1 INTRODUCTION

Somewhere, somewhen,
a two headed strategic meeting
on face recognition and matters alike:
t: What about using template matching?
r: Template matching?
t: Yes, a simple technique to compare patterns . . .
r: I'll have a look.

Faces' faces – r's virtual autobiography
ROBERTO BRUNELLI

Go thither; and, with unattainted eye,
Compare her face with some that I shall show,
And I will make thee think thy swan a crow.

Romeo and Juliet
WILLIAM SHAKESPEARE

Computer vision is a wide research field that aims at creating machines that see, not in the limited meaning that they are able to sense the world by optical means, but in the more general meaning that they are able to understand its perceivable structure. Template matching techniques, as now available, have proven to be a very useful tool for this intelligent perception process and have led machines to superhuman performance in tasks such as face recognition. This introductory chapter sets the stage for the rest of the book, where template matching techniques for monochromatic images are discussed.

1.1. Template Matching and Computer Vision

The whole book is dedicated to the problem of template matching in computer vision. While template matching is often considered to be a very basic, limited approach to the most interesting problems of computer vision, it touches upon many old and new techniques in the field.

The two terms *template* and *matching* are used in everyday language, but recalling the definitions more closely related to their technical meaning is useful:

template/pattern

1. Anything fashioned, shaped, or designed to serve as a model from which something is to be made: a model, design, plan, outline.

Template Matching Techniques in Computer Vision: Theory and Practice Roberto Brunelli

2. Something formed after a model or prototype, a copy; a likeness, a similitude.
3. An example, an instance; esp. a typical model or a representative instance.

matching

1. Comparing in respect of similarity; to examine the likeness or difference of.

A template may additionally exhibit some variability: not all of its instances are exactly equal (see Figure 1.1). A simple example of template variability is related to its being corrupted by additive noise. Another important example of variability is due to the different viewpoints from which a single object might be observed. Changes in illumination, imaging sensor, or sensor configuration may also cause significant variations. Yet another form of variability derives from intrinsic variability across physical object instances that causes variability of the corresponding image patterns: consider the many variations of faces, all of them sharing a basic structure, but also exhibiting marked differences. Another important source of variability stems from the temporal evolution of a single object, an interesting example being the mouth during speech. Many tasks of our everyday life require that we identify classes of objects in order to take appropriate actions in spite of the significant variations that these objects may exhibit. The purpose of this book is to present a set of techniques by which a computer can perform some of these identifications. The techniques presented share two common features:

- all of them rely on explicit templates, or on representations by which explicit templates can be generated;
- recognition is performed by matching: images, or image regions, are set in comparison to the stored representative templates and are compared in such a way that their appearance (their image representation) plays an explicit and fundamental role.

The simplest template matching technique used in computer vision is illustrated in Figure 1.2. A planar distribution of light intensity values is transformed into a vector $\boldsymbol{x}$ which can be compared, in a coordinate-wise fashion, to a spatially congruent light distribution similarly represented by vector $\boldsymbol{y}$:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2 = \frac{1}{N} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \tag{1.1}$$

$$s(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{1 + d(\boldsymbol{x}, \boldsymbol{y})}. \tag{1.2}$$

A small value of $d(\boldsymbol{x}, \boldsymbol{y})$ or a high value of $s(\boldsymbol{x}, \boldsymbol{y})$ is indicative of pattern similarity. A simple variation is obtained by substituting the L_2 norm with the L_p norm:

$$d_p(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^p = \frac{1}{N} \|\boldsymbol{x} - \boldsymbol{y}\|_p^p. \tag{1.3}$$

If $\boldsymbol{x}$ is representative of our template, we search for other instances of it by superposing it on other images, or portions thereof, searching for the locations of lowest distance $d(\boldsymbol{x}, \boldsymbol{y})$ (or highest similarity $s(\boldsymbol{x}, \boldsymbol{y})$).

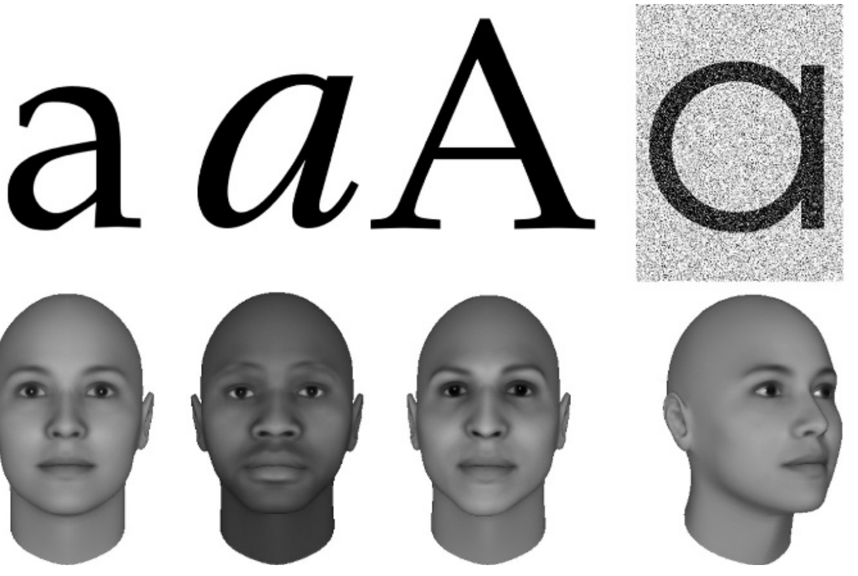

Figure 1.1. Templates from two very common classes: characters and ‘characters’, i.e. faces. Both classes exhibit intrinsic variability and can appear corrupted by noise.

The book shows how this simple template matching technique can be extended to become a flexible and powerful tool supporting the development of sophisticated computer vision systems, such as face recognition systems.

While not a face recognition book, its many examples are related to automated face perception. The main reason for the bias is certainly the background of the author, but there are at least three valid reasons for which face recognition is a valid test bed for template matching techniques. The first one is the widespread interest in the development of high-performing face recognition systems for security applications and for the development of novel services. The second, related reason is that, over the last 20 years, the task has become very popular and it has seen a significant research effort. This has resulted in the development of many algorithms, most of them of the template matching type, providing material for the book. The third reason is that face recognition and facial expression interpretation are two tasks where human performance is considered to be flawless and key to social human behavior. Psychophysical experiments and the evolution of matching techniques have shown that human performance is not flawless and that machines can, sometimes, achieve super human performance.

1.2. The Book

A modern approach to template matching in computer vision touches upon many aspects, from imaging, the very first step in getting the templates, to learning techniques that are key to the possibility of developing new systems with minimal human intervention. The chapters present a balanced description of all necessary concepts and techniques, illustrating them with examples taken from face processing tasks.

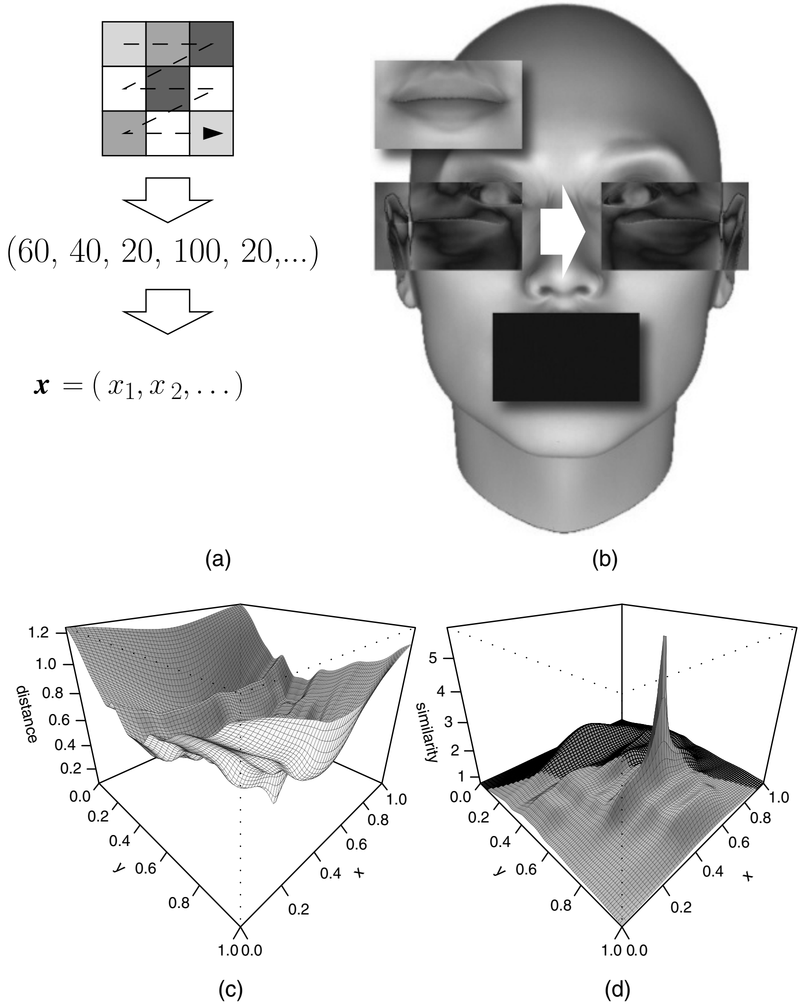

Figure 1.2. The simplest template matching technique: templates are represented as vectors (a) and they are matched by computing their distance in the associated vector space. The template is moved over the image, as a sliding window (b), and the difference between the template and the image is quantified using Equation 1.1, searching for the minimum value (c), or Equation 1.2, searching for the maximum value (d).

A complete description of the imaging process, be it in the case of humans, animals, or computers, would require a (very large) book by itself and we will not attempt it. Chapter 2 discusses some aspects of it that turn out to be critical in the design of artificial vision systems. The basics of how images are created using electromagnetic stimuli and imaging devices are considered. Simple concepts from optics are introduced (including distortion, depth of field, aperture, telecentric lens design) and eyes and digital imaging sensors briefly described. The sampling theorem is presented and its impact on image representation and common image processing operations such as resizing is discussed.

Chapter 3 formally introduces template matching as an hypothesis testing problem. The Bayesian and frequentist approaches are considered with particular emphasis on the Neyman–Pearson paradigm. Matched filters are introduced from a signal processing perspective and simple pattern variability is addressed with the normalized Pearson correlation coefficient. Hypothesis testing often requires the statistical estimation of the parameters characterizing the associated decision function; some subtleties in the estimation of covariance matrices are discussed.

A major issue in template matching is the stability of similarity scores with respect to noise extended to include unmodelled phenomena. Many commonly used estimators suffer from a lack of robustness: small perturbations in the data can drive them towards uninformative values. Chapter 4 addresses the concept of estimator robustness in a technical way, presenting applications of robust statistics to the problem of pattern matching.

Linear correspondence measures like correlation and the sum of squared differences between intensity distributions are fragile. Chapter 5 introduces similarity measures based on the relative ordering of intensity values. These measures have demonstrable robustness both to monotonic image mappings and to the presence of outliers.

While finding a single, well-defined shape is useful, finding instances of a class of shapes can be even more useful. Intraclass variability poses new problems for template matching and several interesting solutions are available. Chapter 6 focuses on the use of projection operators on a one-dimensional space to solve the task. The use of projection operators on multidimensional spaces is covered in Chapter 8.

Finding simple shapes, such as lines and circles, in images may look like a simple task but computational issues coupled with noise and occlusions require some not so naive solutions. In spite of the apparent diversity of lines and areas, it turns out that common approaches to the detection of linear structures can be seen as an efficient implementation of matched filters. Chapter 7 describes how to compute salient image discontinuities and how simple shapes embedded in the resulting map can be located with the Radon/Hough transform.

The representation of images of even moderate resolution requires a significant amount of numeric data, usually one (three) values per pixel if the typical array-based method is adopted. Chapter 8 investigates the possibility of alternative ways of representing iconic data so that a large variety of images can be represented using vectors of reduced dimensionality. Besides significant storage savings, these approaches provide significant benefits to template detection and recognition algorithms, improving their efficiency and effectiveness.

Chapter 9 addresses a couple of cases that are not easily reduced to pattern detection and classification. One such case is the detailed estimation of the parameters of a parametric curve: while Hough/Radon techniques may be sufficient, accurate estimation may benefit from specific approaches. Another important case is the comparison of anatomical structures, such as brain sections, across different individuals or, for the same person, over time. Instead of modeling the variability of the patterns within a class as a static multidimensional manifold, we may focus on the constrained deformation of a parameterized model and measure similarity by the deformation stress.

The drawback of template matching is its high computational cost which has two distinct origins. The first source of complexity is the necessity of using multiple templates to accommodate the variability exhibited by the appearance of complex objects. The second source of complexity is related to the representation of the templates: the higher the resolution, i.e. the number of pixels, the heavier the computational requirements. Besides some computational tricks, Chapter 10 presents more organized, structural ways to improve the speed at which template matching can be performed.

Matching sets of points using techniques targeted at area matching is far from optimal, with regard to both efficiency and effectiveness. Chapter 11 shows how to compare sparse templates, composed by points with no textural properties, using an appropriate distance. Robustness to noise and template deformation as well as computational efficiency are analyzed.

When the probability distribution of the templates is unknown, the design of a classifier becomes more complex and many critical estimation issues surface. Chapter 12 presents basic results upon which two interrelated, powerful classifier design paradigms stand: regularization networks and support vector machines.

Many applications in image processing rely on robust detection of image features and accurate estimation of their parameters. Features may be too numerous to justify the process of deriving a new detector for each one. Chapter 13 exploits the results presented in Chapter 8 to build a single, flexible, and efficient detection mechanism. The complementary aspect of detecting templates considered as a set of separate features will also be addressed and an efficient architecture presented.

Template matching techniques are a key ingredient of many computer vision systems, ranging from quality control to object recognition systems among which biometric identification systems have today a prominent position. Among biometric systems, those based on face recognition have been the subject of extensive research. This popularity is due to many factors, from the non-invasiveness of the technique, to the high expectations due to the widely held belief that human face recognition mechanisms perform flawlessly. Building a face recognition system from the ground up is a complex task and Chapter 14 addresses all the required practical steps: preprocessing issues, feature scoring, the integration of multiple features and modalities, and the final classification stage.

The process of developing a computer vision system for a specific task often requires the interactive exploration of several alternative approaches and variants, preliminary parameter tuning, and more. Appendix A introduces AnImAl, an image processing package written for the R statistical software system. AnImAl, which relies on an algebraic formalization of the concept of image, supports interactive image processing by adding to images a self-documenting capability based on a history mechanism. The documentation facilities of the resulting interactive environment support a practical approach to reproducible research.

A key need in the development of algorithms in computer vision (as in many other fields) is the availability of large datasets for training and testing them. Ideally, datasets should cover the expected variability range of data and be supported by high-quality annotations describing what they represent so that the response of an algorithm can be compared to reality. Gathering large, high-quality datasets is, however, time consuming. An alternative is available for computer vision research: computer graphics systems can be used to generate photorealistic images of complex environments together with supporting ground truth information. Appendix B shows how these systems can be exploited to generate a flexible (and cheap) evaluation environment.

Evaluation of algorithms and systems is a complex task. Appendix C addresses four related questions that are important from a practical and methodological point of view: what is a good response of a template matching system, how can we exploit data to train and at the same time evaluate a classification system, how can we describe in a compact but informative way the performance of a classification system, and, finally, how can we compare multiple classification systems for the same task in order to assess the state of the art of a technology?

The exposition of the main chapter topics is complemented by several intermezzos which provide ancillary material or refresh the memory of useful results. The arguments presented are illustrated with several examples from a very specific computer vision research topic: face detection, recognition, and analysis. There are three main reasons for the very biased choice: the research background of the author, the relevance of the task in the development of biometrics systems, and the possibility that a computational solution to these problems helps understanding (and benefits from the understanding of) the way people do it. Some of the images appearing in the book are generated using the computer graphics techniques described in Appendix B and the packages POV-ray (The Povray Team 2008), POVMan (Krouverk 2005), Aqsis (The Aqsis Team 2007), and FaceGen (Singular Inversions 2008).

Intermezzo 1.1. The definition of intermezzo

intermezzo

1. A brief entertainment between two acts of a play; an entr'acte.
2. A short movement separating the major sections of a lengthy composition or work.

References to relevant literature are not inserted throughout chapter text but are postponed to a final chapter section. Their order of presentation follows the structure of the chapter. All papers on which the chapter is based are listed and pointers to additional material are also provided. References are not meant to be exhaustive, but the interested reader can find additional literature coverage in the cited papers.

1.3. Bibliographical Remarks

This book, while addressing a very specific technique of computer vision, touches upon concepts and methods typical of other fields, from optics to machine learning and its comprehension benefits from readings in these fields.

Among the many books on computer vision, the one by Marr (1982) is perhaps the most fascinating. The book by Horn (1986) still provides an excellent introduction to the fundamental aspects of computer vision, with a careful treatment of image formation. A more recent book is that by Forsyth and Ponce (2002).

Basic notions of probability and statistics can be found in Papoulis (1965). Pattern classification is considered in detail in the books by Fukunaga (1990) and Duda *et al.* (2000). Other important reference books are Moon and Stirling (2000) and Bishop (2007).

A very good, albeit concise, reference for basic mathematical concepts and results is the *Encyclopedic Dictionary of Mathematics* (Mathematical Society of Japan 1993). A wide coverage of numerical techniques is provided by Press *et al.* (2007).

Two interesting papers on computer and human face recognition are those by Sinha *et al.* (2006) and O'Toole *et al.* (2007). The former presents several results on human face analysis processes that may provide guidance for the development of computer vision algorithms. The latter presents some results showing that, at least in some situations, computer vision efforts resulted in algorithms capable of superhuman performance.

References

Bishop C 2007 *Pattern Recognition and Machine Learning*. Springer.

Duda R, Hart P and Stork D 2000 *Pattern Classification* 2nd edn. John Wiley & Sons, Ltd.

Forsyth D and Ponce J 2002 *Computer Vision: A Modern Approach*. Prentice Hall.

Fukunaga K 1990 *Statistical Pattern Recognition* 2nd edn. Academic Press.

Horn B 1986 *Robot Vision*. MIT Press.

Krouverk V 2005 POVMan v1.2. http://www.aetec.ee/fv/vkhomep.nsf/pages/povman2.

Marr D 1982 *Vision*. W.H. Freeman.

Mathematical Society of Japan 1993 *Encyclopedic Dictionary of Mathematics* 2 edn. MIT Press.

Moon T and Stirling W 2000 *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall.

O'Toole A, Phillips P, Fang J, Ayyad J, Penard N and Abdi H 2007 Face recognition algorithms surpass humans matching faces over changes in illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(9), 1642–1646.

Papoulis A 1965 *Probability, Random Variables and Stochastic Processes*. McGraw-Hill.

Press W, Teukolsky S, Vetterling W and Flannery B 2007 *Numerical Recipes* 3rd edn. Cambridge University Press.

Singular Inversions 2008 FaceGen Modeller 3.2. http://www.facegen.com.

Sinha P, Balas B, Ostrovsky Y and Russell R 2006 Face recognition by humans: nineteen results all computer vision researchers should know about. *Proceedings of the IEEE* **94**, 1948–1962. Face Recognition.

The Aqsis Team 2007 Aqsis v1.2. http://www.aqsis.org/.

The Povray Team 2008 The Persistence of Vision Raytracer v3.6. http://www.povray.org/.