Figure 5.9 (a-c), left side, it can be easily removed).

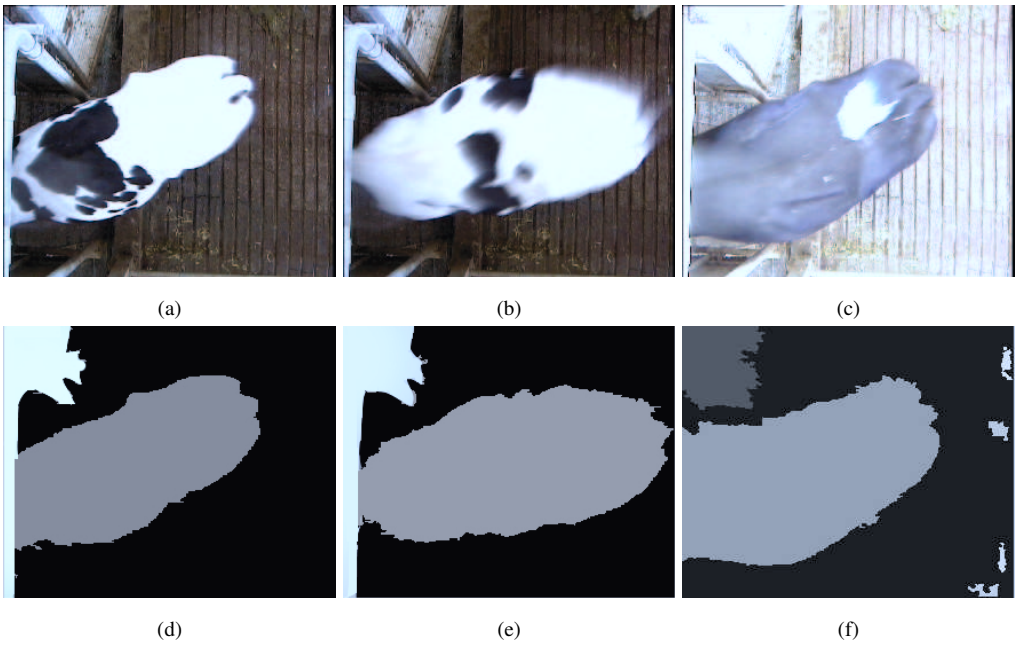
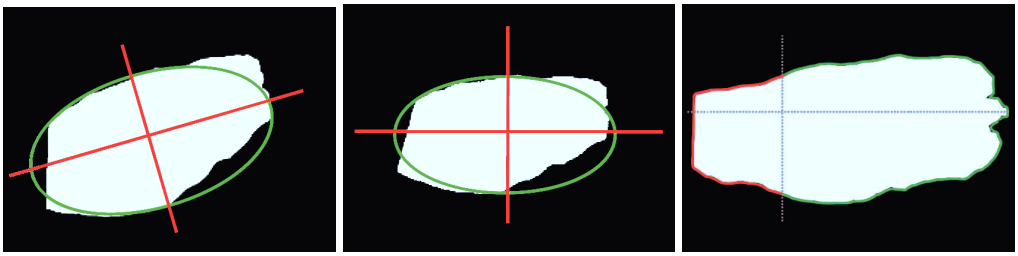



Figure 5.9: Some examples of the output produced by the SRM algorithm applied to cows' images; input images are converted into HSV color space before performing the segmentation. Top row: RGB input images; Bottom row: corresponding SRM output.

The output of this phase is then transformed into a curve representation and used for subsequent analysis; a morphological regularization for noise reduction is also performed.

To obtain a consistent shape representation, shape must be aligned to a common coordinate reference, removing position, scale and rotation. Please note that the procedure described in Section 5.3.1 cannot be applied here, since we don't have the manually labelled points required. Instead, we can consider the shape of a cow

as an ellipsoidal object and use a PCA based algorithm to find its major axis which are then rotated and aligned with respect to $X$ and $Y$ axis, filtering out rotational effect (see Figure 5.10 (a-b)).

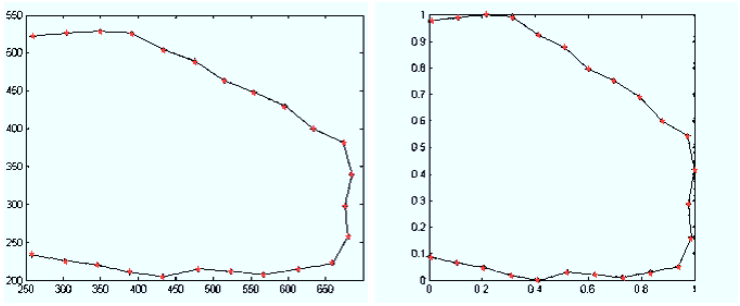


(a) (b) (c)

Figure 5.10: The shape of a cow can be roughly approximated as an ellipsoidal object (a). A simple PCA is applied to all points within the shape, to find the major axis of the ellipsoid; final alignment with respect to $X$ and $Y$ axis filters out most rotational effects (b). To filter out the effect of random neck twist (c) we consider a fixed portion of the whole silhouette (the 70% of the whole curve centered at the cows tail).

### 5.6.2 Shape descriptor and BCS estimation

An uncontrollable source of variation in cows shape is the random neck twist taken by the animal at the moment when it has been photographed. To filter out this effect we consider for further processing only a portion of the whole silhouette, more precisely the 70% of the whole shape contour, centered at the cows tail (5.10c). Taking into account the previous caveat, once alignment of all the curves has been obtained, the contour of the shape is uniformly sampled using *arc-length parameterisations*, leaving us with a set of $n$ points along the contour (Figure 5.11 (a)). Once shapes are aligned to the axis and the neck part ha been removed, shapes are scaled to fit in a unit square (Figure 5.11 (b)). After alignment, all shapes refer to the same coordinate system. Uniform sampling leaves us with each shape described as a set of points coordinates, which can be represented using Equation 5.3.1. Application of Polynomial Kernel PCA is straightforward, and we can safely use Equations 5.3.1 - 5.3.7 to derive a shape descriptor and a BCS estimated for each involved shape, as described in previous sections.



(a) (b)

Figure 5.11: The contour of the shape is uniformly sampled (a) and shapes are then scaled to fit in a unit square (b).

### 5.6.3 Results and Conclusion

To validate our automatic pipeline we adopted the evaluation methodology described in Section 5.4. We fixed the number $n$ of sampled point along the shape contour to 24, to better compare this method with other approach, all based on the 23 anatomical points described in [84]; the parabolic fitting method proposed by Halachmi is not further considered because of its poor performances, as reported in Section 5.4.2. We tested our technique on a subset of the BCS dataset, since some images included in the whole dataset are excessively blurred and/or the automatic white balancing performed by the network camera produced a wrong result (e.g. bluish aloes). In this cases only manual labeling could be used and even though, it would be error prone (see Figure 5.12 for some examples of images not included in this experiment). Please note that the same subset has been used also with the other techniques, for a fair comparison. Numerical results are reported in Table 5.2. As it can be noted, the removal of blurred and low quality images from the dataset improves performances of all methods included in this experiment (see also Table 5.1), of a percentage about 7-13%. Figure 5.13 (a-e) reports the plots of predicted BCS versus the ground truth, for each method considered in our experiments, as well as the REC curves. The automatic approach generally gives better results than other models, including our semi-automatic techniques, except for BCS values $2.5 < BCS < 3$. The reason could be the low number of samples, with a BCS values within this range, available in our dataset. This observation suggests to extend our dataset including more images of cows with a low BCS

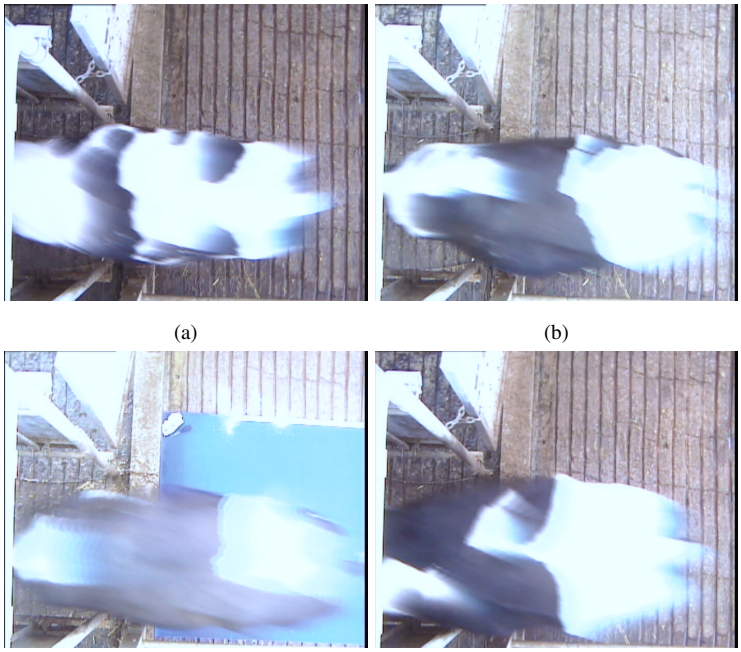
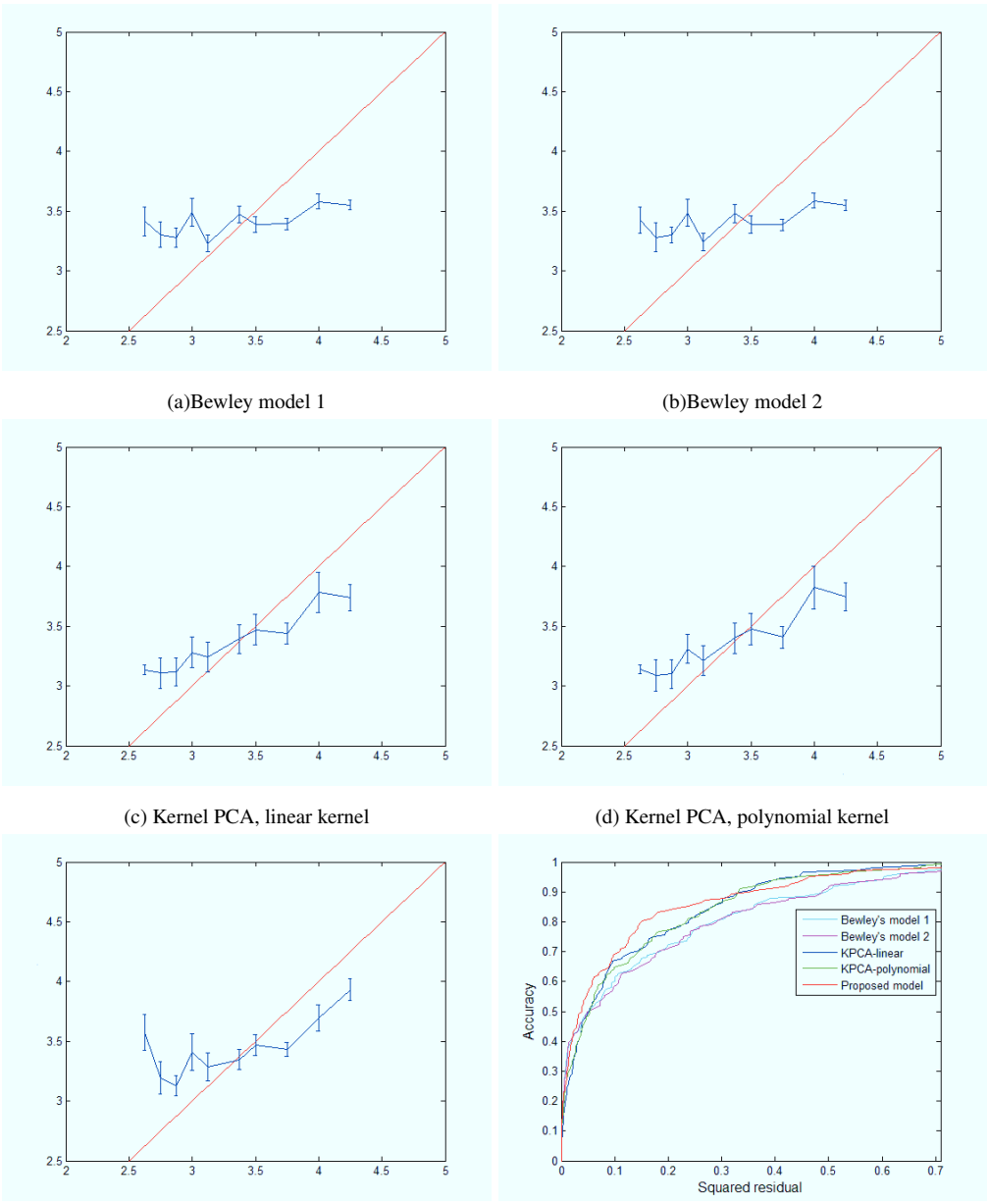



Figure 5.12: Sample images affected by blurring and/or wrong automatic white balancing, which have been excluded from the BCS dataset. In almost all cases images were took during daytime.

value; moreover the use of a camera without automatic white balancing could be beneficial, allowing us to perform this step in post-processing, employing ad-hoc techniques. At the best of our knowledge, the fully automated technique we reported in the second part of this chapter, is the first technique that can provide reliable BCS estimation without the direct human intervention and as such it has promising practical applications.

| **Method** | **Mean BCS error** |
|---|---|
| Bewley (model 1) | 0.3008 |
| Bewley (model 2) | 0.3039 |
| KPCA - Linear Kernel | 0.2813 |
| KPCA - Polynomial Kernel | 0.2787 |
| KPCA - Linear Kernel, Automatic Pipeline | 0.2613 |

Table 5.2: Fully automatic pipeline: mean BCS error comparison.




(a)Bewley model 1

(b)Bewley model 2

(c) Kernel PCA, linear kernel

(d) Kernel PCA, polynomial kernel

(e) Automatic segmentation, Kernel PCA, linear kernel

(f) REC curve

Figure 5.13: (a)-(e) BCS predicted by the compared approaches versus actual BCS, estimated by technicians. Mean values and standard deviation are plotted. (a) Linear kernel PCA; (b) Polynomial kernel PCA; (c) Bewley model 1; (d) Bewley model 2; (e) Halachmi model. (f) REC curves.

# Chapter 6

# Shape Classification for Articulated and Deformed Shapes

## 6.1 Introduction

In the past decades, the application of the computer science to archaeological research, in particular in the field of prehistory, turned out from a simple auxiliary technology into a cognitive strategy influencing the approach to ancient artifacts. Even if the introduction of the database enhance the possibility of dealing with large amount of text data, the problem of taking into account huge groups of visual data still remains unsolved. One common case that makes hard the initial steps of an archaeological study is for example the analysis of decorative repertoires of some prehistoric pottery classes, characterized by the exuberant use of a multiplicity of motifs. The most complicated artistic production of the Mediterranean prehistory is certainly the Kamares style pottery (Fig. 6.1), flourished in Crete in the first half of the second millennium, whose main feature is a complex system of polychrome painted decoration with a rich decorative alphabet that is combined, according to an elaborate visual syntax, to produce an almost endless variation of results. The high number of vessels and potsherds available of this class all over Cretan territory and the fact that a complete framework for the Kamares decorative grammar and the associations motif/vessel/site/chronological layer is far from being understood, determined a stop in the cognitive progress of

Figure 6.1: Examples of Kamares style vessels (this image has been obtained as a collage from several public sources).

the Minoan civilization.

In this perspective Computer Vision and Pattern Recognition can provide a great support in automatically assisting the archaeologists in classification according time and place of production of Kamares pottery fragments (Fig. 6.2), also taking into account that, in some cases, the visual information can be of several kinds, like watercolors, black and white and color photographs of unequal resolution and quality, pencil sketches and high quality digital photographs.

The application of this research strategy on the available visual corpus of image data is the object of the research that we present in this chapter. The chapter is structured as follow: in the remainder of this Section we introduce the Kamares pottery. Section 6.2 shows the proposed technique and a short review of existing shape similarity techniques. Section 6.3 reports the experimental results; a conclusion Section ends the chapter.

**The Kamares Pottery.** The Kamares style pottery represents the main artistic feature of the Minoan civilization between the XXth and XVIIth century BC and its technical and stylistic level achieved was never equaled in the Aegean Bronze Age [97], [98], [99], [100]. This class, with a large lifespan that covers the Pro-

Figure 6.2: Selection of Kamares pottery sherds from Phaistos (courtesy of Prof. V. La Rosa).

topalatial and Neopalatial periods, can be divided in four main phases related to the local chronology arranged by A. Evans [101]: Pre-Kamares or 1 (Middle Minoan IA), Early Kamares or 2 (Middle Minoan IB/IIA), Classical Kamares or 3 (Middle Minoan IIA/IIB/IIIA), Post-Kamares or 4 (Middle Minoan IIIA/IIIB). Kamares ware is named for finds first identified at the end of XIX century in the Kamares cave sanctuary [102] in Mount Ida, and its diffusion on the Cretan territory can be distinguished between the elegant production of the palatial centers of Knossos and Phaistos (Palatial Kamares) and the less impressive creations of the provincial workshops of Eastern, Western, Central and Eastern-Central Crete (Provincial Kamares areas 1-4) as in Fig. 6.3.

Finer clay, thrown on the wheel, sometimes with very thin bodies (2-3 mm like in the egg shell varieties), painted with polychrome pigments on dark background, the Kamares style has a vast assemblage of shapes and its distinctive feature is a huge repertoire of exuberant decorative motifs often resembling naturalistic atmospheres (Fig. 6.1). About the colors, the black slip which covers most of Kamares vessels has been identified as a clayish paint layer containing iron oxides of spinel type with a high content of potassium. It is produced by exposure to high temperatures in a reducing atmosphere, thereby turning in into a black sintered layer. The red color consists of red ochre, which in contrast with the black slip does not contain potassium at all. As a result of the lower content of potassium, the red paint layer does not melt and turn into a sintered layer in a reducing atmosphere

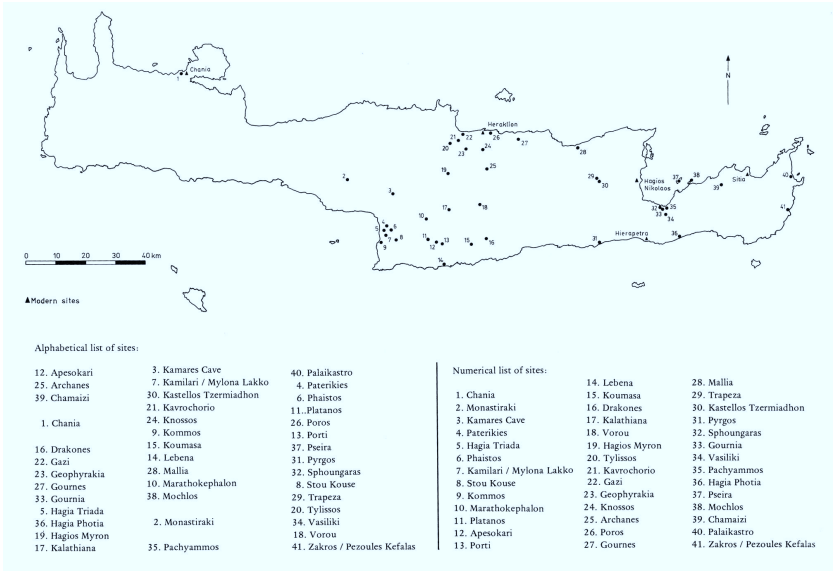



Figure 6.3: Map of Crete showing principal sites with Kamares style pottery.

but it remains porous and turns bright red when exposed to oxidation. The white pigment is composed by calcium silicate or talc depending on the chronological period of the production and of the production centre itself. In the decoration of large vessels of domestic type also a less elaborated and much lasting dark on light decoration system is used as it was found appropriate for vases which had to be strong and practical rather than attractive.

In a preliminary survey, seventy-four different shapes were identified including both the decorated and fine ware and domestic and specialized forms. During the four chronological phases, open and closed shapes are equally preferred and even if there is a general conservativeness in the use of the same formal typologies some specific tendencies for each phase can be defined (Fig. 6.4).

For what concerns the decoration, the Kamares style is basically characterized by the application of thirty-one different core motifs (abstract, rectilinear, pictorial, pictorialized motifs) and patterns (stone, rock, sponge and metal surface patterns), each of them with several levels of varieties, composed together in order to obtain complex decorative syntaxes joined to create much more complicated composite designs. In general the repertoire of core motifs maintains itself the same in both the main production districts, the Palatial and Provincial, and

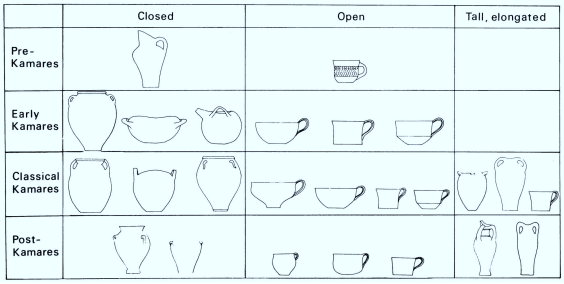



Figure 6.4: Schematic shape taxonomy of the Kamares pottery repertoire.

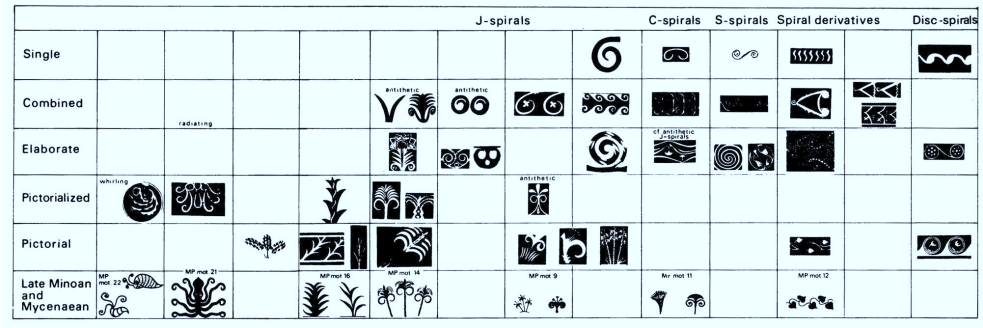



Figure 6.5: Schematic taxonomy and possible elaborations of some Kamares decorative motifs.

basically remains unaltered during all the chronological development, just with few elements peculiar for some phases. But it must be also considered that a core motif, even simple, can deeply change shape when transformed from a 2D sketch to a 3D design on the vessel body, when combined with copies of the same motif or joined with other motifs, when elaborated or pictorialized and in relation to different chronological phases. As a result, in some cases, the classification of elements can be very hard (Fig. 6.5).

About the composition of decorative elements and its relation with the tectonic of the vessel, two different method can be identified. First is the unity decoration that depends on the shape of the vase and that can be circumcurrent, that means planned with regard to the whole vessel body and composed by encircling zones, facial, that means planned with regard only to one part of the vessel body as seen from a special angle, and zonal, that means composed by a system of decorated

zones related to each other in order to create coherent designs. Second is the structural decoration that consists in separation of the vase into different parts and accentuates the horizontal and vertical axes. In the both type of decoration a field division for the location of the motifs is applied, selecting different parts of the body vessels for the definition of the limits of the decorated area. In the field division the choice of the decoration strategy for the accessory parts of the vase, like rim, lip, handle, spout, is very peculiar of each period and production centre. It must be also considered that both the shapes and the decoration contain dynamic effects and the decoration cannot merely be described as a movement across a surface, as in the case of a static background. It must be described as a movement in the same direction or opposed to the movement inherent in the vessel shape. The dynamic effects of the vase and the decoration together create the overall effect, that can give, for example, the illusion of the contraction or expansion of some vessel parts. In the selection of the composition strategies it's possible to find difference between the Palatial and Provincial production and to define tendencies from the Pre-Kamares to the Post-Kamares phases.

The overall effect of a Kamares vessel is sometimes completed by the application of an accessory decoration represented by incision, relief and applied plastic details or by the use of complementary embellishment techniques like the ripple, barbotine and the trickle decoration. Finally, it must be noticed that differences can be find also within the same centre and in the same period because, besides the guidelines of this production, the essence of the style itself was strongly influenced by the creativity and aesthetic concept of the single artisan, that constantly elaborated the core motifs and experimented new ways of expression. As a result, the Kamares style is more than the simple combination of syntaxes of elementary motifs with different vessel type. It is a complicated interplay, carried out to the smallest detail, where every elements has an important part, especially in the Palatial production. The fact that each Kamares style product depends by many quantifiable variables (time, place, vessel shape, high number of probable composition of core decorative elements, colors and accessory decoration techniques) and just by one variable not quantifiable at all that is the unpredictable will of the artisan, it made quite impossible for the archaeologists specialized in the Middle Minoan pottery to carry out a exhaustive study of the main features of this style,

also because of the huge amount of data represented by millions of decorated potsherds.

The first study of classification of the Kamares pottery was carried out by G. Walberg in 1978 and it was dedicated to the Palatial production of Phaistos and Knossos [97], [99] (Fig. 6.3). In this work, the position of Phaistos, as the most important production centre and also the only site with a clear chronological progression in the development of the production itself, were pointed out and corpus of core motifs and patterns was published. Furthermore, the features characterizing the climax of the Kamares style were find between Middle Minoan IIA and Middle Minoan IIIA, within the phase named Classical Kamares. In 1983 a second reassessment of the Kamares materials from Provincial districts was carried out by the same author [98], emphasizing assonances and dissonances between Palatial and Provincial production and trying to rebuild the decorative repertoire, as whole as possible, in order to enrich the main assemblage and obtain a complete study and classification of the Kamares pottery.

After the Walberg's corpora, a large quantity of new Kamares vessels were found and published together with groups of pottery coming from old excavations, in both cases often fragmented potsherds. This re-opened the problem for the archaeologists of dealing with a class of materials so variable present in huge amount of specimens. In particular the interpretation of the decorative motifs and syntaxes partially preserved on the sherds, in order to ascribe them to the repertoire of a specific workshop, trying to match the fragmentary information with a standard collection of visual references arranged by the scholars, has become the hardest part of the research.

Even if the goal of an accurate and exhaustive classification according time and place of production of Kamares pottery fragments is likely to be unobtainable with the present state of the art, Computer Vision and Pattern Recognition could provide a great support in automatically assisting the archeologists in the classification task. Our contribution illustrates a complete pipeline to automatically process these data. The processing starts with the extractions of a clean representation of the decorative designs. Our goal is to create a standardized database of "shapes" that could be successively automatically investigated with Pattern Recognition methods.

**Iconographical Documentation Available.** Besides the large amount of specimens available, another significant problem in an exhaustive study of the Kamares style pottery is dealing with an heterogeneous group of iconographical data. For the two most important Palatial production centers, Phaistos and Knossos, the documentation is basically composed by watercolors, black and white pictures, technical drawings in scale with front view and section and reconstructive unscaled perspective drawings. The best preserved vessels representing the highest stylistic level of the Kamares style, in the editions of the two sites, were illustrated by unscaled watercolors carried out by E. Stafani, R. Oliva and Th. Fanourakis for Phaistos and D. Mackenzie for Knossos. In the reproduction of the complex Kamares painted decoration, each modern artist or technician at work was influenced by his own personal taste and this caused, in many cases, alteration and distortion of the original shape of the motifs and of the original scheme of the syntaxes.

It must be also considered that the draftsmen working in Crete in the first half of the XX century, had different formation and artistic education. For the reproduction of the Phaestian Kamares, E. Stefani, chosen by the director L. Pernier for the documentation of the excavation, was an architect. Decades later, D. Levi, new head of the expedition, chose R. Oliva and Th. Fanourakis, two painters strongly influenced by his archaeological point of view. Otherwise at Knossos, a large part of the graphical documentation, including the watercolors, was carried out by D. Mackenzie [103], that was an archaeologist, second in charge after A. J. Evans, head of the mission. In this case he demonstrated a more straight and scientific method of depicting artistic features [104].

Furthermore, the fact that the documentation of Kamares vessels from Knossos dated back to 1921-1935 [101] while that of Phaistos was performed in 1935-1951 and 1976 [105], [106], [107], it determined different choices in the reproduction strategies, like different line thickness or colors, due to the changing taste of the time (Fig. 6.6). Finally, another misleading problem in the interpretation of the Kamares pottery from Knossos come also from the reconstructive drawings, where in many cases the hypothesis of the missing part of a motif or of a decorative outline was denied by the subsequent findings.

In this chapter the results achieved working basically on the visual archive provided by Walberg's corpora [98], [99] will be discussed.

(a) (b) (c)

Figure 6.6: Watercolors of Kamares style pottery and potsherds: (a) Phaistos 1935 drawings by E. Stefani; (b) Phaistos 1976 drawings by R. Oliva; (c) Knossos 1935 drawings by D. Mackenzie.

## 6.2 Kamares Shape Analysis: System Overview

The automated system that we propose to assist the archaeologist in assessing similarities among the Kamares findings, follows a general pipeline that could be easily adapted to other similar case studies. As in most Computer Vision applications the idea is to go from the raw data (pixels) to a symbolic representation of the image content. These information may, in turn, be fed to an intelligent system that assists the experts in formulating and checking working hypotheses about the scenes captured in the original images.

More specifically, in this case study the system starts with a digitalized format of a pictorial representation of a vase or of a sherd. Fig. 6.7 shows the successive steps performed. The digital data are the input of a contour extraction algorithm. Please, notice that in this way we are disregarding any 3D information: perspective deformation, and occlusions together with noise and artifacts due to low resolution are simply ignored at this stage.

Contour extraction produces a collection of contours: some of them are not relevant for our application and should be filtered out [108],[109],[110]. We choose to ask the interventation of the expert for this ROI identification task because of

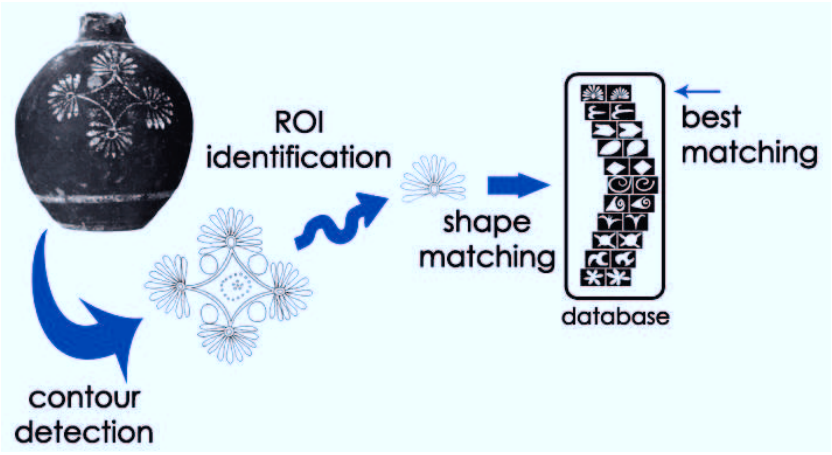



Figure 6.7: The complete pipeline of the proposed technique.

the occurrence of cases like those reported in Fig. 6.8, in which several occurrences of the "raquet" decorative pattern are shown. It is clear that the isolation of this complex motif is the product of an informed semantic choice that is not easily mimicked with a fully automatic approach.

ROI identification most commonly reduces the set of contour shapes to examine. Even so Kamares decorative motifs are reduced by this step into a smaller but yet complex set of elementary shapes. Fig. 6.9 shows several instances of the patterns that are obtained after ROI selection. Observe that in some case it is more convenient to refer to the "flattened" version of the motifs whenever these, manually produced by an expert draftsman, are available. These lucky cases reduce the problems due to 3-D distortion and occlusion but introduces in the pipeline the bias of a draftsman.

The proposed system deals with simple shapes one at each time. They are analyzed by the shape analyzer and reduced to a set of numerical and geometrical invariants. These invariants are the final features that the matching module of the system considers for search and similarity retrieval into a reference shape database.

The reason to deal with elementary shapes instead of the set of contours in a ROI has the following motivations:

- Algorithms that deal with simple contours and extract suitable invariants

(a) (b)

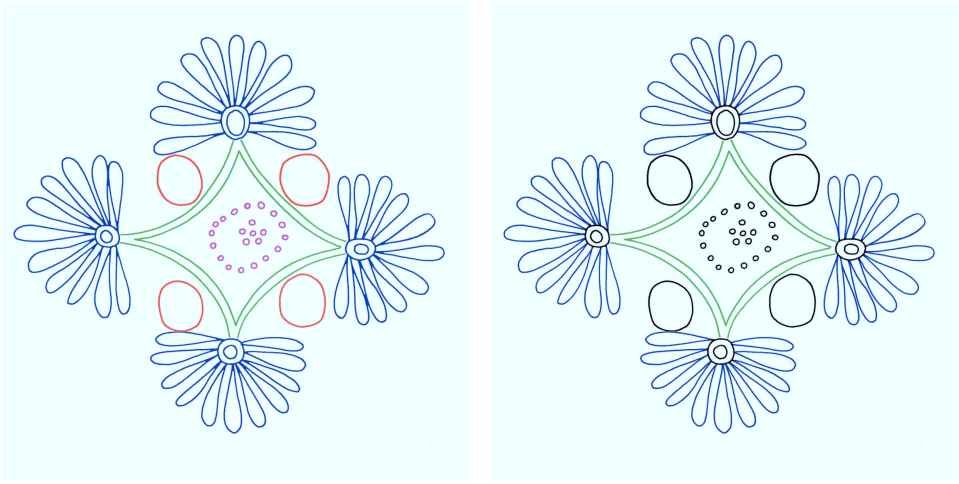
Figure 6.8: Examples of occurrences of the “raquet” motif in different vases.

Figure 6.9: Kamares potteries and relative decorations.

from them are much better understood and robust at the actual state of art;

- A ROI represents a complex figurative “proposition” composed of simpler elementary shapes. Unfortunately, the “order of reading” of such complex visual structures is not linear. Indeed it is not easy to “read” those drawing in a canonical un-ambiguous way.

Fig. 6.10 reports a case of two different ways to isolate ”visual word” within the same ROI. Archaeologists would in this case probably prefer the visual hierarchy on the left (Fig. 6.10(a)) while, on the other hand, a computer algorithm would find more appropriate the right one (Fig. 6.10(b)). It is hard, at this stage of the research to provide a way to resolve these ambiguities. It is hence a safer choice, for now, to find and match only elementary shapes. Next steps of our research will address this issue.



(a) (b)

Figure 6.10: (a) figure interpretation by the archaeologist. (b) figure interpretation by the computer scientist. The same color indicate the same symbol.

### 6.2.1 Shape Similarity Measures

To apply the technique presented in this chapter a contour line representation of the motifs is required. Hence, a decorative pattern is previously translated into a digital sequence of consecutive points on a raster plane. Several issues about resolution and standardization of these rasterized contours arise. Going from digital images to digital shapes is far from being an easy task: different media (water colors, photos and hand drawn schematic lines) require different methods of contour extraction and present different algorithmic challenges. Once the pictorial data have been translated into a more abstract "contour shape representation" it is possible to apply a shape similarity measures. The idea is to present the shape database with a "query" which represents a decorative element that the user (typically an archaeologist) wishes to recognize. The answer to the query is an estimate of the similarity distance between the query shape and the shapes in the database. In this chapter er propose the use of the following shape similarity algorithms: contour flexibility technique [39], Shape Context [38], and Circular Blurred Shape Model descriptor [111]; these methods are described in Sections 2.2.2-2.2.2.

Notice that even if the published ideas in shape recognition are overabundant

in the present scientific literature most of them are relative to complex vision tasks where one wishes to recognize real objects that move and deform within a complex real environment. This is not the case for our application where the most common deformations observed are affine distortions, noise, irregularities, lacunae, and non linear scaling.

## 6.3 Results

To test our system we have built an experimental image database, created as follows. First of all, a selection of images from [97] have been inserted in the collection. A total of 36 motifs have been chosen. The selection of these motifs among the much larger catalogue in [97] has been done randomly, although in this initial stage of our research we choose to sort out the most complex motifs. Instances of these selected patterns show a high intra-class variability in terms of scale, rotation, rigid and elastic deformations when observed in the pictorial reference corpus. Some of the shapes exhibits a low inter-class variability (this holds in particular for simple shapes), explaining for instance some resulting ambiguities of the classification. The selected images are in a way the "canonical" reference for each of the decorative motifs to recognize. To use as a reference only the clean and canonical drawing prepared by an expert is too demanding for any shape recognition algorithm. Following a general praxis in the shape recognition community we hence enriched the database with a number of variations for each decorative motif. These variations have been artificially obtained applying the following transformations: rotation in clockwise sense of about 33 degrees, perspective camera distortion, where camera orientation has been assigned at random, and random warping. Examples of the shape included in the database are reported in Fig. 6.11, and an example of variations over a given shape is reported in Fig. 6.12. Eventually the database is made of 144 shapes. Those shapes are stored in binary images of $500 \times 500$ pixels, which are processed for contour extraction in Matlab [112, 113].

The resulting contours are resampled and for each shape a vector of 100 points is obtained. Using a higher sampling rate does not improve the performance of the system, perhaps because a denser sampling of a contour tends to preserve some

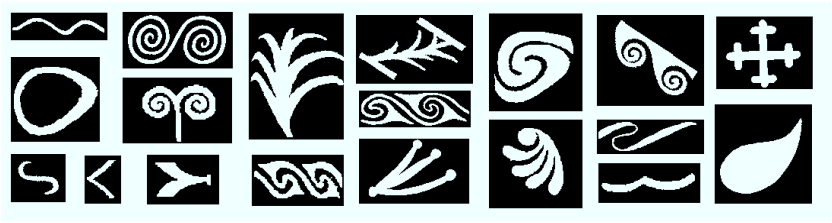


Figure 6.11: Examples of decoration used in our database.

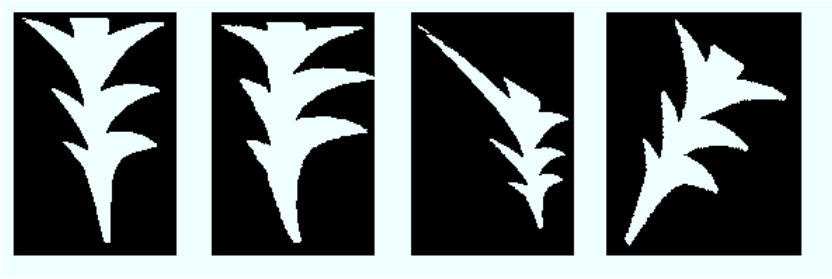


Figure 6.12: Examples of distortion for a single motif in our database. Starting on the left: original, warping, distortion and rotation.

of the noise of the original image making recognition and matching a harder task. The down-sampling has been performed in two different ways, and the results of both selections have been stored in the reference database. The first way to sample 100 points from a closed contour is to compute 100 equally spaced pints along the contour. Notice that the sample points are obtained with cubic interpolations of close by points on the contour. This uniform sampling procedure is applied to be able to compute CBSM distance and to estimate local contour flexibility values. A second non-uniform sampling has been also performed. This second sampling takes 100 points according to contour flexibility values: more points are allocated in the most flexible segments of the shape. This non-uniform sampling procedure is applied to be able to compute contour flexibility distance. Contour flexibility value in each of the sampled point is also stored. In summary our system refers to this collection of $144 \times 2$ point vectors and 144 contour flexibility values to compute similarity distances of queries from the database. Observe that this mathematical representation is quite compact and easily allows the scaling up of

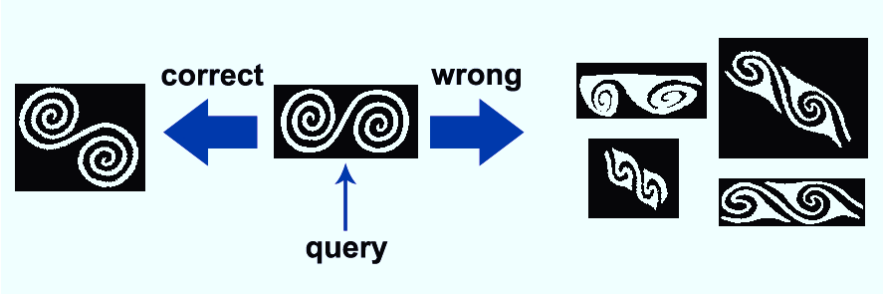
the system up to thousands of shapes.

We choose to check the performance only of the CSBM algorithm and of the Contour Flexibility approach. Indeed CSBM is a refinement of the initial Belongie's proposal and it is safe to assume that it will perform better.

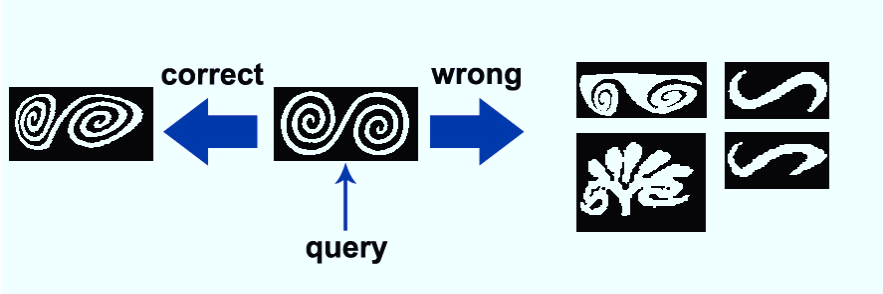
| | % of perfect matches within the four closest shapes | % of majority of correct matches within the four closest shapes |
|---|---|---|
| CBSM | 48% | 72% |
| Contour Flexibility | 75% | 99% |

Table 6.1: Performance of the shape matching procedures.

The experiments have been carried out as follows. Each shape in the database has been, in turn, considered as a query and the best matching shapes (other than the query shapes itself) have been considered as the results of the retrieval operation. In Table 6.1 we report the results obtained with this set of experiments. In particular the first column of the Table indicates the percentage of total queries in which the algorithms finds, within the four most similar shapes, the query shape itself and the three variations of it in the database. The second column in the table indicates the percentage of total queries in which the algorithms finds, within the three most similar shapes (other than the query shape itself) the majority of shapes of the correct class. Each shape of the database is, in turn, taken as a query and the seven shapes with the smallest distances are retrieved from the database, including the query itself. Using Contour Flexibility, on the average 75% of the retrieved shapes is of the same class of the query pattern images. This rate is close to $95 \sim 100\%$ for categories of articulated shapes, whereas for simpler shapes (diamonds, petals, etc) it is significantly lower: in these shapes the flexibility along the boundary varies insignificantly and does not help for object matching, because the global matching reduces to Procrustean distance matching of uniformly sampled contours. On the other hand, Circular Blurred Shape Model has better performances in case of simple shapes, hence a possible remedy to this problem is to analyze the sequence of contour flexibility and to automatically switch to a simpler descriptor for matching (e.g., Circular Blurred Shape Model or Shape Context) which can manage these shapes better than Contour Flexibility. It can




(a)



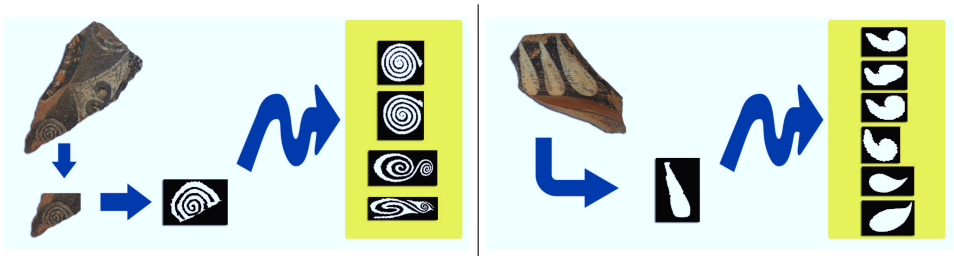

(b)

Figure 6.13: Instance of false positives produced by the system. (a) CF; (b) CBSM

be observed that most of the errors are "near misses", in particulare in case of articulated shapes used as query and using the Contour Flexibility algorithm. An example of these is given in Fig. 6.13.

The results are very encouraging, although the system is far from being error free it provides a valuable help to the expert. When trying to match a novel query an archeologist, in almost all of the cases, has to look only at a very small (4 to 6) reference shapes provided as an answer from the system. This greatly reduces the time that he should otherwise spend to look among a larger set of candidate reference patterns. On the other hand at this stage of development of the system the human expert intervention is still necessary.

To further support our claim about the robustness and the usefulness of our approach we present two real world examples. The queries and the results are shown in Fig. 6.14. The application of the algorithm for the automatic classification of the Kamares style pottery revealed itself particularly efficient for the immediate discrimination of the motifs on the chronological and geographical scales. Basing



(a) (b)

Figure 6.14: Examples of real queries to the database.

upon the Walberg's corpora it's possible to offer two examples of the procedure working. In the first case in Fig. 6.14(a), we have a potsherd with a spiral type motif partially preserved. From the image the algorithm automatically extracts the contour and checking for matches in the available database it suggests three hypothesis of interpretation, the simple j-spiral (Walberg 2.i.1), the running j-spiral (Walberg 2.i.2), even possibly deformed, and the j-spiral with filled angle (Walberg 6.2), again with a certain degree of alteration. The immediate restriction of the range of possible motifs gives to the archaeologists the chance to easily check the available data on the geographical and chronological distribution of the motifs themselves and obtain confirmation for their initial hypothesis. In fact, while the simple J- spirals has a wide diffusion in both Provincial and Palatial areas and in the four chronological phases, the j-spiral with filled angle is present in the Palatial area just in the phase 3 and in the Provincial district exclusively in the phases 2 and 3 and just in the East and East-Central Crete (in [98] pp. 38-39, 41, pll. 28, 31; [99], pp. 48-49, 51, 180, 183). Again in the second example in Fig. 6.14(b), in the potsherd is visible a series of teardrops motifs. After the extraction of the contour, the algorithm suggests two different matches, the spiral derivatives (Walberg 8.33), both the original motif and its altered versions, and the petaloid loops (Walberg 12.i.1). Also in this case, the two motifs have exclusive distribution features. The spiral derivatives are present in the Palatial area in phases 2-4 and are absent in phase 1, while in the Provincial district they occur in all the chronological phases but restricted just to East, East-Central and Central Crete. The petaloid loops have same distribution in the Palaces as the above mentioned derivatives but

in the Provincial areas they are exclusive of phase 1 and 3 also having the same geographical diffusion ([98], pp. 43-44, 48-49, pll. 35, 41; [99], pp. 52-53, 57-58, 183, 188).

## 6.4 Conclusions

In this chapter a pipeline to automatically classify simple Kamares decoration has been presented. In particular, we focused on the shape matching of the pipeline showing how contour flexibility together with Procustean distance, circular blurred shape model or shape context may solve this problem. The proposed system will be of great support in automatically assisting the archeologists in classification operations according time and place of production of Kamares pottery fragments.

This first example of application of Computer Vision and Pattern Recognition techniques to a specific topic of the prehistoric archaeology as the Minoan Kamares pottery opened a completely new field of investigation that in the future can deeply change the approaching of the scholars to iconographical problems. The satisfactory results obtained with the classification of single motifs, even deformed by the preservation of the specimens or altered by the chronological and geographical variability of the production and by the creativity of the Minoan artisans, let expect the extension of this technique also to the decorative syntax of the Kamares style, in order to allow the archaeologist to look forward for the long waited exhaustive study of this so significant feature of the Mediterranean prehistory. The content of this chapter has been published in [114, 115, 116].

# Chapter 7

# Shape Classification through Template Matching

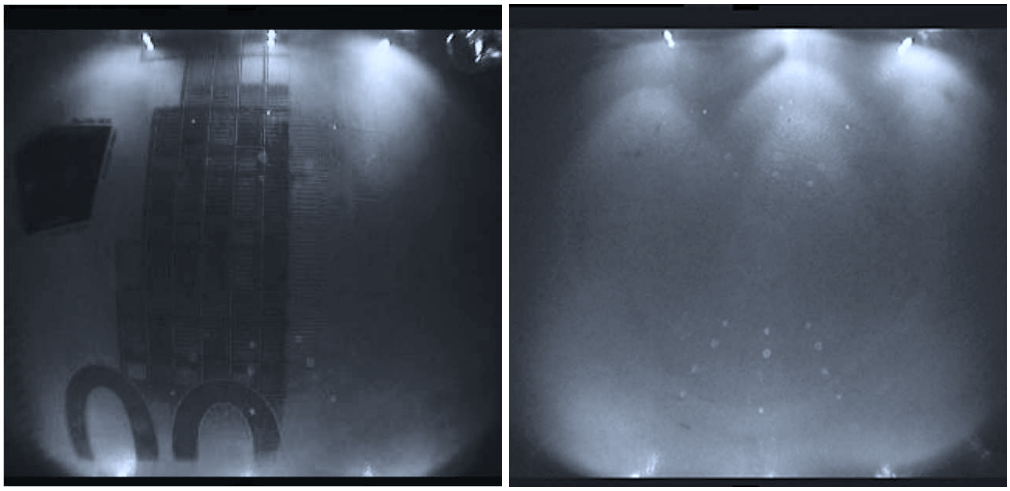
## 7.1 Introduction

The detection of counterfeit banknotes is one of the most important task in billing machines. Usually it is performed by employing several image analysis techniques (transmittance analysis, reflectance analysis, etc) on different light spectra (visible, infrared and ultraviolet). Unfortunately these systems are very expensive and hence they are used only for ATM machines where a high degree of reliability is required since those machines are fully unsupervised by humans. In the last years, cheaper systems for validation and classification of banknotes have been commercialized. They are usually based on motors aimed to let the banknote pass through light emitters and sensors in a dark area. This allows to have a controlled lighting. On the other hand, motors are expensive and the mechanical parts tend to frazzle in few years.

Is well known that, in order to avoid forgeries, the security system of a banknote is typically encoded inside the banknote itself. Different security systems are used depending on the currency, hence an algorithm must be carefully designed for a specific currency. In [117] the authors make use of light transmittance and pattern recognition techniques to recognize the value of banknotes. The pro-

posed solution requires the banknote to pass through light emitters (LED) and receivers (photo transistors) placed in opposite side. Hence, the system requires of a motor or a moving part; moreover the method is not designed to handle counterfeit banknotes. In [118] an algorithm for Korean Won bills classification is described. It makes use of images acquired with visible light. The face value recognition is performed by extracting features in the wavelet domain. Banknote images are first processed with a Sobel filter and then the wavelet feature extraction is performed. Also in this case the problem of detecting counterfeit banknotes is not considered. In the approach proposed in [119] the banknote is segmented into different regions and a classifier for each segmented region is employed. A consensus on the results obtained from the different classified regions provide the final decision. Specifically, genetic algorithms are used for classification of the segmented regions. A neural network and genetic algorithms are exploited in [120, 121] to address the problem of banknote recognition, whereas in [122] the authors propose an Intelligent Banknote Identification System (IBIS) based on neural networks. The system is specifically designed for Turkish Lira and Cyprus Pounds identification. Despite different approaches have been presented in literature, most of them do not take into account Euro Banknotes. All the techniques above could be extended to other currencies (e.g., Euro), but their adaptation to counterfeit detection is not straightforward because, under visible light, many counterfeit banknotes may look like the valid ones. In Euro banknotes there are different features which make forgery difficult: the sheet of paper, the watermarks, the special inks with different behavior in visible, infrared and ultraviolet lights, etc. However, all these features are designed to work in synergy, i.e.a system designed for counterfeit detection should check all features in order to classify a banknote as genuine or counterfeit. Hence, a system which performs only a few check, e.g. concentrating on the banknote appearance under visible light will fail on ad-hoc counterfeit banknotes.

In this chapter we present a complete system, with hardware and software components useful to detect counterfeit banknotes and to recognize their face value. Our system has been developed together with an industrial partner and hence some major constraints were taken into account:



(a) (b)

Figure 7.1: An euro banknote (500 EURO) as seen the infrared camera (a) and an example of the calibration map from a different prototype (b).

1. the proposed solution must be effective in recognising banknotes face values and detecting counterfeits;

2. our system must emphasize ease of use and it must work in quasi-real time;

3. the system must use cheap components in order to have a very low final cost for the buyer.

As consequences, moving parts are not allowed, lighting condition cannot be controlled and the software components have the key role to ensure reliability and low computational time, despite of hardware limitations (e.g. lack of memory and storage, low CPU clock speed and so on). Moreover, only a few features can be checked without requiring specific and expensive extra hardware, hence a careful selection of the features to use is required. In the proposed system an infrared (IR) image of the banknote is used, since it provides good and robust features for counterfeit detection of Euro banknotes.

A prototype of the proposed system has been built and tested on a dataset composed by genuine and fake euro banknotes provided by an Italian Bank. The main components of the proposed system are:

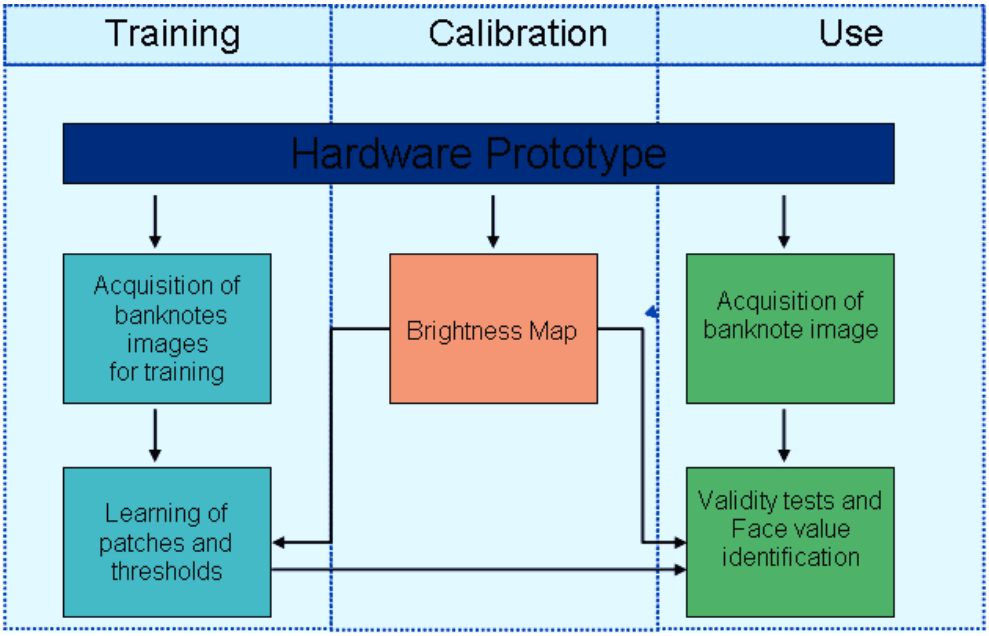



Figure 7.2: Scheme of the proposed system.

- an infrared camera;
- a microprocessor
- a software component.

A glass is placed in the focal plane of the camera in order to acquire a sharp image. The acquired image is then processed through the designed algorithms. The only task which is left to the user is to lean a banknote on the glass and the system provides the information on the validity and the related face value. An example of a Euro banknote, as seen by the IR camera, is shown in the Figure 7.1(a). The two different inks present on the euro banknotes (one reflective to infrared and one absorbing infrared) allow to obtain useful features for the considered task. The reminder of the paper is organized as follows: Section 7.2 summarizes the proposed algorithms for counterfeit detection and currency value recognition of euro banknotes. In Section 7.3 the realized hardware prototype is described, whereas experimental results are presented in Section 7.4. Finally, conclusions and future works are given in Section 7.5.

## 7.2 Proposed methodology

The overall scheme of the developed software for infrared based counterfeit detection and currency value recognition is shown in Figure 7.2. The software consists of three main blocks: Calibration, Training and Use Module.

### 7.2.1 Calibration

Since the employed LED illumination is not spatially uniform (see Figure 7.1(b)), and taking into account its variabilities from prototype to prototype, a calibration phase is required to remove illumination bias. Specifically, in this phase a brightness map is acquired and a compressed copy is stored inside the flash memory. To obtain the brightness map we capture an image of a white sheet of paper under a idealized lighting condition (i.e. dark room, with no source of infrared light). The map is then used by Training and Use Modules to normalize the input banknotes images as following:

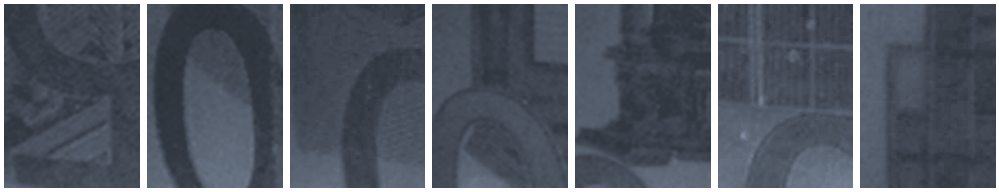
$$NormalisedImage(i,j) = \frac{InputImage(i,j)}{BrightnessMap(i,j)} \tag{7.2.1}$$

### 7.2.2 Training

The Training block is used to learn optimal features of the banknotes (i.e., patches), which will be used determine the validity and the value of each banknotes. The parameters are learned using a large data set of genuine and counterfeit banknotes, which come in several face values $D_i \in D = \{5,\ 10,\ 20,\ 50,\ 100,\ 200,\ 500\}$.

A training dataset of 1000 images has been acquired taking into account typical contexts of use of the apparatus, in which there are very different lighting conditions (e.g. neon, sunlight, incandescent and fluorescent lamps, etc.) and a high degree of misalignment with respect to the proscenium (the system must be robust to slightly translated and or rotated banknotes).

Genuine Euro banknotes are made such that only specific visual features are visible under infrared lighting (see Figure 7.1(a)). In particular banknotes must show darker areas in different zones depending on their value. Those areas always show characteristic patterns. However, there is some overlap among all banknote



(a) 5 Euro (b) 10 Euro (c) 20 Euro (d) 50 Euro (e) 100 Euro (f) 200 Euro (g) 500 Euro

Figure 7.3: A sample set of discriminative patches, one for each possible face value for the Euro currency. The output of the learning process is a set of patches, together with the corresponding positions within the image.

values, for both dark and bright areas, which can be used to develop a reliable validity check.

The learning phase is aimed to find the best area to check banknote validity by using a training dataset composed by genuine and counterfeit euro banknotes.

Given the set of all genuine banknotes $G = \{G_1, G_2, \ldots, G_n\}$ the training block is devoted to search for the largest common dark area $F_g$, and the largest common bright area $T_g$.

Let $F_{gi} = IRSignal(G_i)$ be the infrared highlighted dark area of each $G_i$ , then $F_g = \bigcap_i F_{gi}$.

Let $T_{gi} = G_i \setminus F_{gi}$ be the unresponsive bright area of each $G_i$ , then $T_g = \bigcap_i T_{gi}$. The regions $T_g$ and $F_g$ need to be refined in order to ensure robustness with respect to counterfeit banknotes, which in some case might show a slightly similar infrared response. In a similar way, given the set $C = \{C_1, C_2, \ldots, C_n\}$ of all counterfeit banknotes in our data set, we define $F_{cj} = IRSignal(C_j)$ and $T_{cj} = C_j \setminus F_{cj}$.

Let $C^*$ be the set of all $C_j$ such that $\left(\left(F_{cj} \bigcap F_g \neq \varnothing\right) AND \left(T_{cj} \bigcap T_g \neq \varnothing\right)\right)$. We can finally define $T$ and $F$ considering $C^*$ as follows: $T = T_g \setminus \bigcup_j T_{cj}$ and $F = F_g \setminus \bigcup_j F_{cj}$.

For each banknote $B_k$ in $G \bigcup C$, for each possible threshold $s \in \{0, \ldots, 255\}$, we computed the percentage $percB_{ks}$ of pixels above $s$ in region $T_k$ and the percentage $percD_{ks}$ of pixels below $s$ in regions $F_k$. Using those values we computed optimum $percB^*$ and $percD^*$ thresholds to separate genuine and counterfeit classes.

Template matching is a classical technique for locating the position of a given small subimage inside a large image. It has been frequently used in the applica-

tions of object detection, image registration and pattern recognition. The matching process involves shifting a template image over a search area, measuring the similarity between the template and the current search area, and locating the best match position.

We tested various correlation and histogram based matching with our hardware and software, in order to find a suitable similarity measure. The sum of absolute distances (*SAD*), carried out on pixel gray values, has proven to be of particular interest, since it gives good matching results and has a very simple structure. Let the search area consists of $W \times H$ pixels whereas the template object (i.e. a patch) consists of $w \times h$ opaque pixels. For all $(W-w+1) \times (H-h+1)$ possible positions $(y,x)$ of the template within the image, calculate

$$SAD(y,x) = \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} \left( |I(i+y, j+x) - T(i,j)| \cdot M(i,j) \right); \tag{7.2.2}$$

where $I$ is the pixel value of the search frame, and $T$ is the pixel value of the template object, respectively. For one frame, absolute distance calculations need to be computed. A match is found at a position $(y,x)$ where $SAD(y,x)$ is minimum and also smaller than a certain threshold determined by the user.

To classify the banknote value, during training phase the software determines the discriminative patterns $P_i$ and corresponding search areas for each possible banknote value $D_i$ such that the intra-class *SAD* distance is minimized whereas the inter-class distance is maximized. For some banknote values, which share similar patterns (e.g., 5 Euro and 20 Euro), the process is time consuming.

The output of this process is a set of patches, with the corresponding position, related to the different banknote values. In Figure 7.3 an exemplary set of patches is shown. To take into account slightly translated and/or rotated images, the search area of each patch has the same shape and center coordinates of the patch itself, but it is wider: approximately the search area is 2.5 times the corresponding patch area.

Once the training phase is performed, the selected banknote features (patches, locations and thresholds learned) are stored in the flash memory and used during usage stage (see Figure 7.2) to infer validity and value of input banknote images.

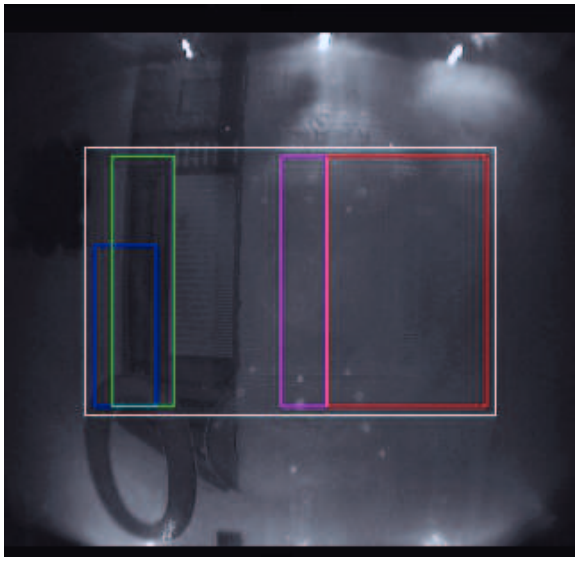


Figure 7.4: A banknote and the set of regions used to test its genuineness. White: ROI. Blue: the threshold to binarize the image is calculated in this area. Green: when seen under infrared light, this area must be dark. Red: when seen under infrared light, this area must be bright and without noticeable patterns. Magenta: additional area used to check the genuineness of some denominations.

### 7.2.3 Banknote authentication

Once the image has been corrected for the nonuniform LED illumination (see Section 7.2.1), a proper threshold needs to be found according to the actual input data. In our experiments we tested the hypothesis that the average grey value (indicated as $MeanRef$ ) of the blue region shown in Figure 7.4 can be robustly used for this purpose.

To check the genuineness of a banknote, the percentage of pixels inside the green region with a grey value below $MeanRef$ ($percLT$) is calculated, together with the percentage of pixels inside the magenta and red regions with a grey value above $MeanRef$ ($percGT$). If $((percLT > percD^*) AND (percGT > percB^*))$ the banknote is classified as genuine.

To identify the face value, the training block provides us with a set of patches (templates related to the different banknotes values) and the corresponding search areas coordinates. Search areas are wider than the patches in order to be robust to

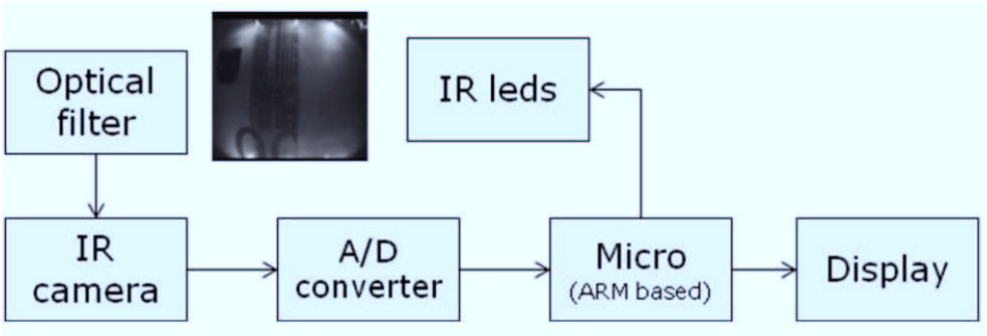



Figure 7.5: The block based schema of the hardware prototype.

a small misalignment. Each template patch is placed at the center of its search area and the correlation measure $\frac{1}{SAD}$ between the pattern itself and the corresponding pixels on the search area is calculated. The procedure then search if a translation around the neighborhood of the current position could increase the correlation. This is performed by moving the pattern position. This step is then repeated until the correlation reaches a local maximum. The pattern with the highest correlation determines the final value to be assigned to the input banknote. Information about genuineness and face value are then sent to the display, in order to inform the user about the results of the banknote analysis. The overall computational speed is of about 1 second for validity and 2 seconds for value recognition using the microprocessor described in the Section 7.3. It should be noted that the proposed software has been designed with the aim to work on a low cost hardware, with the low computational resources (e.g., $32Kbyte$ of Random Access Memory). The hardware components used in our system are described in the following Section.

## 7.3 Hardware Prototype

The hardware prototype has been designed to demonstrate the effectiveness of the proposed framework. The overall system is summarized in the block diagram reported in Figure 7.5.

The Infrared LEDs (IR leds block) illuminate the scene. It is composed by 6 LEDs placed around the proscenium (the area where the banknote is placed).

The illumination is not uniform in the real system. Moreover, it can vary from prototype to prototype. Hence, a calibration is needed to take into account the non uniformity and the performances decay in the system life (see Section 7.2.1). An optical filter is inserted to avoid the external light source to influence the image acquisition system. It is placed on top of the infrared camera (IR camera block) that acquires the image. It is basically a common camera with $CCTV$ output. Since the output is analog, an analog to digital converter ($A/D$ converter block) is used to obtain a digital standardized format (i.e., $CCIR-656$). The digital image is acquired by the microprocessor. In the prototype we have used an $ATMEL\,AT91SAM9XE256$, containing a $ARM926EJ-S^{TM}$ processor, $200MHz$ with $256KByte$ internal high-speed flash memory in which is stored the program. It contains also an Image Sensor Interface ($ISI$) port able to capture video sequences compliant with the standard $ITU-RBT601/656$. The program contains, beside the control logic for the entire subsystems (e.g. IR led, IR camera settings, Display, etc.), also the related algorithms (for both validation and classification) described in the Section 7.2. The prototype has been also equipped with external SRAM memory, since the microprocessor contains only $32KByte$ of internal SRAM. In Figure 7.6 the hardware prototype is shown. In the center there is the $ATMEL$ microprocessor; on the right side there is the $A/D$ converter, in the upper part there is the external SRAM memory, in the lower/right side there is the power supply.

| **True positive** (counterfeit banknote correctly classified) | **True negative** (genuine banknotes correctly classified) |
|---|---|
| 100% | 95.7% |
| **False positive** (counterfeit banknote misclassified) | **False negative** (genuine banknotes misclassified) |
| 0% | 4.3% |

Table 7.1: Genuine/Counterfeit classification

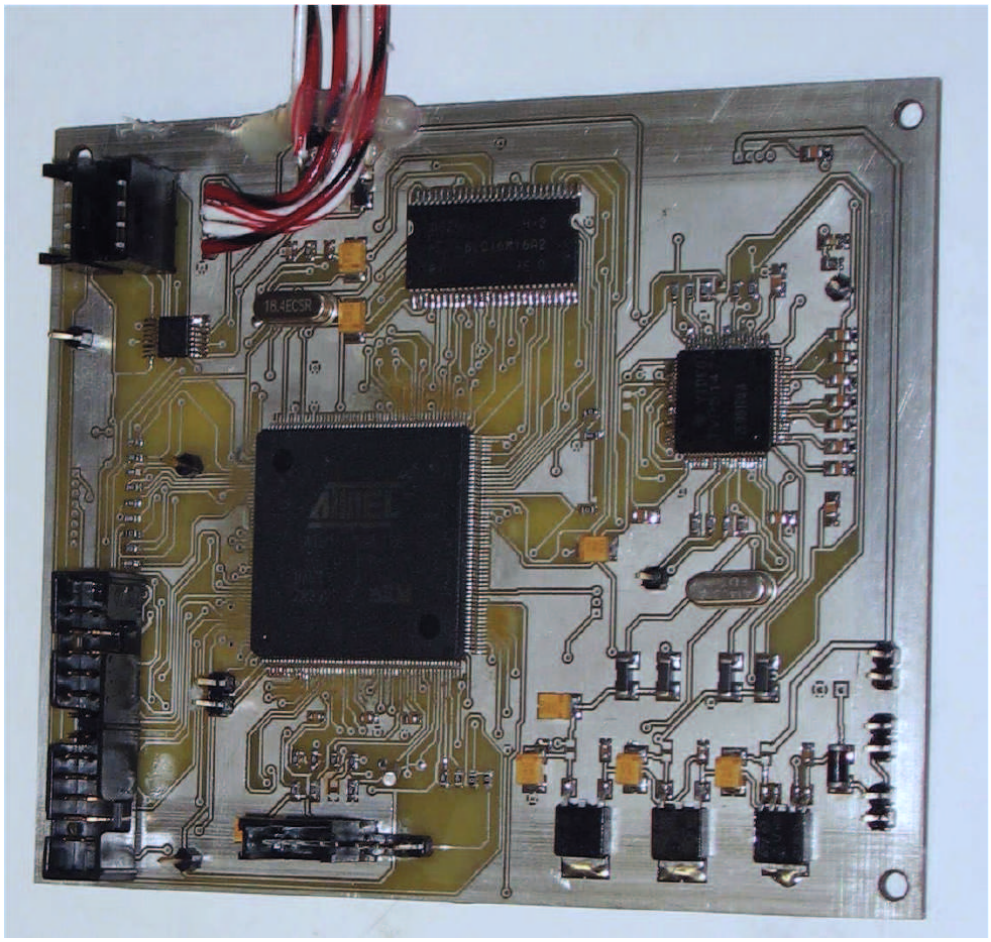


Figure 7.6: The main board containing the microprocessor in the center, the A/D converter on the right side, the external SRAM memory in the upper part, and the power supply in the lower/right side.

## 7.4 Results

To evaluate performances of the proposed technique we used our prototype to acquire a test set of 1750 banknote images, with the same criteria used for the training dataset (i.e. both counterfeit and genuine banknotes have been acquired under several environment lighting conditions, with different illuminants and brightness). In both cases, training and experimental phases, the dataset has been pro-

vided by an Italian Bank.

Acquired counterfeit banknotes include also specimens carefully calibrated to mislead digital counterfeit detectors. To deal with special cases (i.e., fake banknotes provided by the bank), additional procedure has been included in our software. The overall processing time is very close to the base algorithm, since it requires a few additional computations, relative to small areas. Table 7.1 reports the results of the validity assessment. Table 7.2 shows the output of the banknote value classification. Most of the false negatives are observed for images acquired under direct sunlight illumination and for worn banknotes. Wear and tear can be identified as the main cause of misclassified samples for 5 and 10 Euro banknotes, since those banknotes are widely used in everyday life.

## 7.5 Conclusions

In this chapter we have proposed hardware and software modules to detect counterfeit of euro banknotes. Conversely to the state of the art algorithms, the proposed solution makes use of infrared image and low-cost hardware. The proposed system allows recognizing not only the value, but also forgeries. The proposed software components demonstrate the use of very simple shape matching techniques and similarity measures to solve in an effective way an important task for billing machines. he described algorithms are robust to changes in the incident lighting conditions, potentially enabling a widespread diffusion of our system. Thanks to a training phase it is also robust to non-uniformity of the infrared light-

| | **5E** | **10E** | **20E** | **50E** | **100E** | **200E** | **500E** | **Counterfeit** |
|---|---|---|---|---|---|---|---|---|
| **5E** | 88% | 0% | 0% | 0% | 0% | 0% | 0% | 12% |
| **10E** | 1% | 91% | 0% | 0% | 0% | 0% | 0% | 8% |
| **20E** | 0% | 0% | 98% | 0% | 0% | 0% | 0% | 2% |
| **50E** | 0% | 0% | 0% | 99% | 0% | 0% | 0% | 1% |
| **100E** | 0% | 0% | 0% | 0% | 93% | 0% | 0% | 7% |
| **200E** | 0% | 0% | 0% | 0% | 0% | 98% | 1% | 1% |
| **500E** | 0% | 0% | 0% | 0% | 0% | 0% | 95% | 5% |

Table 7.2: Banknote value classification

ing apparatus. The experiments performed on genuine and counterfeit banknotes provided by an Italian Bank show good performances in both validity and value recognition. Future works will be aimed to increase the recognition accuracy and to further reduce the computational time.

# Chapter 8

# Conclusions

In this dissertation the key role of the shape of the objects is emphasized, demonstrating how such a fundamental feature can be efficiently used to solve a wide range of relevant problems in the fields of Computer Vision, Computer Graphics and Pattern Recognition. Typically our input is a set of one or more 2-D image, which convey enough information about the shape if an ad-hoc setup is used for each individual application. The input devices we use range from DSLR cameras to MRI devices, with great differences in the imaging techniques and in the quality of images.

In Chapter 3 we presented a novel technique for estimating surface normals from polarization cues obtained from the Stokes reflectance field captured in just four photographs from a single viewpoint and under a single (unpolarized or circularly polarized) constant spherical incident lighting. We found it to work well for both dielectrics as well as dielectric-metal composites. We demonstrate that both circularly polarized and unpolarized incident lighting can be used to reliably estimate surface normals from observations of the Stokes reflectance field. Accurate shape and appearance estimation is a crucial component in many computer vision and computer graphics applications, which often requires expensive laboratory settings: the main contribution of our work is that we show how our theory can be applied to normal estimation even under uncontrolled outdoor illumination. On the other hand, if the incident lighting is circularly polarized, our method can be used together with the technique described in [54] to obtain a full description

of the BRDF of the object in just four pictures, achieving a long-standing goal of computer graphics: digitally reproducing in a realistic way the appearance of physical objects such as cultural artifacts, consumer products, material samples, and human faces.

In many applications there is a need to deal with classes of objects which are not identical, for example female breasts. A suitable model should maintain the essential characteristics of the class of objects it represents, it should be able to deform to fit a range of examples and it should only generate legal examples. In Chapters 4 and 5 such models are developed to respectively describe the variability observed in 3-D female breast shapes and in the silhouette of cows, as seen from above. In particular we present an embedding of the shape space of the human female breast into a low dimensional linear parameter space, obtained from a set of MRI data processed for noise removal and then analyzed with the PCA technique. Our contributions are a 3-D graphical application to enable the physician to interactively explore the patient data and the potentiality of the proposed modelling technique and a tentative qualitative scale for breast evaluation, helpful to certify the post-operative outcome of breast plastic surgery which is currently a little more than craftsmanship and intuition. We also suggest a new direction for our research, which makes use of a low cost depth camera able to produce accurate 3-D models, potentially enabling accurate shape analysis and avoiding the use of the MRI device.

Our method to parameterize the shape of cows produces a description of the shape to be used for automatic estimation of the health condition of the involved animals (Body Condition Score). The first contribution of this work is a new benchmark dataset useful for research purpose, publicly available through the Internet. BCS estimation systems that work fully automatically (with no user intervention) or at least semi-automatically (with minimal user intervention) are desired to cut down time and costs of the traditional BCS estimation techniques. Moreover, these systems can produce an objective evaluation of BCS in a way that is less invasive for the cows. We proposed both solutions, an automatic and a semi-automatic approach for BCS estimation, able to outperform the state of the art techniques.

In chapter 6 we present a pipeline to automatically classify simple Kamares

decoration, focusing on the shape matching step of an ideal pipeline to assist the archaeologists in classification of Kamares pottery fragments, according time and place of production. This first example of application of Computer Vision and Pattern Recognition techniques to a specific topic of the prehistoric archaeology as the Minoan Kamares pottery opens a completely new field of investigation that in the future can deeply change the approaching of the scholars to iconographical problems. The satisfactory results obtained with the classification of single motifs, even deformed by the preservation of the specimens or altered by the chronological and geographical variability of the production and by the creativity of the Minoan artisans, show how just the 2-D shape contour is able to convey enough information to solve complex problems. Future directions of our research will be devoted to the implementation of the outlined pipeline, starting from the identification of the Region of Interest of each fragment.

In some cases there are specific project constraints which prevent from using an exact solution or state of the art algorithms to obtain a good approximation, e.g. when the hardware employed to execute the algorithms has noticeable limitation or quasi-real time is required. In Chapter 7 we describe an instance of this class of problems, in which a banknote must be efficiently classified in terms of face value and genuineness. We propose a complete solution, which consists in hardware and software modules to detect counterfeit of euro banknotes. Conversely to the state of the art algorithms, the proposed solution makes use of infrared image and low-cost hardware and the software components demonstrate the use of very simple shape matching techniques and similarity measures to solve in an effective way this important task for billing machines. The described algorithms are robust to changes in the incident lighting conditions, potentially enabling a widespread diffusion of our system. Thanks to a training phase it is also robust to non-uniformity of the infrared lighting apparatus. The experiments performed on genuine and counterfeit banknotes provided by the Italian Central Bank show good performances in both validity and value recognition. Future works will be aimed to increase the recognition accuracy and to further reduce the computational time.

# Bibliography


[1] H. G. Barrow and J. M. Tenenbaum, "Recovering intrinsic scene characteristics from images," in *International Conference on Computer Vision Systems*, 1978.

[2] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 1459–1472, Sept. 2005.

[3] R. J. Woodham, "Photometric stereo: A reflectance map technique for determining surface orientation from image intensity," in *Proc. SPIE's 22nd Annual Technical Symposium*, vol. 155, 1978.

[4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. New York, NY, USA: Cambridge University Press, 2 ed., 2003.

[5] Y. Yu, P. Debevec, J. Malik, and T. Hawkins, "Inverse global illumination: recovering reflectance models of real scenes from photographs," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '99, pp. 215–224, 1999.

[6] B. K. Horn, "Shape from shading: a method for obtaining the shape of a smooth opaque object from one view," tech. rep., 1970.

[7] J. Barron and J. Malik, "Shape, albedo, and illumination from a single image of an unknown object," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 334 –341, june 2012.

[8] J. T. Barron and J. Malik, "Color constancy, intrinsic images, and shape estimation," in *Computer Vision ECCV 2012* (A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, eds.), vol. 7575 of *Lecture Notes in Computer Science*, 2012.

[9] M. K. Johnson and E. H. Adelson, "Retrographic sensing for the measurement of surface texture and shape," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 1070–1077, 2009.

[10] J. Ackermann, M. Ritz, A. Stork, and M. Goesele, "Removing the example from photometric stereo by example," in *Proceedings of the ECCV 2010 Workshop on Reconstruction and Modeling of Large-Scale 3D Virtual Environments*, Sept. 2010.

[11] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Shape from interreflections," *International Journal of Computer Vision*, vol. 6, pp. 173–195.

[12] K. Schlüns, "Photometric stereo for non-lambertian surfaces using color information," in *Proceedings of the 5th International Conference on Computer Analysis of Images and Patterns*, CAIP '93, pp. 444–451, 1993.

[13] S. P. Mallick, T. E. Zickler, D. J. Kriegman, and P. N. Belhumeur, "Beyond lambert: Reconstructing specular surfaces using color," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02*, CVPR '05, pp. 619–626, 2005.

[14] S. Barsky and M. Petrou, "The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, pp. 1239–1252, Oct. 2003.

[15] K. Ikeuchi, "Determining surface orientations of specular surfaces by using the photometric stereo method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 3, pp. 661–669, June 1981.

[16] M. A. Halstead, B. A. Barsky, S. A. Klein, and R. B. Mandell, "Reconstructing curved surfaces from specular reflection patterns using spline sur-

face fitting of normals," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '96, pp. 335–342, 1996.

[17] T. Chen, M. Goesele, and H.-P. Seidel, "Mesostructure from specularity," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, pp. 1825–1832, 2006.

[18] R. Basri and D. Jacobs, "Photometric stereo with general, unknown lighting," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 374–381, 2001.

[19] S. Roth and M. J. Black, "Specular flow and the recovery of surface structure," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, pp. 1869–1876, 2006.

[20] Y. Vasilyev, T. Zickler, S. Gortler, and O. Ben-Shahar, "Shape from specular flow: Is one flow enough?," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, 2011.

[21] S. Rahmann and N. Canterakis, "Reconstruction of specular surfaces using polarization imaging," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001.*

[22] G. A. Atkinson and E. R. Hancock, "Multi-view surface reconstruction using polarization," in *ICCV*, pp. 309–316, 2005.

[23] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Transparent surface modeling from a pair of polarization images," *PAMI*, vol. 26, no. 1, pp. 73–82, 2004.

[24] L. B. Wolff, "Surface orientation from two camera stereo with polarizers," in *Proc. SPIE Conf. Optics, Illumination and Image Sensing for Machine Vision IV*, vol. 1194, pp. 287–297, 1989.

[25] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Polarization-based transparent surface modeling from two views," in *ICCV*, p. 1381, 2003.

[26] G. A. Atkinson and E. R. Hancock, "Two-dimensional brdf estimation from polarisation," *Comput. Vis. Image Underst.*, vol. 111, no. 2, pp. 126–141, 2008.

[27] G. A. Atkinson and E. R. Hancock, "Shape estimation using polarization and shading from two views," *PAMI*, vol. 29, no. 11, pp. 2001–2017, 2007.

[28] M. Saito, Y. Sato, K. Ikeuchi, and H. Kashiwagi, "Measurement of surface orientations of transparent objects by use of polarization in highlight," *J. Opt. Soc. Am. A*, vol. 16, no. 9, pp. 2286–2293, 1999.

[29] D. Miyazaki and K. Ikeuchi, "Shape estimation of transparent objects by using inverse polarization ray tracing," *PAMI*, vol. 29, no. 11, pp. 2018–2030, 2007.

[30] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, and P. Debevec, "Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination," in *Rendering Techniques*, pp. 183–194, 2007.

[31] K. Koshikawa, "A polarimetric approach to shape understanding of glossy objects," *Radiometry*, pp. 190–192, 1992.

[32] "Region-based shape descriptor invariant to rotation, scale and translation," *Signal Processing: Image Communication*, vol. 16, no. 12, pp. 87 – 93, 2000.

[33] T. Cootes, C. Taylor, D. H. Cooper, and J. Graham, "Training models of shape from sets of examples," in *In Proc. British Machine Vision Conference*, pp. 9–18, Springer-Verlag, 1992.

[34] J. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, pp. 33–51.

[35] K. Fukunaga and W. L. G. Koontz, "Application of the karhunen-loève expansion to feature selection and ordering," *IEEE Trans. Comput.*, vol. 19, pp. 311–318, Apr. 1970.

[36] B. M. Scassellati, S. Alexopoulos, and M. D. Flickner, "Retrieving images by 2d shape: a comparison of computation methods with human perceptual judgments," pp. 2–14, 1994.

[37] W. J. Rucklidge, "Efficiently locating objects using the hausdorff distance," *International Journal of Computer Vision*, vol. 24, pp. 251–270, 1997.

[38] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 509–522, 2002.

[39] C. Xu, J. Liu, and X. Tang, "2d shape matching by contour flexibility," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 180–186, 2008.

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

[41] I. Biderman, "Human image understanding: Recent research and a theory," *Comput. Vision Graphics Image Processing*, vol. 32, pp. 29–73, 1985.

[42] S. Escalera, A. Fornés, O. Pujol, A. Escudero, and P. Radeva, "Circular blurred shape model for symbol spotting in documents," in *ICIP*, pp. 2005–2008, 2009.

[43] L. B. Wolff and T. E. Boult, "Constraining object features using a polarization reflectance model," *PAMI*, vol. 13, no. 7, pp. 635–657, 1991.

[44] V. Müller, "Elimination of specular surface-reflectance using polarized and unpolarized light," in *ECCV*, 1996.

[45] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in *ACM SIGGRAPH*, 2000.

[46] S. Nayar, X. Fang, and T. Boult, "Removal of Specularities using Color and Polarization," in *CVPR*, 1993.

[47] S. K. Nayar, X.-S. Fang, and T. Boult, "Separation of reflection components using color and polarization," *IJCV*, vol. 21, no. 3, 1997.

[48] S. Umeyama and G. Godin, "Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images," *PAMI*, vol. 26, no. 5, 2004.

[49] A. Ghosh, T. Chen, P. Peers, C. A. Wilson, and P. E. Debevec, "Estimating specular roughness and anisotropy from second order spherical gradient illumination," *Comput. Graph. Forum*, vol. 28, no. 4, pp. 1161–1170, 2009.

[50] A. Ghosh, G. Fyffe, B. Tunwattanapong, J. Busch, X. Yu, and P. Debevec, "Multiview face capture using polarized spherical gradient illumination," in *Proceedings of the 2011 SIGGRAPH Asia Conference*, SA '11, pp. 129:1–129:10, 2011.

[51] L. B. Wolff, "Polarization-based material classification from specular reflection," *PAMI*, vol. 12, no. 11, pp. 1059–1071, 1990.

[52] H. Chen and L. B. Wolff, "Polarization phase-based method for material classification in computer vision," *IJCV*, vol. 28, no. 1, pp. 73–83, 1998.

[53] S. Tominaga and T. Yamamoto, "Metal-dielectric object classification by polarization degree map," in *CVPR*, pp. 1–4, Dec. 2008.

[54] A. Ghosh, T. Chen, P. Peers, C. A. Wilson, and P. Debevec, "Circularly polarized spherical illumination reflectometry," *ACM Trans. Graph.*, vol. 29, pp. 162:1–162:12, December 2010.

[55] O. G. Cula, K. J. Dana, D. K. Pai, and D. Wang, "Polarization multiplexing and demultiplexing for appearance-based modeling," *PAMI*, vol. 29, no. 2, pp. 362–367, 2007.

[56] L. B. Wolff, A. Lundberg, and R. Tang, "Image understanding from thermal emission polarization," in *CVPR*, (Washington, DC, USA), p. 625, IEEE Computer Society, 1998.

[57] D. Miyazaki, R. T. Tan, K. Hara, and K. Ikeuchi, "Polarization-based inverse rendering from a single view," in *ICCV*, p. 982, 2003.

[58] D. Miyazaki and K. Ikeuchi, "Inverse polarization raytracing: Estimating surface shapes of transparent objects," in *CVPR*, pp. 910–917, 2005.

[59] G. C. Guarnera, P. Peers, P. Debevec, and A. Ghosh, "Estimating surface normals from spherical stokes reflectance fields," in *Computer Vision ECCV 2012. Workshops and Demonstrations*, pp. 340–349, 2012.

[60] G. C. Guarnera, P. Peers, P. Debevec, and A. Ghosh, "Estimating specular normals from spherical stokes reflectance fields," in *ACM SIGGRAPH 2012 Talks*, SIGGRAPH '12, pp. 7:1–7:1, 2012.

[61] G. C. Guarnera, P. Peers, P. Debevec, and A. Ghosh, "Estimating specular normals from spherical stokes reflectance fields," in *ACM SIGGRAPH 2012 Posters*, SIGGRAPH '12, pp. 55:1–55:1, 2012.

[62] E. Collett, *Field Guide to Polarization, SPIE Field Guides vol. FG05*. SPIE, 2005.

[63] C. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*. Wiley Science Paperback Series, 1998.

[64] V. V. Tuchin, "Light scattering study of tissues," *Physics-Uspekhi*, vol. 40, no. 5, p. 495, 1997.

[65] S. Nayar and S. Narasimhan, "Assorted Pixels: Multi-Sampled Imaging With Structural Models," in *European Conference on Computer Vision (ECCV)*, vol. IV, pp. 636–652, May 2002.

[66] C. A. Wilson, A. Ghosh, P. Peers, J.-Y. Chiang, J. Busch, and P. Debevec, "Temporal upsampling of performance geometry using photometric alignment," *ACM Trans. Graph.*, vol. 29, pp. 17:1–17:11, April 2010.

[67] C. Clauser, I. Tebbets, B. Bradtmiller, J. McConville, and C. Gordon, *Measurer's Handbook: U.S. Army Anthropometric Survey, 1987-1988*. Natick, Ma: U.S. Army Natick RD&E Center, 1988.

[68] K. Robinette, H. Daanen, and E. Paquet, "The caesar project: a 3-d surface anthropometry survey," in *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on*, pp. 380 –386, 1999.

[69] B. Allen, B. Curless, and Z. Popović, "The space of human body shapes: reconstruction and parameterization from range scans," *ACM Trans. Graph.*, vol. 22, pp. 587–594, July 2003.

[70] H.-Y. Lee, K. Hong, and E. A. Kim, "Measurement protocol of womens nude breasts using a 3d scanning technique," *Applied Ergonomics*, vol. 35, no. 4, pp. 353 – 359, 2004.

[71] G. Catanuto, A. Spano, A. Pennati, E. Riggio, G. M. Farinella, G. Impoco, S. Spoto, G. Gallo, and M. B. Nava, "Experimental methodology for digital breast shape analysis and objective surgical outcome evaluation," *J Plast Reconstr Aesthet Surg*, vol. 61, no. 3, pp. 314–8, 2008.

[72] G. M. Farinella, G. Impoco, G. Gallo, S. Spoto, and G. Catanuto, "Unambiguous analysis of woman breast shape for plastic surgery outcome evaluation," in *Eurographics Italian Chapter Conference'06*, pp. 255–261, 2006.

[73] D. Chen, I. Kakadiaris, M. Miller, R. Loftin, and C. Patrick, "Modeling for plastic and reconstructive breast surgery," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2000* (S. Delp, A. DiGoia, and B. Jaramaz, eds.), vol. 1935 of *Lecture Notes in Computer Science*, pp. 146–162, Springer Berlin / Heidelberg.

[74] G. Gallo, G. C. Guarnera, F. Milanese, D. Modica, G. Catanuto, and F. Pane, "Parametric representation of human breast shapes," in *Proceedings of the 2009 IEEE International Workshop on Medical Measurements and Applications*, MEMEA '09, pp. 26–30, 2009.

[75] G. Gallo, G. C. Guarnera, and G. Catanuto, "Human breast shape analysis using pca," in *BIOSIGNALS 2010 - Proceedings of the Third International Conference on Bio-inspired Systems and Signal Processing, Valencia, Spain, January 20-23, 2010*, pp. 163–167, 2010.

[76] J. Serra, *Image Analysis and Mathematical Morphology*. Orlando, FL, USA: Academic Press, Inc., 1983.

[77] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *SIGGRAPH Comput. Graph.*, vol. 21, pp. 163–169, Aug. 1987.

[78] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, "Depth mapping using projected patterns," 2009.

[79] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '11, pp. 127–136, 2011.

[80] C. Heindl and C. Kopf, "Reconstructme," Jan. 2012.

[81] J. D. Ferguson, D. T. Galligan, and N. Thomsen, "Principal descriptors of body condition score in holstein cows," *Journal of Dairy Science*, vol. 77, no. 9, pp. 2695 – 2703, 1994.

[82] W. R. Ward, "Body condition scoring - technique and application," *Cattle Practice*, vol. 11, pp. 111–116, 2003.

[83] J. Ferguson, G. Azzaro, and G. Licitra, "Body condition assessment using digital images," *Journal of Dairy Science*, vol. 89, no. 10, pp. 3833 – 3841, 2006.

[84] J. Bewley, A. Peacock, O. Lewis, R. Boyce, D. Roberts, M. Coffey, S. Kenyon, and M. Schutz, "Potential for estimation of body condition scores in dairy cattle from digital images," *Journal of Dairy Science*, vol. 91, no. 9, pp. 3439 – 3453, 2008.

[85] I. Halachmi, P. Polak, D. Roberts, and M. Klopcic, "Cow body shape and automation of condition scoring," *Journal of Dairy Science*, vol. 91, no. 11, pp. 4444 – 4451, 2008.

[86] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, pp. 681–685, June 2001.

[87] E. Persoon and K. S. Fu, "Shape discrimination using fourier descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, pp. 388–397, Mar. 1986.

[88] G. Azzaro, M. Caccamo, J. D. Ferguson, S. Battiato, G. M. Farinella, G. C. Guarnera, G. Puglisi, R. Petriglieri, and G. Licitra, "Objective estimation of body condition score by modeling cow body shape from digital images.," *Journal fo Dairy Science*, vol. 94, no. 4, pp. 2126–37, 2011.

[89] S. Battiato, G. M. Farinella, G. C. Guarnera, G. Puglisi, G. Azzaro, and M. Caccamo, "Assessment of cow's body condition score through statistical shape analysis and regression machines," *Journal of Machine Learning Research - Proceedings Track*, vol. 11, pp. 66–73, 2010.

[90] S. Battiato, G. M. Farinella, G. C. Guarnera, G. Puglisi, G. Azzaro, Caccamo, G. Licitra, and J. D. Ferguson, "Estimation of cows body condition score from images," in *International Workshop on Visual observation and analysis of animal and insect behavior (VAIB) 2010 - Satellite event of the 20th IAPR International Conference of Pattern Recognition (ICPR), Turkey, 22 Aug 2010*, pp. 1–4, 2010.

[91] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of Calcutta Mathematical Society*, vol. 35, 1943.

[92] I. Dryden and K. Mardia, *Statistical shape analysis*. Wiley series in probability and statistics, Chichester [u.a.]: Wiley, 1998.

[93] Y. Rathi, S. Dambreville, and A. Tannenbaum, "Statistical shape analysis using kernel pca," *Proc. SPIE 6064, Image Processing: Algorithms and*

*Systems, Neural Networks, and Machine Learning*, pp. 60641B–60641B–8, 2006.

[94] H. Sahbi, "Kernel pca for similarity invariant shape recognition," *Neurocomput.*, vol. 70, pp. 3034–3045, Oct. 2007.

[95] J. Bi and K. P. Bennett, "Regression error characteristic curves," in *ICML'03*, pp. 43–50, 2003.

[96] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, pp. 1452–1458, Nov. 2004.

[97] G. Walberg, *The Kamares style. Overall effects*, vol. 10. Acta Universitatis Upsaliensis. Boreas, Uppsala, 1978.

[98] G. Walberg, *Provincial Middle Minoan pottery*. Mainz, 1983.

[99] G. Walberg, *Kamares. A study of the character of Palatial Middle Minoan pottery*. Göteborg, 1987.

[100] G. Walberg, "The role and individuality of Kamares ware," *in Aegean Archeaology*, vol. 5, pp. 9–18, 2001.

[101] A. Evans, *The Palace of Minos at Knossos*, vol. I-IV. London, 1921-1935.

[102] R. Dawkins and M. Laistner, "The excavation of the Kamares cave in Crete," *in Annual of the British School at Athens*, vol. 19, pp. 1–34, 1912-1913.

[103] N. Momigliano, *Duncan Mackenzie: a Cautious Canny Highlander and the Palace of Minos at Knossos.*, vol. 72. Bulletin of the Institute of Classical Studies, London, 1999.

[104] D. Mackenzie, *The Middle Minoan pottery of Knossos*, vol. I-IV. Journal of Hellenic Studies, 1906.

[105] L. Pernier, *Il palazzo minoico di Festós*, vol. I. 1935.

[106] L. Pernier and L. Banti, *Il palazzo minoico di Festós*, vol. II. 1951.

[107] D. Levi, *Festós e la Civiltá Minoica I, Roma*. 1976.

[108] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.

[109] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*. Gatesmark Publishing, 2009.

[110] G. Moore, "Automatic scanning and computer processes for the quantitative analysis of micrographs and equivalent subjects," pp. 275–326, 1968.

[111] S. Escalera, A. Fornes, O. Pujol, A. Escudero, and P. Radeva, "Circular blurred shape model for symbol spotting in documents," pp. 2005–2008, 2009.

[112] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," in *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pp. 163–169, 1987.

[113] C. Maple, "Geometric design and space planning using the marching squares and marching cube algorithms," in *Geometric Modeling and Graphics, 2003. Proceedings. 2003 International Conference on*, pp. 90 – 95, july 2003.

[114] G. Guarnera, F. Stanco, D. Tanasi, and G. Gallo, "Classification of decorative patterns in kamares pottery," in *In proceedings of SCCG 26th Spring Conference on Computer Graphics*, pp. 20–23, Comenius University, Bratislava, 2010.

[115] S. Filippo, D. Tanasi, G. C. Guarnera, and G. Gallo, "Automatic classification of decorative patterns in the minoan pottery of kamares style," in *Pattern Recognition and Signal Processing in Archaeometry: Mathematical and Computational Solutions for Archaeology* (C. Papaodysseus, ed.), pp. 186–211.

[116] S. Filippo, D. Tanasi, G. C. Guarnera, and G. Gallo, "Classification of decorative patterns in kamares pottery," in *Radamante al computer. Archeologia e informatica nel mondo minoico: lesperienza catanese* (G. Gallo, V. La Rosa, F. Stanco, and D. Tanasi, eds.), pp. 19–50.

[117] A. Hinwood, P. Preston, G. Suaning, and N. Lovell, "Bank note recognition for the vision impaired," *Australasian Physics & Engineering Sciences in Medicine*, vol. 29, pp. 229–233, 2006.

[118] E. Choi, J. Lee, and J. Yoon, "Feature extraction for bank note classification using wavelet transform," in *Proceedings of the 18th International Conference on Pattern Recognition - Volume 02*, ICPR '06, pp. 934–937, 2006.

[119] C. He, M. Girolami, and G. Ross, "Employing optimized combinations of one-class classifiers for automated currency validation," *Pattern Recognition*, vol. 37, no. 6, pp. 1085 – 1096, 2004.

[120] F. Takeda, T. Nishikage, and S. Omatu, "Banknote recognition by means of optimized masks, neural networks and genetic algorithms," *Engineering Applications of Artificial Intelligence*, vol. 12, no. 2, pp. 175 – 184, 1999.

[121] F. Takeda, L. Sakoobunthu, and H. Satou, "Thai banknote recognition using neural network and continues learning by dsp unit," in *Knowledge-Based Intelligent Information and Engineering Systems*, 2003.

[122] A. Khashman and B. Sekeroglu, "Multi-banknote identification using a single neural network," in *Proceedings of the 7th international conference on Advanced Concepts for Intelligent Vision Systems*, ACIVS'05, pp. 123–129, 2005.

[123] M. C. G. G. T. S. Center, "Lambert's cosine law," dec 2012.

[124] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis, "Geometric considerations and nomenclature for reflectance," *National Bureau of Standards Monograph 160*, 1977.

# Appendix A

# Light and Reflectance Models

The visual ability of humans and other animals is the result of the complex interaction of light, eyes and brain. We are able to see because light from an object can move through space and reach our eyes. The objects that we are able to see can be either luminous objects, which generate their own light, or illuminated objects, capable of reflecting light to our eyes. Objects are visible because they reflect light to the eyes of those who look its way: it is only thanks to reflection that objects in our physical world can be seen.

**Specular Reflection.** Light is known to behave in a very predictable manner. If a ray of light could be observed approaching and reflecting off of a flat mirror, then the behavior of the light as it reflects would follow a predictable law known as the law of reflection (Figure A.1). Specular reflection is the mirror-like reflection of light (or of other kinds of wave) from a surface, in which light from a single incoming direction (a ray) is reflected into a single outgoing direction. Such behavior is described by the law of reflection, which states that the direction of incoming light (the incident ray $I$), and the direction of outgoing light reflected (the reflected ray $R$) make the same angle with respect to the surface normal $N$, thus the angle of incidence $\theta_i$ equals the angle of reflection $\theta_r$, and that the incident, normal, and reflected directions are coplanar. To view an image of an object in a mirror, a person (or a camera) must sight along a line at the image location (Figure A.2(a)).

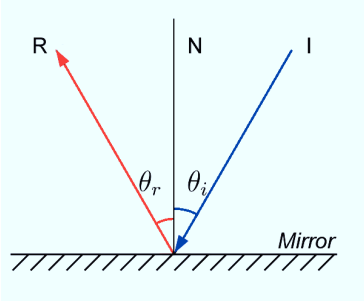
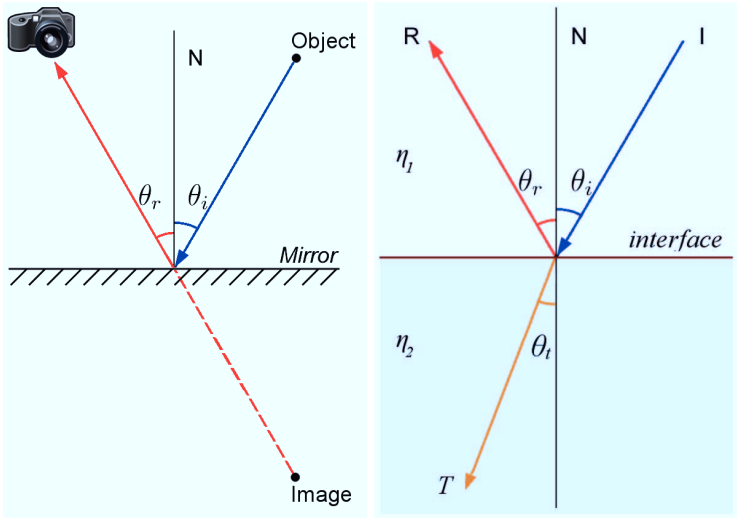



Figure A.1: A diagram illustrating the law of reflection.

**Fresnel Equations.** Even when a surface exhibits only specular reflection with no diffuse reflection, not all of the light is necessarily reflected, since the surface material may absorb some light and depending on the type of material behind the surface, some of the light may be transmitted through the surface (Figure A.2(b). For most interfaces between materials, the fraction of the light that is reflected increases with increasing angle of incidence. The relationship between the angle of incidence $\theta_i$ and the angle of refraction $\theta_t$, when referring to light passing through a boundary between two different isotropic media, is described by the Snell's Law:

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{\eta_2}{\eta_1}, \tag{A.0.1}$$

where $\eta_1$ is the refractive index of the initial medium through which the light propagates, and $\eta_2$ is the index of the other medium. The reflectivity of a surface is the ratio of reflected power to incident power. The reflectivity is a material characteristic, depends on the wavelength, and is related to the refractive index of the material through Fresnel equations, which describe the behaviour of light when moving between media of differing refractive indices. Consider the case of reflections of a mirror-like specular dielectric material. If the other media is the air, which has a index of refraction $\eta_{air} \approx 1$, the exact amount of total reflected radiance is governed by the well-known Fresnel equations, which are a function




(a) (b)

Figure A.2: (a)diagram which shows how the light reflects off the mirror. (b) When the incident light ray $I$ strikes the interface between two media of refractive indices $\eta_1$ and $\eta_2$ , part of the ray is reflected as ray $R$ and part refracted as ray $T$. The angles that the incident, reflected and refracted rays make to the normal of the interface are given as $\theta_i$, $\theta_r$ and $\theta_t$, respectively.

of incident angle $\theta_i$ and the index of refraction $\eta$ of the surface:

$$\mathbf{R}_{\perp} = \left( \frac{\cos\theta_{\mathbf{i}} - \eta\sqrt{\mathbf{1} - \left(\frac{\mathbf{1}}{\eta}\sin\theta_{\mathbf{i}}\right)^{\mathbf{2}}}}{\cos\theta_{\mathbf{i}} + \eta\sqrt{\mathbf{1} - \left(\frac{\mathbf{1}}{\eta}\sin\theta_{\mathbf{i}}\right)^{\mathbf{2}}}} \right)^{\mathbf{2}} \tag{A.0.2}$$

$$\mathbf{R}_{\parallel} = \left( \frac{\sqrt{\mathbf{1} - \left(\frac{\mathbf{1}}{\eta}\sin\theta_{\mathbf{i}}\right)^{\mathbf{2}}} - \eta\cos\theta_{\mathbf{i}}}{\sqrt{\mathbf{1} - \left(\frac{\mathbf{1}}{\eta}\sin\theta_{\mathbf{i}}\right)^{\mathbf{2}}} + \eta\cos\theta_{\mathbf{i}}} \right)^{\mathbf{2}} \tag{A.0.3}$$

**Brewster's Angle.** Brewster's angle (also known as the polarization angle) is an angle of incidence at which light with a particular polarization is perfectly transmitted through a transparent dielectric surface, with no reflection. When unpolarized light is incident at this angle, the light that is reflected from the surface is therefore perfectly polarized. When light encounters a boundary between two me-

dia with different refractive indices, some of it is usually reflected as shown in the figure above. The fraction that is reflected is described by the Fresnel equations, and is dependent upon the incoming light's polarization and angle of incidence. The Fresnel equations predict that light with the $\mathbf{R}_{\parallel}$ polarization will not be reflected if the angle of incidence is

$$\theta_B = \arctan\left(\frac{\eta_2}{\eta_1}\right), \tag{A.0.4}$$

where $\eta_1$ is the refractive index of the initial medium through which the light propagates, and $\eta_2$ is the index of the other medium; the angle defined by this equation is the Brewster's angle. For a glass medium, with $\eta_2 \approx 1.5$ in air, Brewster's angle for visible light is approximately 56; the refractive index for a given medium changes depending on the wavelength of light, hence Brewster's angle will also vary with wavelength.

**Diffuse Reflection.** Ideal diffuse reflectors reflect light according to Lambert's cosine law. Lambert's law states that the reflected energy from a small surface area in a particular direction is proportional to the cosine of the angle between that direction and the surface normalA.3. Lambert's law determines how much of the incoming light energy is reflected in any one direction, which is constant. In other words, the reflected intensity is independent of the viewing direction, differently from the case of specular reflection. The intensity does, however, depend on the light source's orientation relative to the surface, and it is this property that is governed by Lambert's law [123]:

$$I_{diffuse} = k_d \cdot I_{light} \cdot \cos\theta_i, \tag{A.0.5}$$

where $I_{light}$ represents the intensity of the incoming light at the particular wavelength (which determines the light's color) and $k_d$ represents the diffuse reflectivity of the surface at that wavelength. In practice, many common materials exhibit a mixture of specular and diffuse reflection and it is necessary to model both kind of reflections in order to achieve realistic image synthesis.

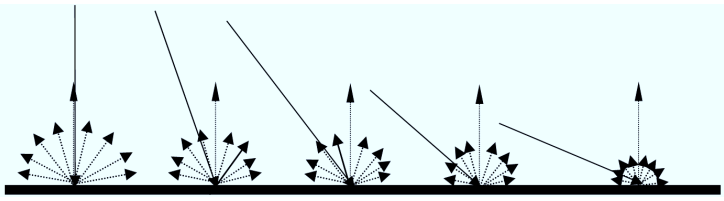


Figure A.3: Lambert's law states that the reflected energy from a small surface area in a particular direction is proportional to the cosine of the angle between that direction and the surface normal.

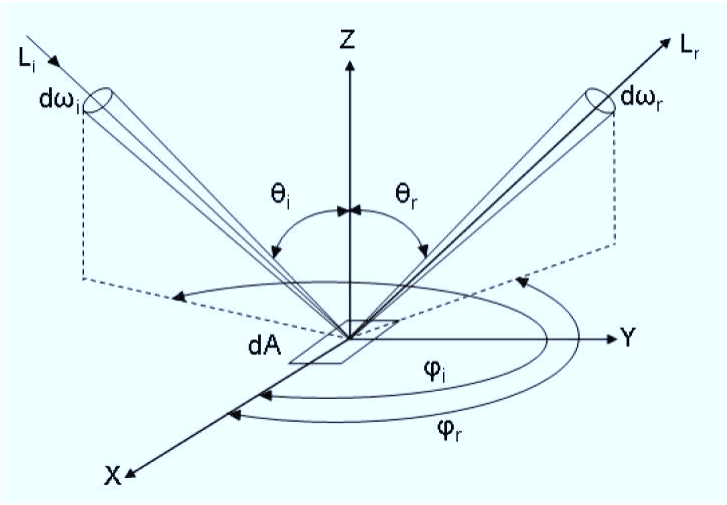



Figure A.4: The BRDF geometry.

**Bidirectional Reflectance Distribution Function.** Accurate descriptions of how light reflects off a surface are a fundamental prerequisite for realistic rendering. Real world materials exhibit characteristic surface reflectance, such as glossy or specular highlights, anisotropy, or retro-reflection, which need to be modeled for visual realism. The surface reflectance of a material is formalized by the notion of the Bidirectional Reflectance Distribution Function (BRDF) as described by Nicodemus et. al. [124], which is a four dimensional function (ignoring the wavelength dependence) describing the response of a differential surface $dA$ in a certain exitant direction $(\theta_r, \phi_r)$, centered within a cone of solid angle $d\omega_r$, to illumination from a certain incident direction $(\theta_i, \phi_i)$, through a solid angle $d\omega_i$, over a

hemisphere of directions (Figure A.4). The BRDF is mathematically defined as the ratio of the directionally reflected radiance $dL_r$ to the directionally incident irradiance $dE_i$:

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{dE_i(\omega_i)} \tag{A.0.6}$$