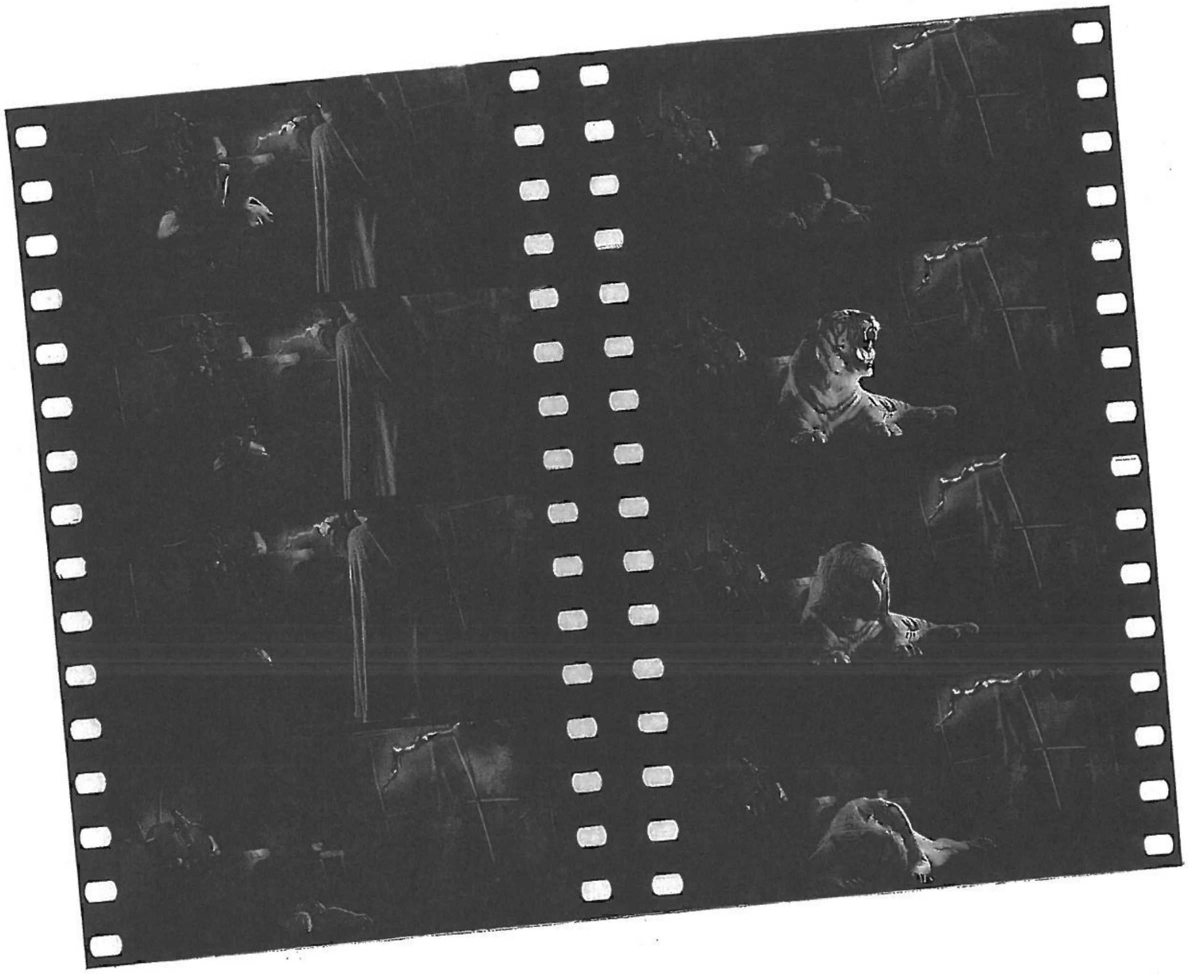
Digital Image Warping

George Wolberg



IEEE Computer Society Press Monograph

VALEO EX. 1029_001

DIGITAL IMAGE WARPING

George Wolberg

Department of Computer Science
Columbia University
New York

IEEE Computer Society Press
Los Alamitos, California

Washington • Brussels • Tokyo

Published by

IEEE Computer Society Press
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1264

Cover layout by Vic Grenrock

Printed in United States of America

IEEE Computer Society Press Order Number 1944
IEEE Catalog Number EH0322-8
ISBN 0-8186-8944-7 (case)
ISBN 0-8186-5944-0 (microfiche)
SAN 264-620X

Additional copies can be ordered from:

IEEE Computer Society Press
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1264

IEEE Computer Society
13, Avenue de l'Aquilon
B-1200 Brussels
BELGIUM

IEEE Computer Society
Ooshima Building
2-19-1 Minami-Aoyama,
Minato-Ku
Tokyo 107, JAPAN

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331

PREFACE

Digital image warping is a growing branch of image processing that deals with geometric transformation techniques. Early interest in this area dates back to the mid-1960s when it was introduced for geometric correction applications in remote sensing. Since that time it has experienced vigorous growth, finding uses in such fields as medical imaging, computer vision, and computer graphics. Although image warping has traditionally been dominated by results from the remote sensing community, it has recently enjoyed a new surge of interest from the computer graphics field. This is largely due to the growing availability of advanced graphics workstations and increasingly powerful computers that make warping a viable tool for image synthesis and special effects. Work in this area has already led to successful market products such as real-time video effects generators for the television industry and cost-effective warping hardware for geometric correction. Current trends indicate that this area will have growing impact on desktop video, a new technology that promises to revolutionize the video production market in much the same way as desktop publishing has altered the way in which people prepare documents.

Digital image warping has benefited greatly from several fields, ranging from early work in remote sensing to recent developments in computer graphics. The scope of these contributions, however, often varies widely owing to different operating conditions and assumptions. This state is reflected in the image processing literature. Despite the fact that image processing is a well-established subject with many textbooks devoted to its study, image warping is generally treated as a peripheral subject with only sparse coverage. Furthermore, these textbooks rarely present image warping concepts as a single body of knowledge. Since the presentations are usually tailored to some narrow readership, different components of the same conceptual framework are emphasized. This has left a noticeable gap in the literature with respect to a unified treatment of digital image warping in a single text. This book attempts to redress this imbalance.

The purpose of this book is to introduce the fundamental concepts of digital image warping and to lay a foundation that can be used as the basis for further study and research in this field. Emphasis is given to the development of a single coherent framework. This serves to unify the terminology, motivation, and contributions of many disciplines that have each contributed in significantly different ways. The coherent framework puts the diverse aspects of this subject into proper perspective. In this manner, the needs and goals of a diverse readership are addressed.

This book is intended to be a practical guide for eclectic scientists and engineers who find themselves in need of implementing warping algorithms and comprehending the underlying concepts. It is also geared to students of image processing who wish to apply their knowledge of that subject to a well-defined application. Special effort has been made to keep prerequisites to a minimum in the hope of presenting a self-contained treatment of this field. Consequently, knowledge of elementary image processing is helpful, although not essential. Furthermore, every effort is made to reinforce the discussion with an intuitive understanding. As a result, only those aspects of supporting theory that are directly relevant to the subject are brought to bear. Interested readers may consult the extensive bibliography for suggested readings that delve further into those areas.

This book originally grew out of a survey paper that I had written on geometric transformation techniques for digital images. During the course of preparing that paper, the large number of disparate sources with potential bearing on digital image warping became strikingly apparent. This writing reflects my goal to consolidate these works into a self-contained central repository. Since digital image warping involves many diverse aspects, from implementation considerations to the mathematical abstractions of sampling and filtering theory, I have attempted to chart a middle path by focusing upon those basic concepts, techniques, and problems that characterize the geometric transformation of digital images, given the inevitable limitations of discrete approximations. The material in this book is thus a delicate balance between theory and practice. The practical segment includes algorithms which the reader may implement. The theory segment is comprised of proofs and formulas derived to motivate the algorithms and to establish a standard of comparison among them. In this manner, theory provides a necessary context in which to understand the goals and limitations of the collection of algorithms presented herein.

The organization of this book closely follows the components of the conceptual framework for digital image warping. Chapter 1 discusses the history of this field and presents a brief overview of the subsequent chapters. A review of common terminology, mathematical preliminaries, and digital image acquisition is presented in Chapter 2. As we shall see later, digital image warping consists of two basic operations: a spatial transformation to define the rearrangement of pixels and interpolation to compute their values. Chapter 3 describes various common formulations for spatial transformations, as well as techniques for inferring them when only a set of correspondence points are known. Chapter 4 provides a review of sampling theory, the mathematical framework used to describe the filtering problems that follow. Chapter 5 describes image resampling, including several common interpolation kernels. They are applied in the discussion of antialiasing in Chapter 6. This chapter demonstrates several approaches used to avoid artifacts that manifest themselves to the discrete nature of digital images. Fast warping techniques based on scanline algorithms are presented in Chapter 7. These results are particularly useful for both hardware and software realizations of geometric transformations. Finally, the main points of the book are recapitulated in Chapter 8. Source code, written in C, is scattered among the chapters and appendices to demonstrate implementation details for various algorithms.

It is often difficult to measure the success of a book. Ultimately, the effectiveness of this text can be judged in two ways. First, the reader should appreciate the difficulties and subtleties in actually warping a digital image. This includes a full understanding of the problems posed due to the discrete nature of digital images, as well as an awareness of the tradeoffs confronting an algorithm designer. There are valuable lessons to be learned in this process. Second, the reader should master the key concepts and techniques that facilitate further research and development. Unlike many other branches of science, students of digital image warping benefit from the direct visual realization of mathematical abstractions and concepts. As a result, readers are fortunate to have images clarify what mathematical notation sometimes obscures. This makes the study of digital image warping a truly fascinating and enjoyable endeavor.

George Wolberg

TABLE OF CONTENTS

1

INTRODUCTION

1.1. BACKGROUND

Digital image warping is a growing branch of image processing that deals with the geometric transformation of digital images. A *geometric transformation* is an operation that redefines the spatial relationship between points in an image. Although image warping often tends to conjure up notions of highly distorted imagery, a warp may range from something as simple as a translation, scale, or rotation, to something as elaborate as a convoluted transformation. Since all warps do, in fact, apply geometric transformations to images, the terms ''warp'' and ''geometric transformation'' are used interchangeably throughout this book.

It is helpful to interpret image warping in terms of the following physical analogy. Imagine printing an image onto a sheet of rubber. Depending on what forces are applied to that sheet, the image may simply appear rotated or scaled, or it may appear wildly distorted, corresponding to the popular notion of a warp. While this example might seem to portray image warping as a playful exercise, image warping does serve an important role in many applied sciences. Over the past twenty years, for instance, image warping has been the subject of considerable attention in remote sensing, medical imaging, computer vision, and computer graphics. It has made its way into many applications, including distortion compensation of imaging sensors, decalibration for image registration, geometrical normalization for image analysis and display, map projection, and texture mapping for image synthesis.

Historically, geometric transformations were first performed on continuous (analog) images using optical systems. Early work in this area is described in [Cutrona 60], a landmark paper on the use of optics to perform transformations. Since then, numerous advances have been made in this field [Horner 87]. Although optical systems offer the distinct advantage of operating at the speed of light, they are limited in control and flexibility. Digital computer systems, on the other hand, resolve these problems and potentially offer more accuracy. Consequently, the algorithms presented in this book deal exclusively with digital (discrete) images.

The earliest work in geometric transformations for digital images stems from the remote sensing field. This area gained attention in the mid-1960s, when the U.S. National Aeronautics and Space Administration (NASA) embarked upon aggressive earth observation programs. Its objective was the acquisition of data for environmental research applicable to earth resource inventory and management. As a result of this initiative, programs such as Landsat and Skylab emerged. In addition, other government agencies were supporting work requiring aerial photographs for terrain mapping and surveillance.

These projects all involved acquiring multi-image sets (i.e., multiple images of the same area taken either at different times or with different sensors). Immediately, the task arises to align each image with every other image in the set so that all corresponding points match. This process is known as *image registration*. Misalignment can occur due to any of the following reasons. First, images may be taken at the same time but acquired from several sensors, each having different distortion properties, e.g., lens aberration. Second, images may be taken from one sensor at different times and at various viewing geometries. Furthermore, sensor motion will give rise to distortion as well.

Geometric transformations were originally introduced to invert (correct) these distortions and to allow the accurate determination of spatial relationships and scale. This requires us to first estimate the distortion model, usually by means of reference points which may be accurately marked or readily identified (e.g., road intersections and land-water interface). In the vast majority of cases, the coordinate transformation representing the distortion is modeled as a bivariate polynomial whose coefficients are obtained by minimizing an error function over the reference points. Usually, a second-order polynomial suffices, accounting for translation, scale, rotation, skew, and pincushion effects. For more local control, affine transformations and piecewise polynomial mapping functions are widely used, with transformation parameters varying from one region to another. See [Haralick 76] for a historical review of early work in remote sensing.

An example of the use of image warping for geometric correction is given in Figs. 1.1 and 1.2. Figure 1.1 shows an example of an image distorted due to viewing geometry. It was recorded after the Viking Lander 2 spacecraft landed on Mars in September 1976. A cylindrical scanner was used to acquire the image. Since the spacecraft landed with an 8° downward tilt, the level horizon appears curved. This problem is corrected in Fig. 1.2, which shows the same image after it was rectified by a transformation designed to remove the tilt distortion.

The methods derived from remote sensing have direct application in other related fields, including medical imaging and computer vision. In medical imaging, for instance, geometric transformations play an important role in image registration and rotation for digital radiology. In this field, images obtained after injection of contrast dye are enhanced by subtracting a mask image taken before the injection. This technique, known as digital subtraction angiography, is subject to distortions due to patient motion. Since motion causes misalignment of the image and its subtraction mask, the resulting produced images are degraded. The quality of these images is improved with transformation algorithms that increase the accuracy of the registration.

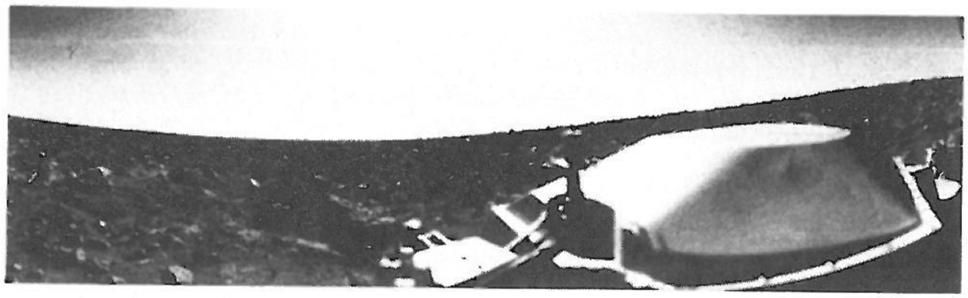


Figure 1.1: Viking Lander 2 image distorted due to downward tilt [Green 89].

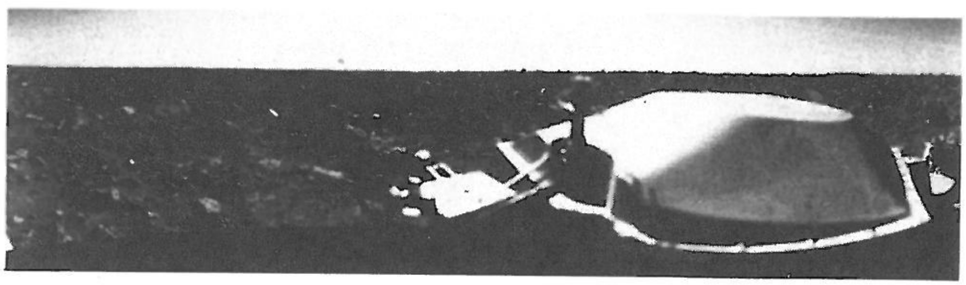


Figure 1.2: Viking Lander 2 image after distortion correction [Green 89].

Image warping is a problem that arises in computer graphics as well. However, in this field the goal is not geometric correction, but rather inducing geometric distortion. Graphics research has developed a distinct repertoire of techniques to deal with this problem. The primary application is texture mapping, a technique to map 2-D images onto 3-D surfaces, and then project them back onto a 2-D viewing screen. Texture mapping has been used with much success in achieving visually rich and complicated imagery. Furthermore, additional sophisticated filtering techniques have been promoted to combat artifacts arising from the severe spatial distortions possible in this application. The thrust of this effort has been directed to the study and design of efficient spatially-varying low-pass filters. Since the remote sensing and medical imaging fields have generally attempted to correct only mild distortions, they have neglected this important area. The design of fast algorithms for filtering fairly general areas remains a great challenge.

Image warping is commonly used in graphics design to create interesting visual effects. For instance, Fig. 1.3 shows a fascinating sequence of warps that depicts a transformation between two faces, a horse and rider, two frogs, and two dancers. Other examples of such applications include the image sequence shown on the front cover, as well as other effects described in [Holzmann 88].

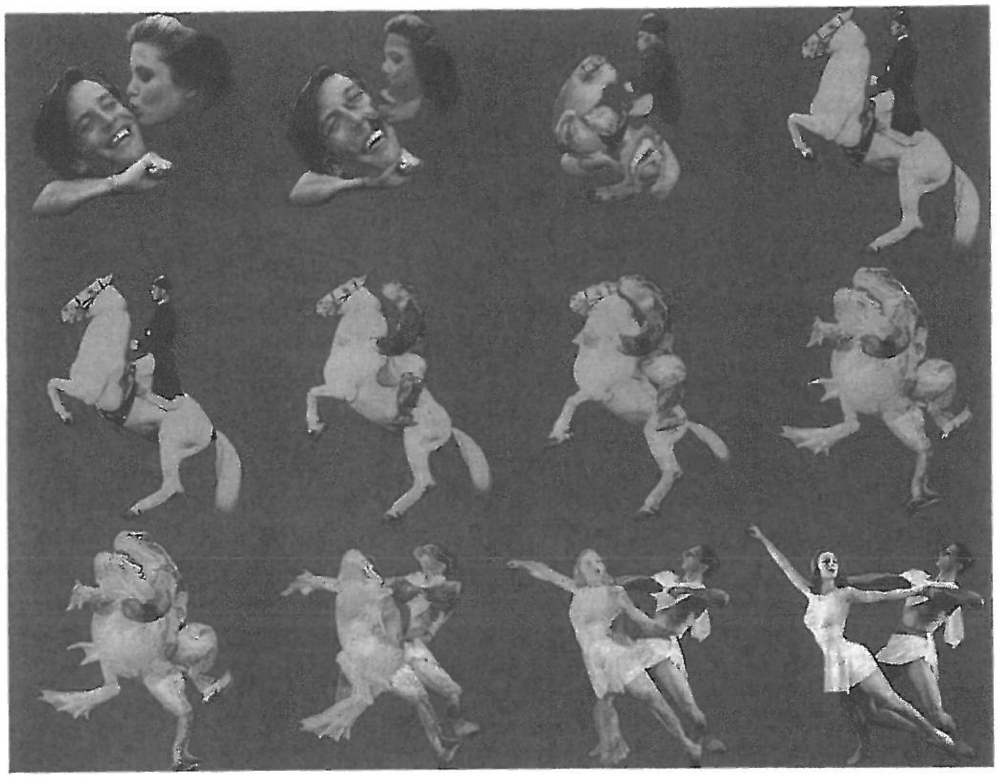


Figure 1.3: Transformation sequence: faces → horse/rider → frogs → dancers.

The continuing development of efficient algorithms for digital image warping has gained impetus from the growing availability of fast and cost-effective digital hardware. The ability to process high resolution imagery has become more feasible with the advent of fast computational elements, high-capacity digital data storage devices, and improved display technology. Consequently, the trend in algorithm design has been towards a more effective match with the implementation technology. This is reflected in the recent surge of warping products that exploit scanline algorithms.

It is instructive at this point to illustrate the relationship between the remote sensing, medical imaging, computer vision, and computer graphics fields since they all have ties to image warping. As stated earlier, image warping is a subset of image processing. These fields are all connected to image warping insofar as they share a common usage for image processing. Figure 1.4 illustrates these links as they relate to images and mathematical scene descriptions, the two forms of data used by the aforementioned fields.

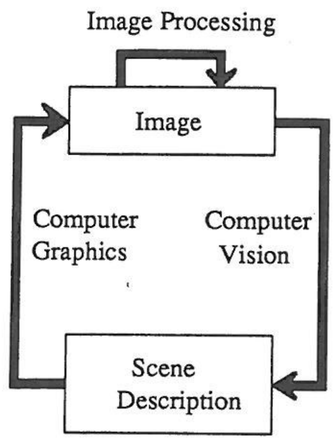


Figure 1.4: Underlying role of image processing [Pavlidis 82].

Consider the transition from a scene description to an image, as shown in Fig. 1.4. This is a function of a renderer in computer graphics. Although image processing is often applied after rendering, as a postprocess, those rendering operations requiring proper filtering actually embed image processing concepts directly. This is true for warping applications in graphics, which manifests itself in the form of texture mapping. As a result, texture mapping is best understood as an image processing problem.

The transition from an input image to an output image is characteristic of image processing. Image warping is thereby considered an image processing task because it takes an input image and applies a geometric transformation to yield an output image. Computer vision and remote sensing, on the other hand, attempt to extract a scene description from an image. They use image registration and geometric correction as preliminary components to pattern recognition. Therefore, image warping is common to these fields insofar as they share images which are subject to geometric transformations.

1.2. OVERVIEW

The purpose of this book is to describe the algorithms developed in this field within a consistent and coherent framework. It centers on the three components that comprise all geometric transformations in image warping: spatial transformations, resampling, and antialiasing. Due to the central importance of sampling theory, a review is provided as a preface to the resampling and antialiasing chapters. In addition, a discussion of efficient scanline implementations is given as well. This is of particular importance to practicing scientists and engineers.

In this section, we briefly review the various stages in a geometric transformation. Each stage has received a great deal of attention from a wide community of people in many diverse fields. As a result, the literature is replete with varied terminologies, motivations, and assumptions. A review of geometric transformation techniques, particularly in the context of their numerous applications, is useful for highlighting the common thread that underlies their many forms. Since each stage is the subject of a separate chapter, this review should serve to outline the contents of this book. We begin with some basic concepts in spatial transformations.

1.2.1. Spatial Transformations

The basis of geometric transformations is the mapping of one coordinate system onto another. This is defined by means of a *spatial transformation* — a mapping function that establishes a spatial correspondence between all points in the input and output images. Given a spatial transformation, each point in the output assumes the value of its corresponding point in the input image. The correspondence is found by using the spatial transformation mapping function to project the output point onto the input image.

Depending on the application, spatial transformation mapping functions may take on many different forms. Simple transformations may be specified by analytic expressions including affine, projective, bilinear, and polynomial transformations. More sophisticated mapping functions that are not conveniently expressed in analytic terms can be determined from a sparse lattice of control points for which spatial correspondence is known. This yields a spatial representation in which undefined points are evaluated through interpolation. Indeed, taking this approach to the limit yields a dense grid of control points resembling a 2-D spatial lookup table that may define any arbitrary mapping function.

In computer graphics, for example, the spatial transformation is completely specified by the parameterization of the 3-D object, its position with respect to the 2-D projection plane (i.e., the viewing screen), viewpoint, and center of interest. The objects are usually defined as planar polygons or bicubic patches. Consequently, three coordinate systems are used: 2-D texture space, 3-D object space, and 2-D screen space. The various formulations for spatial transformations, as well as methods to infer them, are discussed in Chapter 3.

1.2.2. Sampling Theory

In the continuous domain, a geometric transformation is fully specified by the spatial transformation. This is due to the fact that an analytic mapping is bijective — one-to-one and onto. However, in our domain of interest, complications are introduced due to the discrete nature of digital images. Undesirable artifacts can arise if we are not careful. Consequently, we turn to sampling theory for a deeper understanding of the problem at hand.

Sampling theory is central to the study of sampled-data systems, e.g., digital image transformations. It lays a firm mathematical foundation for the analysis of sampled signals, offering invaluable insight into the problems and solutions of sampling. It does so by providing an elegant mathematical formulation describing the relationship between a continuous signal and its samples. We use it to resolve the problems of image reconstruction and aliasing. Note that reconstruction is an interpolation procedure applied to the sampled data and that aliasing simply refers to the presence of unreproducibly high frequencies and the resulting artifacts.

Together with defining theoretical limits on the continuous reconstruction of discrete input, sampling theory yields the guidelines for numerically measuring the quality of various proposed filtering techniques. This proves most useful in formally describing reconstruction, aliasing, and the filtering necessary to combat the artifacts that may appear at the output. The fundamentals of sampling theory are reviewed in Chapter 4.

1.2.3. Resampling

Once a spatial transformation is established, and once we accommodate the subtleties of digital filtering, we can proceed to resample the image. First, however, some additional background is in order.

In digital images, the discrete picture elements, or *pixels*, are restricted to lie on a sampling grid, taken to be the integer lattice. The output pixels, now defined to lie on the output sampling grid, are passed through the mapping function generating a new grid used to resample the input. This new resampling grid, unlike the input sampling grid, does not generally coincide with the integer lattice. Rather, the positions of the grid points may take on any of the continuous values assigned by the mapping function.

Since the discrete input is defined only at integer positions, an interpolation stage is introduced to fit a continuous surface through the data samples. The continuous surface may then be sampled at arbitrary positions. This interpolation stage is known as *image reconstruction*. In the literature, the terms "reconstruction" and "interpolation" are used interchangeably. Collectively, image reconstruction followed by sampling is known as *image resampling*.

Image resampling consists of passing the regularly spaced output grid through the spatial transformation, yielding a resampling grid that maps into the input image. Since the input is discrete, image reconstruction is performed to interpolate the continuous input signal from its samples. Sampling the reconstructed signal gives us the values that are assigned to the output pixels.

The accuracy of interpolation has significant impact on the quality of the output image. As a result, many interpolation functions have been studied from the viewpoints of both computational efficiency and approximation quality. Popular interpolation functions include cubic convolution, bilinear, and nearest neighbor. They can exactly reconstruct second-, first-, and zero-degree polynomials, respectively. More expensive and accurate methods include cubic spline interpolation and convolution with a sinc function. Using sampling theory, this last choice can be shown to be the ideal filter. However, it cannot be realized using a finite number of neighboring elements. Consequently, the alternate proposals have been given to offer reasonable approximations. Image resampling and reconstruction are described in Chapter 5.

1.2.4. Aliasing

Through image reconstruction, we have solved the first problem that arises due to operating in the discrete domain — sampling a discrete input. Another problem now arises in evaluating the discrete output. The problem, related to the resampling stage, is described below.

The output image, as described earlier, has been generated by *point sampling* the reconstructed input. Point (or zero-spread) sampling refers to an ideal sampling process in which the value of each sampled point is taken independently of its neighbors. That is, each input point influences one and only one output point.

With point sampling, entire intervals between samples are discarded and their information content is lost. If the input signal is smoothly varying, the lost data is recoverable through interpolation, i.e., reconstruction. This statement is true only when the input is a member of a class of signals for which the interpolation algorithm is designed. However, if the skipped intervals are sufficiently complex, interpolation may be inadequate and the lost data is unrecoverable. The input signal is then said to be *undersampled*, and any attempt at reconstruction gives rise to a condition known as *aliasing*. Aliasing distortions, due to the presence of unreproducibly high spatial frequencies, may surface in the form of jagged edges and moire patterns.

Aliasing artifacts are most evident when the spatial mapping induces large-scale changes. As an example, consider the problem of image magnification and minification. When magnifying an image, each input pixel contributes to many output pixels. This one-to-many mapping requires the reconstructed signal to be densely sampled. Clearly, the resulting image quality is closely tied to the accuracy of the interpolation function used in reconstruction. For instance, high-degree interpolation functions can exactly reconstruct a larger class of signals than low-degree functions. Therefore, if the input is poorly reconstructed, artifacts such as jagged edges become noticeable at the output grid. Note that the computer graphics community often considers jagged edges to be synonymous with aliasing. As we shall see in Chapter 4, this is sometimes a misconception. In this case, for instance, jagged edges are due to inadequate reconstruction, *not* aliasing.

Under magnification, the output contains at least as much information as the input, with the output assigned the values of the densely sampled reconstructed signal. When minifying (i.e., reducing) an image, the opposite is true. The reconstructed signal is sparsely sampled in order to realize the scale reduction. This represents a clear loss of data, where many input samples are actually skipped over in the point sampling. It is here where aliasing is apparent in the form of moire patterns and fictitious low-frequency components. It is related to the problem of mapping many input samples onto a single output pixel. This requires appropriate filtering to properly integrate all the information mapping to that pixel.

The filtering used to counter aliasing is known as *antialiasing*. Its derivation is grounded in the well established principles of sampling theory. Antialiasing typically requires the input to be blurred *before* resampling. This serves to have the sampled points influenced by their discarded neighbors. In this manner, the extent of the artifacts is diminished, but not eliminated.

Completely undistorted sampled output can only be achieved by sampling at a sufficiently high frequency, as dictated by sampling theory. Although adapting the sampling rate is more desirable, physical limitations on the resolution of the output device often prohibit this alternative. Thus, the most common solution to aliasing is smoothing the input prior to sampling.

The well understood principles of sampling theory offer theoretical insight into the problem of aliasing and its solution. However, due to practical limitations in implementing the ideal filters suggested by the theory, a large number of algorithms have been proposed to yield approximate solutions. Chapter 6 details the antialiasing algorithms.

1.2.5. Scanline Algorithms

The underlying theme behind many of the algorithms that only approximate ideal filtering is one recurring consideration: speed. Fast warping techniques are critical for numerous applications. There is a constant struggle in the speed/accuracy tradeoff. As a result, a large body of work in digital image warping has been directed towards optimizing special cases to obtain major performance gains. In particular, the use of scanline algorithms has reduced complexity and processing time. Scanline algorithms are often based on separable geometric transformations. They reduce 2-D problems into a sequence of 1-D (scanline) resampling problems. This makes them amenable to streamline processing and allows them to be implemented with conventional hardware. Scanline algorithms have been shown to be useful for affine and perspective transformations, as well as for mappings onto bilinear, biquadratic, bicubic, and superquadric patches. Recent work has also shown how it may be extended to realize arbitrary spatial transformations. The dramatic developments due to scanline algorithms are described in Chapter 7.

1.3. CONCEPTUAL LAYOUT

Figure 1.5 shows the relationship between the various stages in a geometric transformation. It is by no means a strict recipe for the order in which warping is achieved. Instead, the purpose of this figure is to convey a conceptual layout, and to serve as a roadmap for this book.

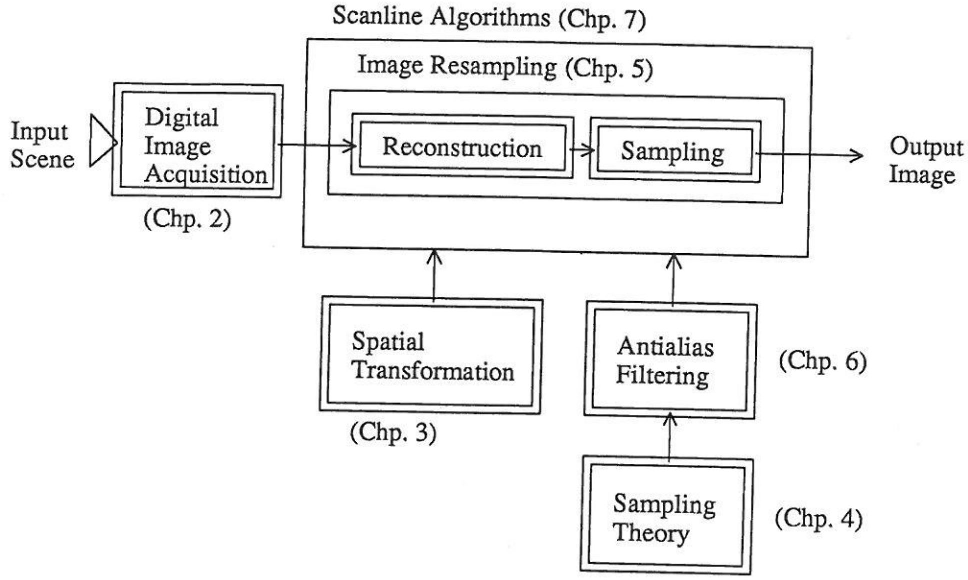


Figure 1.5: Conceptual layout.

An image is first acquired by a digital image acquisition system. It then passes through the image resampling stage, consisting of a reconstruction substage to compute a continuous image and a sampling substage that samples it at any desired location. The exact positions at which resampling occurs is defined by the spatial transformation. The output image is obtained once image resampling is completed.

In order to avoid artifacts in the output, the resampling stage must abide by the principles of digital filtering. Antialias filtering is introduced for this purpose. It serves to process the image so that artifacts due to undersampling are mitigated. The theory and justification for this filtering is derived from sampling theory. In practice, image resampling and digital filtering are collapsed into efficient algorithms which are tightly coupled. As a result, the stages that contribute to image resampling are depicted as being integrated into scanline algorithms.