# Matching, Archiving and Visualizing Cultural Heritage Artifacts Using Multi-Channel Images

Corey Toler-Franklin

A Dissertation

Presented to the Faculty

of Princeton University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance

by the Department of

Computer Science

Adviser: Szymon Rusinkiewicz

May 2011

# Abstract

Recent advancements in low-cost acquisition technologies have made it more practical to acquire real-world datasets on a large scale. This has lead to a number of computer-based solutions for reassembling, archiving and visualizing cultural heritage artifacts. In this thesis, we combine aspects of these technologies in novel ways and introduce algorithms to improve upon their overall efficiency and robustness. First, we introduce a 2-D acquisition system to address the challenge of acquiring higher resolution color and normal maps for large datasets than those available with 3-D scanning devices. Next, we incorporate our normal maps into a novel multi-cue matching system for reassembling small fragments of artifacts. We then present a non-photorealistic rendering pipeline for illustrating geometrically complex objects using images with multiple channels of information.

State-of-the-art 3-D acquisition systems capture 3-D geometry at archeological sites using affordable, off-the-shelf scanners. Although multiple scans at varying viewpoints are required to assemble a complete model, robust registration and alignment algorithms, as well as new work-flow methodologies, significantly reduce the post-processing time. However, the color and normal maps obtained from these systems lack the subtle sub-millimeter details necessary for careful analysis, and high fidelity documentation. We introduce an algorithm that generates higher resolution normal maps and diffuse reflectance (true color texture), while minimizing acquisition time. Using shape from shading, we compute our normal maps from high resolution color scans of the object taken at four orientations on a 2-D flatbed scanner. A key contribution of our work is a novel calibration process to measure the observed brightness as a function of the surface normal. This calibration is important because the scanner's light is linear (rather than a point), and we cannot solve for the surface normal using the traditional formulation of the Lambertian lighting law. High resolution digital SLR cameras provide alternative solutions when objects are too large or fragile to place on a scanner. However, they require more control over the ambient light in the environment and additional manual effort to continually re-position a hand-held flash.

They lack the high resolutions we obtain from the scanner.

Several projects have been explored to leverage these newly acquired datasets for digital reassembly, and have proven successful in some domains. However, current matching algorithms do not perform well when artifacts have deteriorated over many years. One limitation is their reliance on previous acquisition methods that do not capture fine surface details. These details are often important matching cues when features such as color, 2-D contours or 3-D geometry are no longer reliable. We introduce a set of feature descriptors that are based not only on color and shape, but also normal maps with a high data quality. Rather than rely exclusively on one form of data, we use machine-learning techniques to combine descriptors in a multi-cue matching framework. We have tested our system on three datasets of fresco fragments: Theran Frescoes from the site of Akrotiri, Greece; Roman frescoes from Kerkrade in the Netherlands; and a Synthetic fresco created by conservators in a style similar to Akrotiri frescoes. We demonstrate that multi-cue matching using different subsets of features leads to different tradeoffs between efficiency and effectiveness. We observe that individual feature performance varies from dataset to dataset and discuss the implications of feature importance for matching in this domain. Our results show good retrieval performance, significantly improving upon the match prediction rate of state-of the-art 3-D matching algorithms.

The Illustrative depictions found in biology or medical textbooks are one possible method of archiving and distributing historic information. Using a datatype that stores both color and normals, RGBN images, we develop 2-D analogs to 3-D NPR rendering equations. Our approach extends signal processing tools such as scale-space analysis and segmentation for this new data type. We investigate stylized depiction techniques such as toon shading, line drawing and exaggerated shading. By incorporating some 3-D information, we reveal fine details while maintaining the simplicity of a 2-D implementation. Our results achieve levels of detail that are impractical to create with more conventional methods like manual 3-D modeling or 3-D scanning.

# Acknowledgements

Most importantly, I would like to thank my family. No words can express my gratitude and love for my husband, Brian W. Franklin. Your countless sacrifices have allowed me to pursue my dreams. We are partners in all things. To my daughter Eava Franklin, you have brought me the greatest joy. You kept a smile on mommy's face during her most challenging Princeton days. It is all worth it because of you. I am forever indebted to my mother, Olga Toler for her encouragement. I am grateful for your willingness to travel anywhere with me so that Eava is always by my side. To my father Herbert Toler, Thank you for always keeping me focused on my long term goals. I have learned so much from you. To all my family members from the Bahamas, thank you for your support.

To Brian and Eava for their unwavering patience and unconditional support.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

In recent years, researchers have introduced several innovations in data acquisition methods that have pushed the envelope of 3-D scanning technology. These techniques permit the capture of objects that are challenging in size and complexity. As a result, cultural heritage practitioners now benefit from a variety of digital capture methods that are both afford-able and intuitive. This is evidenced by the rise in the number of new projects devoted to accurately documenting and studying collections of artifacts using a digital medium. How-ever, there are several limitations with current acquisition devices. Although some 3-D scanners are capable of generating surfaces with sub-millimeter accuracy (required for sci-entific analysis), they do not perform well on large objects and require time intensive post-processing procedures. These complex systems do not scale to the large volumes of data often found at archeological excavations and are not suitable for non-laboratory environ-ments. Under such conditions, it is common to choose simpler solutions that compromise accuracy for scalability. In addition, the advantages of these acquired 3-D datasets (even when captured at high resolutions) are not fully leveraged in this domain. In particular, they encode more information about the surface structure of an object than color-only rep-resentations. Yet cultural heritage applications often rely on image space features or edge geometry rather than surface orientation. The 3-D information stored in these datasets are

also amenable to a wide variety of rendering algorithms that do not apply to simple photographs. Unfortunately, these algorithms are not fully explored for archival purposes. If processed with efficiency and accuracy, these characteristics of full 3-D datasets have the potential to significantly improve cultural heritage applications that rely on the close visual inspection of surface detail.

In this thesis we improve the quality and robustness of acquired datasets of real-world objects using novel capture methods that are efficient, scalable and cost-effective. Throughout our work, we combine techniques that exhibit the ease and simplicity characteristic of 2-D approaches with results that have the complexity associated with full 3-D models. First, we introduce a 2-D acquisition pipeline that uses a flatbed scanner to generate accurate surface normals from high resolution color scans. We record our datasets as multi-channel images that store both a color and a surface normal at each pixel. These new data types are less sensitive to deterioration than color photographs and easier to acquire with higher accuracy than full 3-D models captured at excavation sites today. Unlike traditional images, however, these new data types express information about surface orientation in the form of $x$, $y$, and $z$ coordinates in Cartesian space and require special handling to maintain their accuracy. We thus extend basic image processing operators to account for a number of situations associated with viewing 3-D geometry at different vantage points (such as foreshortening) to prevent adverse affects like over-smoothing. Next, we develop a set of useful algorithms that operate on these images and leverage the rich sources of information they provide for archaeological study. We validate our claim that systems incorporating multi-channel images are more effective than ones that use color-only or 3-D-only information, by analyzing their performance in two applications: computer-assisted matching of fresco fragments and illustrative depiction of acquired objects.

To provide a foundation for our work, we begin with an overview of methods for acquiring digital representations of real-world objects. First, in Section 1.1, we present a brief introduction to techniques that capture a full 3-D geometric description. One of the pri-

mary goals of this thesis is to contrast the simplicity, robustness and efficiency of our 2-D pipeline with these more complex 3-D examples. We then provide descriptions of methods that recover surface orientation, rather than depth (Section 1.2). These techniques are more closely tied to our approach. Next, in section 1.3, we present two influential cultural heritage projects that incorporate the discussed acquisition methods, *The Pietà Project* and *The Digital Michelangelo Project*. We introduce the key challenges of our matching and visualization problem with a discussion of *The Thera Fresco Project*. We conclude the chapter with a summary of our contributions.

## 1.1   3-D scanning

Several scanning techniques recover full 3-D geometry including the depth and position of points. We focus our discussion on methods that use range scanners to record the depth to the surface of an object for each point in an image. The non-invasive nature of these approaches make them well suited for digitizing fragile objects. Depending on the complexity of the object (the amount of recessed or self occluding regions), several scans may be required to accurately reconstruct a single viewpoint. Once the scans are captured, alignment algorithms are applied to determine the relative position of each scan (previously unknown). The complete 3-D model of the object is typically generated by combining multiple scans taken from multiple viewpoints.

There are two types of range scanners commonly used for acquiring cultural heritage artifacts: time-of-flight scanners and triangulation stereo scanners. Time-of-flight scanners use optical remote sensing technology called *Light Detection And Ranging* (LIDAR). They send a laser pulse toward the target object, and then record the amount of time it takes for the light to bounce off of the surface and return to the sensor. One advantage of time-of-flight range finders is their ability to operate over very long distances (in the order of kilometers). For this reason, they are often preferred for scanning buildings or large statues.

Unfortunately, it is difficult to accurately measure the round-trip travel time of light which travels at a very high speed ($\approx 3x10^8$ m/s). As a result, time-of-flight scanners are only accurate within the range of a few millimeters for certain types of objects.

Triangulation stereo scanners operate on a different principle. Given two cameras, both viewing the same point on an object (but from different positions), the exact position of the point can be computed as the intersection of two camera rays (defined by the line between the camera position and the point). Although these systems are more accurate than time-of-flight approaches (in the order of tens of micrometers) their range is limited. The accuracy of such approaches is closely linked to the size of the viewing volume and the robustness of the calibration methods used to determine the camera parameters. To mitigate the difficulty of finding point correspondences, structured light may be projected into the scene as an aid. Laser triangulation scanners substitute one camera for a laser that is projected at a known angle onto the surface. By observing the laser stripe the remaining camera can more easily detect correspondences and triangulate the model.

## 1.2 Photometric Stereo

Photometric stereo is an acquisition method used to determine the orientation (rather than the depth) of points on a surface. It operates by examining the relationship between image intensity values and the shape of a surface which is determined by the way surfaces reflect light. Unlike traditional stereo approaches, it observes differences in radiance values for corresponding points under varied lighting across different images (rather than geometric displacements). *Shape from shading* methods obtain surface normals given one or more images with known and unknown lighting directions. Photometric stereo is a special case of shape from shading where there are a minimum of three images and the lighting directions are known.

The fundamentals of photometric stereo lie in an understanding of image optics. Before

we explain the mechanics behind these algorithms, recall that the orientation of a surface $f(x, y)$ at a point $p$ can be described by a 3-D vector that is perpendicular to the tangent plane that intersects the surface at $p$. We refer to this orientation as the surface normal vector $n$ at the point $p$. When light from a point source hits a surface at $p$ in an incident direction $l$, several properties determine the fraction of light that will reflect off of the surface in a given direction. They include the optical qualities of the surface material as well as the spatial and spectral qualities of the light. Also considered is the orientation of the light wave's electric field with respect to one period of its direction of propagation (polarization). For highly specular (shiny) objects, light will be reflected in the mirror direction. This will result in a bright spot or specular highlight. For diffuse (Lambertian) objects however, the light falling on the surface is scattered such that it appears equally bright when viewed from any direction (isotropic). According to the *The Lambertian Lighting Law*, the amount of light reflected in any given direction will be proportional to $n \cdot l$. Photometric stereo techniques begin by capturing images from a single viewpoint while varying the direction of illumination between images. The geometry remains unchanged between scenes, and assuming a perfectly diffuse object, $n$ is computed using the *Lambertian Lighting Law* (Section 3.1).

## 1.2.1   Digital SLR Cameras and Light Domes

Traditionally, photometric stereo systems use a high-resolution digital single lens reflex camera (digital SLR camera) and multiple point light sources. The camera is mounted at a fixed position above the object and the lights are positioned at different locations along a hemisphere surrounding the object. Capturing a single image with a digital SLR camera only takes a few milliseconds. During this time, incoming light is directed to an image sensor by a lens for the given exposure time. Other components include a mirror system and a pentaprism that refracts light to an optical view finder for previewing purposes. There are many possibilities for configuring the point light sources. The most basic approaches use a

once subdivided icosahedral template as a guide for manually positioning a light source in several positions relative to the object. Other solutions automate the process with circuitry that controls a set of strobe lights positioned along a supporting dome structure [Malzbender *et al.*, 2001]. In recent years, these camera and light dome structures have become common fixtures in museum conservation labs. The data acquired by such devices may be represented in formats other than per pixel color and normals. In fact, some representations include full lighting models. Examples of other multi-channel image formats are presented in Chapter 2.

## 1.2.2 Linear Light Source Reflectometry

We have reviewed how to capture normals using both calibrated and un-calibrated point light sources. Now we describe *Linear Light Source Reflectometry* as presented by [Gardner *et al.*, 2003]. This technique estimates the spatially-varying reflectance properties of a surface based on its appearance during a single pass of a linear light source (two passes are required for normal estimation). The reflectometry apparatus is inexpensive and simple to construct. During a capture session, the linear light source (a white neon tube) is translated in the $y$ direction at a fixed height above the object using a gantry built from LEGO Mind-Storms (programmable robotics). A laser scanning system consisting of a miniature structured light diode laser with a 60 degree line lens records variations in the surface geometry. The laser stripe is projected parallel to and 38 mm behind the light. Two light passes are taken with the light at different diagonal angles. A digital camera placed at approximately a 55 degree angle of incidence from a perpendicular line to the subject, captures a series of images. Several parameters can be recovered with the system including normals, albedo, height fields, translucency and other high-order reflectance properties. However, we focus our discussion on the method used for normal acquisition.

Using the two light source directions, both components of the normal are computed. The intuition is that when normals point toward the approaching light, the specular com-

ponent is in a earlier time sequence image, and if it is pointing away, it is in a later one. First, reflectance maps are rendered to determine how specular peaks respond to changes in surface normal. For each reflectance table, the specular lobe with the same characteristics as the pixel's observed specular lobe in the sequence image is located. The means of these lobes over the reflectance maps are used to produce a linear function that maps normal perturbation to mean. The mirror function is used for the opposite diagonal angle. To compensate for the fact that this formulation is more accurate on shiny objects, the final normal is a blend of normals estimated by specular reflection and normals obtained by differentiating a laser-scanned displacement map.

One of our methods of obtaining normals is similar to the linear light source reflectometry. We rely on the flatbed scanner's linear light source, which is placed at a fixed angle with respect to the moving CCD. However, because we obtain a single image per scan, instead of a full time-sequence of images, our setup requires different calibration and normal-computation techniques. Hence, we may compute normals from multiple scans with different object rotations. We typically use four scans, rotating the object $90°$ between each. Combining normal and image acquisition exploits the flatbed scanner's high resolution while keeping the number of discrete scanning steps low.

There are several advantages of photometric stereo approaches, such as the one presented above. For applications that require surface orientation, photometric stereo provides a less complicated system for resolving fine details. The camera is in a fixed position so intensive calibration procedures are not required. Most importantly the resulting normal maps have much higher resolutions than those generated from a triangle mesh.

## 1.3   Cultural Heritage Projects

Now that we have presented a technical overview of the key data acquisition technologies, we examine their application in two significant cultural heritage projects, *The Digital*

*Michelangelo Project* [Levoy *et al.*, 2000] and *The Pietà Project* ( [Bernardini *et al.*, 2002]). Both project teams recognized the value of pushing the limits of available scanning capabilities to create accurate digital representations of artifacts for scientific and art historical research. Their contributions have advanced the state-of-the-art in acquisition methods. These are the techniques upon which we build our work. In addition, they have lead to the development of several software algorithms for post processing including algorithms for alignment, registration, mesh simplification, and interactive navigation of large datasets for cultural heritage study. We describe the novelty of each hardware and software approach, and discuss how each satisfied their individual project specifications. By analyzing the complexities of these approaches, we provide the motivation for our design decisions.

### 1.3.1   The Pietà Project

In the case of [Bernardini *et al.*, 2002], a custom built scanner was used to create a 3-D digital model of Michelangelo's Florentine Pietà. The project's goals extended beyond the technical innovations in scanning hardware design to applications for the field of art history. One of the team's objectives was to construct the model to the specifications of an art historian, Professor Jack Wasserman, who would use the model for his research. The complex history of the sculpture, in particular its previous destruction and repair, made it an intriguing subject.

The design of the acquisition system was driven by several constraints. In addition to the usual requirements to minimize cost, limit access to specific hours and protect the artwork, there were physical factors that made some views difficult to capture. The position of the statue in relation to the walls of the room that housed it, the geometric complexity of the sculpture (composed of a group of human figures), and the non-controlled quality of the surrounding lighting conditions dictated a flexible design. The solution was a small, portable system based on a multibaseline stereo system with a customized *Virtuoso ShapeCamera*. This was supplemented with infrastructure for photometric stereo.

During a scanning session, the camera's flash was used to project vertical stripes onto the target, while six black and white digital cameras captured the illuminated surface from different angles and a digital color camera imaged the color texture map. Photometric stereo was incorporated to capture the high-quality geometric detail (such as tool marks) essential for studying Michelangelo's techniques. The statue stands 2.25 meters tall and a scale of 1 to 2 mm was desired to accurately capture surface geometry. Several scan strips of size 20 cm by 20 cm were taken, repositioning the camera for each scan until all views were completed. By using a tripod and camera arrangement instead of a laser range-finder, the team minimized inconsistencies associated with vibrating or moving parts and laborious calibration procedures usually necessary to ensure repeatability in 3-D procedures. For an accurate representation, both geometry and surface reflectance properties were important. In the post-processing stage, stereo algorithms in the camera's software were used to represent the model as a triangle mesh. Once the shape and appearance of each scan's surface was computed, and normal maps generated, the geometric details were aligned and multiple scanned strips merged into a single model.

Several applications were designed to manipulate the model. The intended users of these tools were art historians with little or no experience with computer graphics software and navigation tools. An interactive viewing application permitted the generation of conventional and non-conventional views. Using a frame of reference, users could zoom into regions of a simplified mesh representation, select views, and then display specific views at higher resolutions. The viewer also included rendering and re-lighting capabilities. Tools were provided to determine precise measurements and to modify the statue. The composition of the statue was modified and studied according to information found in additional resources (like historical records and x-rays). In order to edit such a large model (consisting of millions of vertices) a scheme was designed where regions to be edited (removed or replaced) were painted on a scan image and the model was segmented at the painted regions by removing vertices. For the final model, the process was more cumbersome and

tedious. It required bounding boxes to be inserted around the parts that needed editing. The model was then segmented by identifying cracks in the scans.

We apply many of the lessons learned through the successes and limitations of this early example to our system designs. Although revolutionary and successful in its goals, the Pietà project highlighted several short comings with 3-D acquisition systems. First, for the relatively small scan area, the scanner had a long cycle time. Additional time was also spent in processing the acquired images. Complex alignment and registration algorithms were required to generate and merge numerous scans into the final model. Once the model was assembled, mesh simplification was used to create levels of detail that would permit interactivity. Although sufficient for this purpose, the data resolution in the interactive mode was not accurate enough for close examination by a skilled practitioner. Thus, the tradeoffs between usability, data quality and speed of manipulation were never fully resolved. For these reasons, we explore hybrid approaches that use 2-D acquisition methods but incorporate aspects of this work to capture high quality 3-D information. In particular, we use photometric stereo to capture surface detail and focus on sub-millimeter precision to provide added robustness when there is erosion. In our approach, however, we limit processing time by operating in image space. We also aim to leverage the geometric detail captured on the model's surface for art historical study, but instead of providing navigation tools, we focus on revealing subtle geometric features detectable through high-quality normal maps. We also build upon the ideas of this early example when constructing our non-photorealistic rendering pipeline. Many of our visualization techniques are also inspired by the observation that conservators and art historians often move a small light above the surface at a 'raking' angle to emphasize fine surface detail as well as surface curvature.

### 1.3.2 The Digital Michelangelo Project

Another significant project that focused on data acquisition of historic artifacts is *The Digital Michelangelo Project* [Levoy *et al.*, 2000]. Researchers spent one academic year in

Italy scanning several of Michelangelo's statues. In this case, the emphasis was on acquiring 3-D digital models of large fragile objects in non-controlled environments. We focus on their most notable example, *The David*. As with *The Pietà Project*, (and for many of the same reasons), the physical limitations of the environment lead to a customized system. However, the individual components were quite different from those used for the *The Pietá Project*. The main scan head was a customized version of a commercially available full body range scanner manufactured by Cyberware Inc. It contained a laser, range camera, white spotlight, and a digital color camera. Adjustments were made to reduce the scan head's bulkiness. The most interesting structure of the system was a mobile, reconfigurable gantry, capable of extending to a height above the top of David's head (7.5 meters including pedestal). A truss structure and camera mount permitted four degrees of motion. It utilized a vertically moving arm with a pan and tilt assembly, and a camera mount with its own pan and tilt (up to $100\,^{\circ}$) mechanism . To protect the statue, a conservative distance was maintained between the statue and the equipment.

During a scan session, individual range scans, consisting of several concentric curved shells were captured. Color scans were taken intermittently between these. A resolution of 0.25 mm was necessary to capture chisel marks and surface detail for accurate shadowing and re-lighting effects. An extensive post processing pipeline included algorithms for aligning and merging the individual scans into a single model. The first step in the pipeline involved the manual alignment of overlapping scans. An Interative closest points algorithm (ICP was first proposed by [Besl and McKay, 1992]) then performed localized pairwise alignment. A space efficient global relaxation step was handled separately. The color pipeline mapped color images to the vertices of the 3-D model and computed the reflectance at each point. The final model consisted of 2 billion polygons and 7,000 color images. To facilitate interactive viewing, a pyramid approach was adopted where range images were sub-sampled at different levels of detail.

Although the specifics of our project differ from that of *The Digital Michelangelo*

*Project*, we can learn from its implementation and improve upon its inefficiencies. The accuracy of the normals generated by the 3-D scanning process is closely tied to the resolution of each range scan. As mentioned in Section 1.3.1, several factors such as vibrating parts can propagate errors in the scan results. Thus, intensive camera calibration procedures are required to offset issues of instability in the gantry and its moving parts. The amount of the errors introduced is also proportional to the size of the viewing volume and the accuracy of the calibration process itself. Any inconsistencies introduced may materialize as small warps in the final model. In addition, the initial model of the David, despite careful scanning, contained many holes from deep, hard-to-reach recesses. All of the aforementioned defects are challenging to correct and warranted a very complex post-processing pipeline that extended beyond the necessary scan alignment and registration algorithms. Overall the *The Digital Michelangelo Project* required a high level of technical expertise for the design, calibration and operation of the system. Our goal is to develop a system that does not disrupt the workflow of the practitioner and can be operated by non-technical users. In our case, we capture large numbers (in the order of thousands) of relatively small objects rather than a few (in the order of a dozen) very large complex objects. Thus, we are able to focus our solution on a simpler more economical approach capable of scaling to our needs. Our scan unit is a single fragment of an artifact and takes less than a minute to capture. This is relatively fast when compared to the slow process of acquiring multiple thinly sliced scan strips and repositioning a complex system of components between scans.

## 1.4 Case Study: The Thera Fresco Project

In Section 1.3, we presented some of the challenges researchers have faced with acquiring and processing objects of historic importance within the last decade. We now focus on the details of our specific case study *The Thera Fresco Project*. In recent years, the emphasis has shifted from capturing larger, more complex subjects to acquiring high volumes of

data with increased throughput. In fact, we have seen the deployment of more economical, portable 3-D scanners in real-world settings [Brown, 2008]. Although effective at some levels, these systems still do not produce levels of detail adequate for the proper documentation and study of surface information. This is even more evident when objects are eroded and transformed by the elements. Rather than tackle the challenge of improving the mechanics of these 3-D systems directly, we choose to develop simpler complementary systems that are capable of generating higher quality results. Overall, we aim to combine the benefits of efficient 2-D methods with higher resolution results that exhibit some 3-D qualities.

To motivate our work, we introduce a specific excavation site on the island of Thera. Thera is the southernmost island in the Cyclades, an archipelago in the Aegean Sea. As shown in Figure 1.1, one of its more interesting features is its unique shape. Around 1650 B.C. a volcanic eruption destroyed the central part of the island producing the crescent-shaped wedge we see today. It also eliminated the late-Bronze-Age Aegean civilization that once occupied the island. The volcanic ash from the blast was deposited on areas of the remaining land mass. As a

NASA Visible Earth from Wikimedia Commons
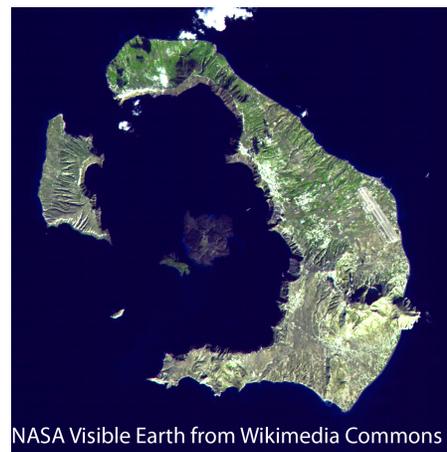
**Figure 1.1:** Thera

result, when archaeologists unearthed the harbor town of Akrotiri, they found it well preserved (comparable to Pompeii). The many objects discovered provided evidence that the inhabitants participated in fishing and agriculture, and traded with other Mediterranean civilizations. However, the most important finds at Akrotiri are extensive wall paintings. They are vibrantly colored, and well preserved with one unfortunate exception; the frescoes were broken into small fragments by an earthquake that preceded the volcanic eruption. Thus, we focus our efforts on the acquisition, matching and archiving of these fragments.

The initial Akrotiri excavation began in 1967 and was led by Professor Spyridon Marinatos under the aegis of the Greek Archaeological Society. In 1974, after the death of Marinatos, the project was taken over by Professor Christos Doumas, who still leads the team today. Presently, much of the effort on the site involves the cleaning, stablization and manual reassembly of these fresco fragments. For a more detailed description of the excavation site, conservation methods and archaeological contexts refer to *The Wall-Paintings of Thera* [Doumas, 1992] and *Santorini: A Guide to the Island and its Archaeological Treasures* [Doumas, 1999].

Our work in developing an acquisition, matching and documentation pipeline is part of a larger collaboration with Professor Doumas, Professor Andreas Vlachopoulos and the conservators at the Akrotiri Laboratory of Wall Paintings. It also includes other researchers at Princeton University, The Katholieke Universiteit Leuven and the University College London. The goal of *The Thera Fresco Project* is to develop new software and scanning technology for helping the archaeologist and conservators reconstruct the Theran Wall Paintings. In this thesis we focus on the 2-D acquisition system and multi-feature matching, as well as visualization techniques. Other components of the larger project include a 3-D acquisition system and a 3-D matching application [Brown, 2008].

### 1.4.1 Problem



**Figure 1.2:** Contour Erosion and Discoloration: Traditional matching cues such as contour shape and color may fail when fragments are eroded.

For decades, the Theran Frescoes have provided essential clues for understanding the culture and lifestyles of a lost civilization. The imagery depicted may range from elaborate geometric patterns to human figures engaged in ritualistic activities. It is also important to note that these frescoes were not simply artwork, but architectural elements that followed the structure of the walls of the each building, often having thickness or curving around wooden beams. The success archaeologists have had with these wall paintings partly lies in their completeness which is quite unique in the ancient Mediterranean. Still, whether using manual or digital methods, the task of matching in this domain is a daunting task. The main challenges that make this fresco matching problem difficult stem from the current physical condition of the fragments themselves. Interestingly enough, these same characteristics serve as a driving force for our solution. In this section, we examine the physical characteristics of Akrotiri fresco fragments. We pay particular attention to how these traits, though challenging, may provide insights into the design of an effective computer-based acquisition and matching pipeline.

We begin with a brief explanation of how ancient wall paintings are created. Frescoes are constructed by applying pigments (usually naturally occurring minerals) to a plaster surface. Underneath the plaster's top surface are several layers including a base layer of mud, on top of which is added a layer of plaster (about 1.5 cm thick). The top-most layer consists of thinner layers of plaster. Once the layers are assembled, the top surface is rubbed smooth. To add color, pigments are brushed onto the surface and eventually bond with the wet lime plaster. The result is a durable, stable surface. This technique, called *buon fresco* is contrasted with another one called *fresco secco*, where the plaster surface is already dry before pigments are applied. In this case, the image is less permanent and the paint easily flakes off.

The fragments found at Akrotiri may range in size from a few millimeters to a couple of meters. However, the majority of the fragments are relatively small with a nearly flat top surface. Close examination of a fragment's surface reveals geometric discontinuities

such as cracks and subtle markings. In addition, there are often subtle impressions on the surface. These impressions were left behind by artists who placed incision markings, or strings in the wet plaster as guides to quickly layout their designs. We observe that string impressions were often placed in locations that separate regions of solid color from regions of pure white. In fact, Akrotiri wall paintings are known for their large fields of white or other solid colors. This particular characteristic makes manual reassembly especially difficult. We also find intricate textures and reed imprints imbedded on the back surface of fragments. Figure 5.1 depicts examples of these physical traits. The small size of the fragments combined with their unique surface impressions suggest an acquisition methodology capable of capturing more accurate detail in a more efficient manner than is available with current 3-D systems.

One of the more interesting characteristics of the Thera frescoes is their current state of deterioration, a result of their resting in the earth for thousands of years. As shown in Figure 1.2 (left), the color on the fragment's surface is often faded and the contour edges are usually crumbling and inconsistent. This suggests that the traditional features used in puzzle matching systems, such as color and curvature (as discussed in Chapter 4), may prove ineffective in this domain. A more obvious example of this is shown in Figure 1.2 (right) where neighboring fragments have become discolored such that adjacent fragments (that clearly have the same imagery) no longer share the same colors. This can occur when a subset of fragments from the same fresco are in contact with soil or objects that have a different pigmentation. The implication here is that a new matching system should have the ability to identify novel cues to aid the matching process. In addition, any such features must be extracted from areas of the fragment surface that are not affected by erosion.

The numbers of fragments at Akrotiri can be measured in the thousands. In fact, there are literally enough fragments to occupy the conservators for centuries. It is evident that any practical acquisition system must be efficient and intuitive without impeding upon the current work-flow on-site. A successful matching pipeline would have to be implemented

incrementally in order to scale to the millions of possible match orientations and alignments. A user friendly method for presenting the most probable matches among many possibilities is a requirement. As mentioned above, each fresco depicts its own themes. Imagery as well the palette of pigments may vary greatly, even among neighboring rooms in the same building. It is easy to observe from these samples that different paintings may have different subsets of characteristics that may be more dominant cues for matching. This suggests a solution that is not limited to any one feature or trait, but rather a combination of cues. Our solution should be able adapt to effectively utilize the best qualities of each feature based on its importance for the given fresco in order to find the best match.

## 1.5    Contributions

Now that we have presented a historical background of the datasets on which we operate, we turn our attention to the set of techniques we have developed for ensuring their preservation through digital documentation, reconstruction and visualization. The focus of our work is two fold. First we develop methods for accurately recording real-world objects that have complex surface geometry (Chapter 3). We build upon previous approaches that store normals and depth information in the form of images. The goal of our work is to take advantage of different types of input data that increase the overall effectives of applications in this domain. The first application we develop is a multi-feature matching system for reassembling fresco fragments (Chapter 5). The second is a non-photorealistic rendering pipeline for generating high quality technical illustrations (Chapter 7).

Our key contributions are:

- Algorithms for obtaining high-resolution 2-D color texture and normal maps using a flatbed scanner [Brown *et al.*, 2008]. Our approach is a variant of photometric stereo that implements a novel calibration phase to accommodate a linear light source (not considered in traditional shape from shading algorithms). With our method, normal

maps are both quick to capture and capable of documenting sub-millimeter detail on the front surfaces of nearly flat objects.

- A computer-assisted matching framework for reassembling fragments of artifacts. First we incorporate normal maps as a new source of information for digital matching systems [Toler-Franklin *et al.*, 2010]. We show that the additional forms of information encoded in our normal maps (surface orientation and fine detail) are more discriminating than color-only solutions when there is deterioration and fading. They can significantly improve solutions that use full 3-D models; particularly when erosion makes geometric approaches unreliable. We develop a set of new feature descriptors that are simple functions computed directly on our color and normal images. Their effectiveness lies in both their efficiency and ability to quantify physical characteristics of the object's surface structure. We introduce an analysis and evaluation methodology that demonstrates the performance of different subsets of features over three different datasets of fresco fragments [Toler-Franklin *et al.*, 2010; Funkhouser *et al.*, 2011]. Our matching framework uses machine learning algorithms to combine multiple features for matching. Our results show improved match prediction rates over existing 3-D algorithms.

- A non-photorealistic rendering pipeline for scientific illustration of artifacts. Our system includes algorithms for adapting signal processing tools for use with multi-channel images [Toler-Franklin *et al.*, 2007]. We create analogs for smoothing, curvature estimation and segmentation. These are essential for creating a variety of stylization effects. We then extend existing stylization techniques for use with images that store a per/pixel color and normal. Specifically, we implement analogs of shading algorithms such as toon shading and exaggerated shading. We also implement line drawing techniques including suggestive contours, discontinuity lines and ridge and valley lines. Finally, we present two new algorithms for simulating shadows us-

ing images with normals; multiscale mean curvature shading and fast discontinuity shadows. These effects help with the overall comprehensibility of objects and are alternatives when there is no depth information. We show several illustrative examples that our results are suitable for archeological study and museum conservation.

# Chapter 2

# Multi-Channel Images

In this thesis we operate on image datasets that store multiple channels of information at each pixel (rather than r, g, b color triplets only). In this section, we describe examples of multi-channel images, emphasizing their advantages over full 3-D representations and traditional color-only photographs. We are particularly interested in how the additional sources of information encoded in these formats can be utilized to improve applications that involve the careful documentation and visual analysis of the surface properties of real-world objects. It is our claim that these datasets are not fully applied in the domain of cultural heritage preservation, despite their potential. In the remaining chapters, we focus on new capture techniques and processing tools for a specific type of multi-channel image that stores both a color and a surface normal at each pixel. These images form the basis of the pattern matching algorithms presented in Chapter 5 and the rendering pipeline presented in Chapter 7.

## 2.1 Hyperspectral Images

Hyperspectral images are muilti-channel images that record information from the electro-magnetic spectrum. In addition to storing the range of visible light, these formats include light with wavelengths ranging from ultraviolet to infrared. The spectral data is typically

stored as a collection of images, each with an electromagnetic range. These spectral bands are usually operated on when combined in a hyperspectral cube. Acquisition methods for this data type incorporate the use of airborne sensors or satellites.

Spectral analysis has several advantages over traditional 2-D signal processing techniques. It permits the study of information in spaces that are independent of color. In fact, many objects have unique spectral signatures that are only recognizable at different wavelengths of light. This is especially useful when identifying unique traits of objects that are indistinguishable with the naked eye (or have similar color). In other scenarios, spectral images may be used to isolate specific properties of interest, while discarding others that may confuse the analysis. This approach may significantly improve a variety of computer vision applications (like segmentation and photometric reconstruction), where features appearing on the surface of an object that are not intrinsically related to its shape (such as specular highlights) may confuse the algorithm. Research by [Zickler *et al.*, 2008] explores the use of invariants derived from illuminant-dependent 'subspaces' of RGB color space to improve such algorithms. The method may also be generalized to a multi-channel hyperspectral image. The project demonstrates how the utility of these invariants can improve applications such as the segmentation of objects by their material properties.

## 2.2 Polynomial Texture Maps

Some capture techniques permit a full lighting model to be stored at each pixel. *Polynomial Texture Maps* (PTMs), introduced by [Malzbender *et al.*, 2001], have been particularly useful for studying ancient archeological clay and stone writings. In this representation, the coefficients of a biquadratic polynomial are stored at each texture pixel. These coefficients are used to reconstruct the surface color under varying lighting conditions. The light orientations may be chosen to emphasize deformations and imperfections on the surface. Although the primary purpose of PTMs is to create more photorealistic texture maps, a set

of operators (reflectance transformations) can be applied to increase the local contrast for close inspection of fine surface details in a more non-photorealistic manor. As described in Section 1.2.1, the digital camera and lighting domes used to acquire this data are becoming more prevalent in museum settings.

## 2.3   Images with Normals

We focus our work on multi-channel images that store both a color and a normal at each pixel (RGBN images [Toler-Franklin *et al.*, 2007]). These datasets are well suited for applications in the cultural heritage domain. First, they are easy to capture in a processes similar to taking digital photographs. Low-cost, off-the-shelf capture devices (digital cameras and 2-D flatbed scanners), make the process practical and significantly easier than 3-D scanning. RGBN images are more informative than traditional color images because they store some information about the object's shape. In addition, they have higher resolution color and normal maps than the 3-D geometry from 3-D laser scanners (under certain conditions). As a result, analysis applications that operate on images with normals are more robust when there is erosion of edge geometry or color. Finally, these data types can be used to produce a wide variety of stylization effects that accentuate subtle patterns on the object surface. Most state-of-the-art non-photorealistic rendering algorithms (Gooch and Gooch [2001]) are simply functions of the surface normal, lighting and viewing directions. Simple extensions to signal processing tools can preserve the integrity of the normals, while introducing a wide range of control for a variety of stylistic effects. Overall, RGBN images are more efficient to process than full 3-D geometry, requiring less storage and computation time. Functions are computed in image space producing powerful 3D results with simpler 2-D methods.

# Chapter 3

# Data Acquisition

In Chapter 1, we introduced a variety of acquisition methods developed in prior work. Our discussion highlighted the fact that the choice of technique is closely dictated by the project goals, capture environment and characteristics of the target datasets. In this section, we explain two different acquisition techniques and capture devices developed for our specific datasets of artifacts. As explained in Section 2.3, our datasets are represented as multi-channel images that store a color and a surface normal at each pixel (RGBN images).

## 3.1   Photometric Stereo

Systems based on the principle of shape from shading [Horn, 1970] are the primary methods for acquiring high-quality, dense normal maps. In particular, several recent systems have used variants of photometric stereo [Woodham, 1980; Rushmeier and Bernardini, 1999], which recovers normals under the assumptions of Lambertian (perfect diffuse) shading and multiple light sources of known position and brightness. The photometric stereo calculation operates on each pixel independently. A surface point of albedo $a$ and normal $\hat{n}$ is illuminated multiple times by light sources with directions $l_i$, yielding intensities $e_i$.

Using the Lambertian lighting law

$$e_i = a\,(\hat{n} \cdot l_i),$$ (3.1)

we may arrange all of the available data into a matrix equation:

$$
\begin{pmatrix}
l_{1,x} & l_{1,y} & l_{1,z} \\
l_{2,x} & l_{2,y} & l_{2,z} \\
l_{3,x} & l_{3,y} & l_{3,z} \\
& \vdots &
\end{pmatrix}
\begin{pmatrix}
a\,\hat{n}_x \\
a\,\hat{n}_y \\
a\,\hat{n}_z
\end{pmatrix}
=
\begin{pmatrix}
e_1 \\
e_2 \\
e_3 \\
\vdots
\end{pmatrix}.
$$ (3.2)

This equation may be solved using least squares for the vector $a\,\hat{n}$, whose length and direction are given by $a$ and $\hat{n}$, respectively. Typically, more than the minimum three lights are used, providing noise reduction and allowing for outlier (specularity and shadow) rejection.

## 3.2  Hand-Held Flash

Our first RGBN acquisition system uses a setup consisting of a standard digital SLR camera and flash (Figure 3.1, left). In order to allow the flash to be moved by hand, we place mirror-reflective and white-diffuse spheres in the scene and use them to solve for flash position and intensity, respectively (a similar arrangement was used by [Masselus *et al.*, 2002]).

We usually capture 30–40 images with different flash positions, and a typical capture session takes only 2–3 minutes. The large number of images in a dataset provides considerable statistical redundancy, allowing for robust exclusion of shadows and specularities during the photometric stereo computation (i.e., equation 3.2). We are thus able to recover normals even for highly-occluded geometry for which only a small percentage of light positions are visible from each pixel (e.g., the pinecone in Figure 7.1) and materials with a significant specular component (e.g., the helmet in Figure 7.8). In addition to extracting
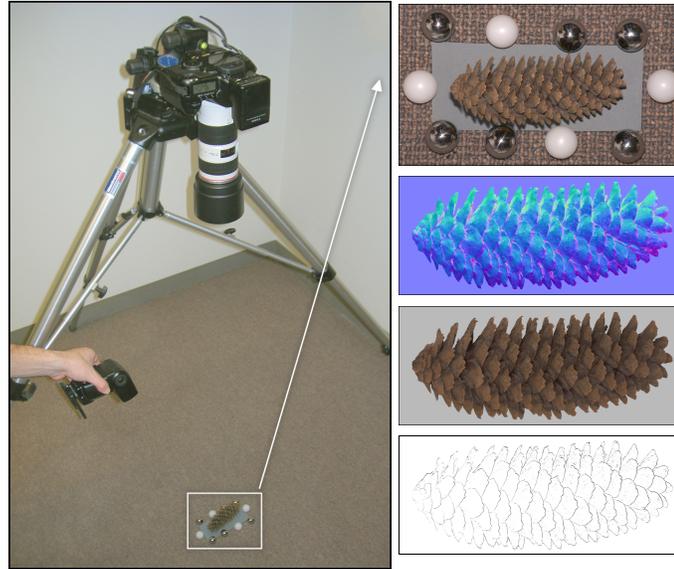
**Figure 3.1:** Digital SLR Camera and Hand-Held Flash: **Left:** our setup for capturing RGBN images, using a digital SLR camera and hand-held flash. White and mirror spheres are used to find the flash intensity and position for each captured image. **Right:** An original image, together with extracted normals, colors, and depth discontinuities. Note that a hand-drawn mask was used to isolate the object of interest.

RGB albedo maps and per-pixel normals, we use the method of [Raskar *et al.*, 2004] to find locations of depth discontinuities in the scene: these are locations of shadow boundaries *in the light direction*. The results are shown in Figure 3.1, right. Note that a hand-drawn mask was used to isolate the object of interest.

## 3.3 Flatbed Scanner

Our second RGBN capture method uses a flatbed scanner to generate normals from high-resolution 2-D color scans of objects. In this section, we explain the advantages of this approach and provide details of our normal reconstruction process.

### 3.3.1 Advantages

Several characteristics of flatbed scanners make them more effective capture devices in some environments and for certain types of artifacts. For example, they do not require

careful control of ambient lighting or meticulous positioning of objects. For these reasons, they are well suited for use by non-experts at an archeological excavation site to capture fragments that are small and relatively flat. The acquired color and normal images have a significantly higher resolution than that of cameras or 3-D scanners. High resolution color and normal maps obtained under controlled illumination are important for archiving, analyzing, digitally restoring, and matching fragments based on surface markings and decoration. In addition, a fixed spatial resolution is always guaranteed as the scanner's sensor is a constant distance from the object's surface. In the following sections, we present algorithms for obtaining 2-D color scans and normals used to document fine relief on the front surfaces of wall paintings. We use these techniques to acquire our fresco datasets.

### 3.3.2   Capturing Non-Traditional Matching Cues

As explained in Section 1.4, frescoes often contain unique surface markings on their front surfaces. Akrotiri wall paintings, for example, display string impressions left behind by artists who frequently placed strings in the wet plaster to outline their drawings and as an aid to demarcate long, straight color edges. More intricate designs such as spirals were sometimes prepared with narrow incisions [Vlachopoulos, 2008]. Brush strokes, patterns left behind by tools used to smooth the wet plaster surface, are another type of visual cue. These impressions are helpful in matching, analysis and restoration, because they survive even where the original pigment does not. Such narrow, shallow markings are difficult to capture with either a digital camera or range scanner, particularly when large volumes of data are involved. In the following sections, we will demonstrate that a flatbed scanner is suitable for documenting these subtle details in the form of high-resolution normal maps. As you will see, these geometric features appear clearly on the surface in regions where there are sharp changes in normal direction.

### 3.3.3 Previous Examples

Several approaches have used 2-D flatbed scanners (or similar devices) to capture images with some 3-D qualities (without depth information). One approach uses the 2-D scanner as if it were a 3-D capture system [Schubert, 2000]. In this case, the system leverages the fact that CCD-based scanners produce orthographic projections in the *y* direction and a perspective projection in *x*. By repositioning the object in the x position, multiple images are taken from different perspective views. The images are then re-arranged and viewed to generate a stereogram. Other approaches use scanning hardware with additional light waves for dust and scratch removal on transparent objects. For example, the *DIGITAL ICE* $^{TM}$ technology from the Eastman Kodak Company introduced infrared hardware. The system is able to detect defects on the surface when the infrared light is blocked by dust or scattered by scratches.

### 3.3.4 Color and Normal Reconstruction

We obtain normal maps of fresco fragments via shape from shading, beginning with several (typically four) scans of a fragment at different orientations. Our approach is conceptually straightforward: during a calibration phase we measure $I(n)$, the observed brightness as a function of the surface normal. Then, given a set of scans, we invert $I$ to solve for the normal. As a side effect, we obtain the true color texture (diffuse reflectance) of the surface.

During a scanning session (Figure 3.2), the user first enters a fragment's ID into the 2-D scanning software, places the fragment face-down on the flatbed scanner, and then clicks the "scan" button. The software automatically finds the fragment's location in a pre-scan using background subtraction. It then scans the fragment (only scanning the detected region occupied by the fragment), and stores it in the fragment database. The user performs four scans, rotated by approximately 90 degrees, then turns the fragment over and repeats the process. A single user can scan approximately 20 fragments per hour. If normals are not required, only a single scan of each face is performed, for a 70 fragment-per-hour throughput.

As previously mentioned, our system is designed to complement 3-D acquisition systems at an excavation site. Although color images may also be obtained with a digital camera, or a 3-D scanner (in conjunction with geometry), we believe that the improved color fidelity and resolution justify the additional effort of performing 2-D scans.
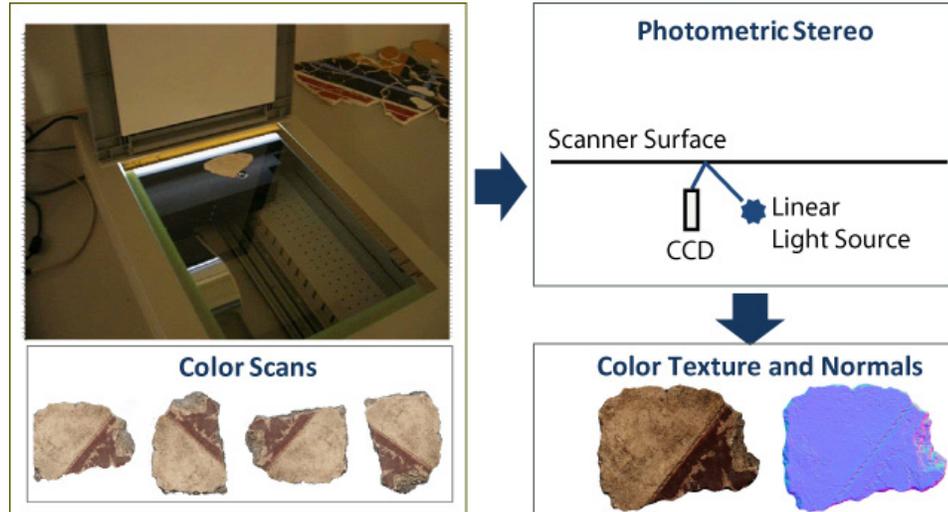


**Figure 3.2:** Acquisition with a 2-D Flatbed Scanner: Capturing RGBN images using a high resolution 2-D flatbed scanner. **Left:** The object is scanned at four orientations. **Top-Right:** Photometric stereo algorithms utilize the slight distance between the scanner's sensor (CCD) and the light source. However, the scanner's light source is linear; a calibration step is used to measure $I(n)$. **Bottom-Right:** The outputs are high resolution color (texture) and normal maps.

### 3.3.5   Calibration Step

Because the scanner's light source is linear, not a point, $I(n)$ will not be of the form $n \cdot l$. We therefore measure it in a calibration step, using a diffuse plane inclined at known angles with respect to the scanner platform. We measure the intensity at several angles ranging from $0°$ to $25°$ of tilt and rotation to ensure good coverage of the space of normals. Because intensity varies with height off the scan bed, we perform all measurements at a constant height.

We fit a spherical-harmonic model to the diffuse plane measurements to obtain a parametric representation of $I(\cdot n)$ that averages out measurement noise and is more easily in-

verted. Using cross-validation, we have determined that second-order spherical harmonics do not substantially decrease the calibration error over a first-order representation containing only constant and linear terms in the normal components:

$$I(\cdot n) = (a_0 \; a_1 \; a_2 \; a_3) \begin{pmatrix} n_x \\ n_y \\ n_z \\ 1 \end{pmatrix} = \cdot a^T \cdot n \,. \tag{3.3}$$

## 3.3.6 Normal Computation

After capture, we select one of our four scans to be the target orientation, and then use the 2-D to 3-D registration algorithm presented in [Brown *et al.*, 2008] to register each scanned image to this orientation. The inverses of the resulting rotations, which we will call $R_i$, may be thought of as rotating the light relative to each pixel; in other words, the $R_i$ act on the calibration coefficient vector $\cdot a$. For each image $i$ we may write

$$I_i = \rho(R_i \cdot a)^T \cdot n \,, \tag{3.4}$$

where $\rho$ is the diffuse albedo. In matrix form, this expands to

$$\begin{pmatrix} - \; (R_1 \cdot a)^T \; - \\ - \; (R_2 \cdot a)^T \; - \\ \vdots \end{pmatrix} \begin{pmatrix} \rho n_x \\ \rho n_y \\ \rho n_z \\ \rho \end{pmatrix} = \begin{pmatrix} I_1 \\ I_2 \\ \vdots \end{pmatrix} \,. \tag{3.5}$$

Note that it is not possible to solve this problem as stated using standard linear least squares, as the leftmost matrix in equation 3.5 (let us call it $A$) is singular: both its last and next-to-last columns will be constant across all rows, because all of the $R_i$ represent planar rotations that leave the $z$ component unchanged. Instead, we define the matrix $\tilde{A}$ to be $A$ with its fourth column removed, giving

$$\tilde{A} \begin{pmatrix} \rho n_x \\ \rho n_y \\ \rho \tilde{n}_z \end{pmatrix} = \begin{pmatrix} I_1 \\ I_2 \\ \vdots \end{pmatrix} \,, \tag{3.6}$$

where $\tilde{n}_z = n_z + a_3 \rho / a_2$. Equation 3.6 can be solved using linear least squares, and by

32

adding the constraint that

$$n_x^2 + n_y^2 + n_z^2 = 1 \,, \tag{3.7}$$

we are able to to solve for $\rho$, $n_x$, $n_y$, and $n_z$.

### 3.3.7 Results
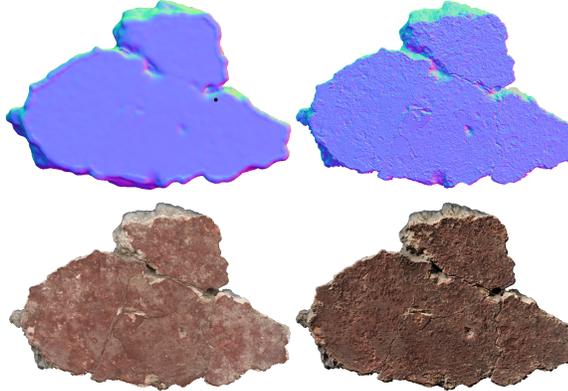


**Figure 3.3:** Acquisition Results - Surface Details: Computed normals (top right) reveal more surface detail than those extracted from the geometry (top left). Combining the high-resolution normals with the extracted RGB color (bottom left) allows for enhanced visualizations [Toler-Franklin *et al.*, 2007] of surface detail (bottom right).
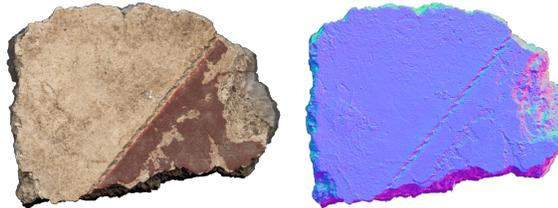


**Figure 3.4:** Acquisition Results - String Impressions: String impressions, most clearly visible in the computed normals, mark boundaries of solid color; they are an important cue for reconstruction, restoration, and archaeological study.

We compared the 2-D color and normal maps generated with our flatbed scanner with the color texture and geometry captured using the 3-D laser triangulation scanner currently used at Akrotiri [Brown *et al.*, 2008]. As shown in Figure 3.3, the color and normal maps we recover have an order of magnitude higher resolution than our geometry. This is because the 3-D scanner's camera provides only limited color fidelity, its lighting is uneven, and its resolution is limited. We are also able to capture unique identifying geometric features

such as the string impression depicted in Figure 3.4. In Chapter 5, we will use such cues to inform a computer-assisted matching system designed to re-assemble frescoes. In Chapter 7, we will demonstrate how to compute a set of non-photorealistic rendering functions on these RGBN images to visualize, and analyze fine surface detail (Figure 3.3), including hairline cracks, string impressions (Figure 3.4), plaster grain, and impressions of reeds or other materials.

### 3.3.8  Limitations

There are some limitations associated with the use of flatbed scanners in this context. First, their depth-of-field is low (only a few millimeters for the hardware we have tested). As a result, we cannot always extract contours for 2-D matching since it may be difficult to distinguish which pixels belong to the front side and which ones come from the fractured edges. Another drawback is the presence of artifacts in the form of 'shadowing' along the fragment edge. Data in this region is unreliable. There are also issues of perspective parallax, which is particularly noticeable if the fragment surface curves upward at distances greater than a few mm from the scanner's surface. If our registration operation is not accurate, defects appear in the normal maps in the form of embossing; particularly around surface decoration. Finally, we are limited to fragments small enough to fit on the scanner's glass window, and stable enough to be handled. Overall, wall painting fragments are well suited to this acquisition method as they are reasonably small and have flat front faces (with few exceptions).

# Chapter 4

# Computer-Assisted Matching

In Chapter 3, we presented methods for acquiring non-traditional images that store more than just three channels of color information at each pixel. Our next goal is to leverage the new forms of data that our digital media provides to improve the fidelity, robustness and efficiency of applications for cultural heritage preservation. The first application we explore is computer-assisted matching. The key goal of any matching system is to analyze the shape and appearance of fragments of objects in search of characteristics that group them by similarity. Approaches range from examining image-space features to exhaustively searching for the optimal alignment along fractured edges of a full 3-D model. In all cases, it is clear that each method has its own strengths and weaknesses and that the effectiveness of each approach is directly linked to the domain of the target dataset. In this chapter, we review several of these methods. We highlight how our approach within the domain of fresco matching improves previous work by combining image and geometry information with additional cues based on normals.

## 4.1   2-D Matching

The first group of techniques apply 2-D matching algorithms to image-space features such as 2-D contours, color and texture. For example, [Kong and Kimia, 2001] reassemble

jigsaw puzzles by examining image-space contours of puzzle pieces. First, curve fitting algorithms are used to align pairs of pieces using local features at both a course and fine scale. To determine the strength of a candidate match, both the difference in curvature between corresponding points on the two contours and their relative lengths are considered. The result is a ranked list of affinity measures for each possible matching pair of pieces. Finally, a global solution is found using a best-first search and backtracking. In this stage, pieces are considered in triplets and ambiguities from local shape matching are resolved. A similar approach by [Leitão and Stolfi, 2002] compares the curvature of fragment outlines at progressively increasing scales of resolution using incremental dynamic programming and sequence-matching. Unlike [Kong and Kimia, 2001], the solution does not constrain matches to triplets, producing fewer false negatives on sparse datasets.

Another contour matching approach by [Papaodysseus *et al.*, 2002] reassembles images of fragments from the Theran Frescoes. In this case, a brute-force approach is used to consider all possible alignments along the contours of fragments. Both gaps between aligned fragments and their overlap are used to determine the likelihood of a match.

Other 2-D approaches like [Fornasier and Toniolo, 2005; Sağiroğlu and Erçil, 2006], apply pattern matching algorithms to image color and texture. In the case of [Fornasier and Toniolo, 2005], the goal was to determine the location and orientation of the remaining fragments of the Chiesa degli Eremitani. These renaissance frescoes from Padua, Italy were destroyed in World War II. Fragments were assembled by correlating color images of the surviving pieces with black and white photographs taken in the early 1900's. Discrete Circular Harmonic expansions were incorporated into a fast and robust image correlation algorithm. [Sağiroğlu and Erçil, 2006] extended this approach to align neighboring fragments using an image in-painting algorithm and eliminating the need for overlapping images.

There are several limiting factors that suggest that traditional 2-D algorithms may be improved upon for our fresco reassembly problem. They are often sensitive to erosion and

discoloration, a significant issue for fragments that have spent thousands of years exposed to natural elements. Moreover, they do not consider the wealth of 3-D information available in geometric representations. Such cues are particularly important in our domain where impressions on the fragment surface provide strong matching cues. We believe that such cues are encoded in our multi-channel images. In addition, we sample our features on the surface and in the center of fragments to avoid acquisition artifacts, a common issue with contour matching solutions.

## 4.2   3-D Matching

Other approaches for assembling fractured objects incorporate full 3-D descriptions. For example, [Huang *et al.*, 2006] reassemble solid objects by first identifying fractured regions, then generating clusters of feature patches for alignment-based matching. Although these feature clusters effectively describe the local geometry of the fracture surface, the algorithm does not consider other physical attributes of the dataset, and is burdened by the complexity of a full 3-D matcher.

[Brown *et al.*, 2008] exploits the orientation constraints of flat fragments to achieve a simple, fast matcher based on edge geometry. This matcher resamples the fragments' edges in a regular grid structure, then exhaustively tests every possible alignment of a pair of fragments in a few seconds, in a correlation-like manner. This approach takes advantage of high resolution geometry to find precise alignments, and mirrors the common technique of finding matching fragments by testing for pairs that physically "lock" together. On the other hand, fragment edges are subject to erosion, and the brute-force nature of the algorithm means there is no early rejection for non-matching pairs.

Our approach retains the efficiency of a special-purpose matcher for flat objects, but focuses on fine surface details rather than edge information. It is complementary to existing geometry-based matchers in two important ways. First, it matches features on the *external*

surface of fragments rather than fitting fractured faces to each other. Because our feature descriptors are computed on images, they are generally quick to compute. Second, it relies on high resolution normals captured with a flatbed scanner that could not be acquired reliably with current stereo-based scanners or fed into an alignment algorithm.

## 4.3 Reassembling Artifacts

Several computer-aided systems have been designed specifically for reassembling broken objects. One notable example is the Forma Urbis Romae project [Koller *et al.*, 2006]. Computer algorithms were developed to analyze incision points and markings on fragments of a marble map of Rome that dates back to the 3$^{\text{rd}}$ Century A.D. One of the key challenges of this dataset is the scarcity of information (only approximately 10% of the map is known). Thus [Koller and Levoy, 2005] extend incision markings beyond fragment boundaries to identify additional matches. While some aspects of these heuristics are of broader applicability (for example, in lining up fragments with string impressions), they are largely tuned to the specific needs of the Forma Urbis Romae.

Another common application in the field of archaeology is the reassembly of broken pottery (sherds). Just as we take advantage of properties of fresco fragments, such as the flatness of the front faces and slowly varying thickness, to obtain an effective, efficient matcher, these algorithms rely on finding the axis of rotation and profile curve common to pottery [Willis, 2004; Karasik and Smilansky, 2007]. For example, [Willis, 2004] fits a polynomial to an initial estimate of the vessel's axis to obtain a profile contour. Next, the algorithm iteratively refines the shape of this curve. Finally, a probabilistic framework is used to find matches among similar axis and profile pairs. The algorithm continually refines the solution as the vessel is reconstructed.

Our approach improves upon these examples because we interpret observed qualities of our domain as a set of functions that are easy to compute, optimizing our system to

use a combination of the most discriminative criteria for matching. Our descriptors can be generated on any collection of color and normal images and are not strictly tied to a specific dataset. Although we incorporate some 3-D quantities, such as thickness, we maintain the ease and simplicity of a 2-D system by only computing in image space.

## 4.4    Contributions: A Comparison With Previous Work

There are four important contributions of our matching approach that distinguish it from the methods described above. First, we address the challenge of operating on eroded objects by introducing new input data for matching in the form of robust, high-resolution *normal maps*. We argue that these are easy to acquire at a higher resolution than 3-D models and are more robust than color and 2-D contour features when there is deterioration. This is particularly important in the case of fresco fragments when 3-D geometry is not always reliable. Our second contribution is a set of novel feature descriptors computed on normal-maps that prove effective for grouping fragments by similarity. A third addition is an evaluation methodology that demonstrates improved match retrieval performance over state-of-the-art 3-D matching algorithms. Finally, unlike other approaches, our system incorporates multiple cues. Rather than rely on a single feature, we generate a collection of features over our datasets and use machine learning algorithms to combine the best qualities of each.

# Chapter 5

# Multi-Feature Matching Framework

In this chapter, we present a multiple-feature approach for determining matches between small fragments of archaeological artifacts such as Bronze-Age and Roman frescoes. In contrast with traditional 2-D and 3-D shape matching approaches, we introduce a set of feature descriptors that are based on not only color and shape, but also *normal maps*. These are easy to acquire and combine high data quality with discriminability and robustness to some types of deterioration. Our feature descriptors range from general-purpose to domain-specific, and are quick to compute and match. We have tested our system on three datasets of fresco fragments, demonstrating that multi-cue matching using different subsets of features leads to different tradeoffs between efficiency and effectiveness. In particular, we show that normal-based features are more effective than color-based ones at similar computational complexity, and that 3-D features are more discriminative than ones based on 2-D or normals, but at higher computational cost. We also demonstrate how machine learning techniques can be used to effectively combine our new features with traditional ones. Our results show good retrieval performance, significantly improving upon the match prediction rate of state-of-the-art 3-D matching algorithms, and are expected to extend to general matching problems in applications such as texture synthesis and forensics.

## 5.1 Overview

Advancements in low-cost, high-volume acquisition systems have made computer-assisted reconstruction of artifacts from small fragments practical. This problem is of particular interest to the field of archaeology, in which the reconstruction of artifacts such as shattered wall paintings reveals information about the history and culture of ancient civilizations. Historically, the process of reconstructing these wall paintings has been manual, occupying a major proportion of the human effort at excavation sites. As a result, wall painting reassembly is not even attempted at countless sites around the world, leading to a significant opportunity to advance our knowledge of ancient societies by improving the practicality of reconstruction.

Several 2-D and 3-D computer-aided matching approaches have been explored and have proven successful in some domains. However, current matching algorithms have difficulty when matching artifacts that have deteriorated over many years. For example, they may consider features such as color, which frequently have changed over time even among neighboring fragments. Alternatively, they may operate exclusively on 3-D geometry, which may not only have deteriorated, but is also challenging to acquire with the same fidelity and resolution as color images.

We address the problem of reconstruction by considering multiple cues based on color, shape, and — most interestingly — normal maps. The latter is a new source of information that has not been used for matching in previous work, and we argue that it combines high data quality and resolution with high discriminability and robustness with respect to certain types of deterioration. As has been recently demonstrated [Brown *et al.*, 2008; Pintus *et al.*, 2009], it is practical to use flatbed scanners to obtain normal maps of mostly-flat objects with 600 or 1200 dpi resolution. These normal maps reveal salient surface characteristics including string impressions, brush strokes, surface roughness, and fine cracks.

Our system begins with scanned images and normals of a collection of fragments, and computes a set of *feature descriptors*. Each descriptor may be computed over an entire

| Brush Strokes | String Impressions | Curvature | Erosion | Decorative Imagery | Texture |

**Figure 5.1:** Fresco Features: Feature descriptors based on surface normal characteristics can capture a variety of physical characteristics such as brush strokes, string impressions, and erosion, that are used by archaeologists when assembling fresco fragments. Combining these with more traditional color-based features and 3-D features using classification trees yields significantly improved matching performance.

fragment or over small patches sampled around the outer contour of the fragment: the tradeoff of sampling patches is greater discriminability for greater computation time. Descriptors range from general-purpose (such as variance in the normal map) to ones designed specifically for the domain of fresco fragment matching, and motivated by visual cues used by archaeologists for reassembly (such as brush stroke direction). They are designed to capture characteristics like shape, surface decoration, surface texture, and deterioration (Figure 5.1).

We use similarity of these descriptors to suggest matches, and evaluate their performance on three different wall-paintings. The first is a geometric scene containing spirals and large areas of constant color, from a late-bronze-age Aegean civilization at Akrotiri, on the island of Thera (Santorini). The second is from a Roman villa at Kerkrade, The Netherlands, and is especially distinctive because of the strong brush marks visible both in the color and the surface relief. The third is a synthetic fresco, professionally created and shattered, for which a ground-truth reconstruction is available.

We perform a cross-validation analysis on databases of dozens to hundreds of fragments, drawn from the three different wall-paintings. Our results demonstrate the discriminative power of our collection of features, and suggest that matching performance is improved by the use of normal maps, in addition to features based on more conventional data sources such as color, thickness, and exterior contour shape. Moreover, we observe that the performance of individual features varies from dataset to dataset, suggesting a future

extension to online learning.

Overall, our matching framework makes the following contributions:

- The introduction of a new type of input, *normal maps*, for matching small fragments of artifacts. We argue that normal maps are easy to acquire with higher resolution than 3-D models, and are more robust to deterioration and discoloration than color.

- A set of easily computable descriptors, of both general purpose and domain specific types, that are effective for matching.

- Analysis and evaluation methods that demonstrate how well our features perform over state-of-the-art 3-D match algorithms.

- A matching framework that is easily extendable to more generalized matching problems used in applications such as texture synthesis and forensics.

## 5.2 System Design

In our system, we focus on obtaining feature descriptors from a database of scanned patches of objects, focusing on an archaeological fragment matching scenario. In this section, we describe our feature descriptors, match classification strategies, and the datasets on which we operate. We use three forms of data: *color maps* acquired using a high-resolution (600 dpi) 2-D scanner, *normal maps* obtained from multiple scans using a variant of shape from shading(Section 3.3.4), and *3-D meshes* from a laser-triangulation range scanner. We use these data types because they can be obtained in situ, at an archaeological excavation or in the context of other digitization efforts, with high fidelity, low cost, and ease of acquisition.

### 5.2.1 Feature Descriptor Generation

Figure 5.2 presents a conceptual overview of our feature descriptor pipeline. Because practical datasets may contain thousands of fragments, we focus on designing a match-

ing pipeline capable of scaling to these data sizes. Indeed, a brute-force solution that tested every possible alignment of every possible pair of fragments would quickly become infeasible, requiring perhaps $10^{10}$ to $10^{12}$ comparisons (a few thousand fragments times a few hundred orientations, squared). To overcome this growth we employ a sequence of matching stages, ranging from ones that can quickly reject a large number of implausible candidates to ones that precisely check individual matches. A key observation is that the early stages should require computation that grows *linearly* with the number of fragments, rather than quadratically.

We thus consider three possible classes of features. The first includes **per-fragment** features: those that are computed (once) for each fragment in our database. Fragments that differ greatly in the computed descriptors are assumed to have a low probability of matching, hence generation of plausible matching pairs of fragments could be accelerated with a fast clustering or indexing technique (we focus on demonstrating the matching accuracy of the descriptors, rather than evaluating their efficiency).

The second class of features we consider are **per-patch** features. These are computed not over entire fragments, but over small regions sampled around the boundary. Because these features consider more localized properties, we expect that they will be more dis-
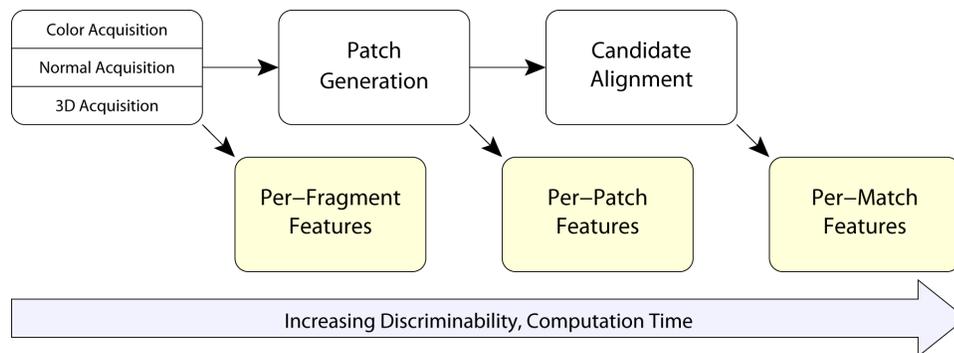


**Figure 5.2:** An overview of feature generation. We extract a variety of feature descriptors from high-resolution color scans, normal maps, and 3D models of each fragment. Descriptors may be extracted at the fragment level, patch level, or for a specific candidate match. Although we can use machine learning to match all features simultaneously (Section 5.8), it is advantageous to first select possible matches with per-fragment and per-patch features, then compute the more expensive per-match features only for these possibilities.

criminative of true matches. On the other hand, they also require more computation than the per-fragment features, since they must be evaluated at dozens to hundreds of locations around the perimeter of a fragment. Nevertheless, the descriptors are precomputed and cached, once per fragment, adding only a few seconds to the preprocessing time. The patches we use are circular and are sampled from the original 600 dpi images, every 5 mm along the perimeter of the fragment. For most features, we use patches 10 mm in diameter, and offset them 8 mm inward from the outside contour. This ensures that the resulting features are not corrupted by the very edges of the fragments, which are frequently broken off, eroded, or shadowed. For a few features we also use larger patches — 20 mm in diameter — to estimate properties more accurately and robustly, and for the area-based curvature descriptor the patches are centered on the contour instead of being offset inward. These variants are discussed below, in the descriptions of individual features.

Finally, we employ **per-match** features, which evaluate the plausibility of a candidate pair of fragments in a particular alignment. This stage is the most expensive, since it involves computing features per *pair* of fragments, and precludes the use of fast clustering or indexing methods. On the other hand, such descriptors (e.g., average distance between the fragments) can be more discriminative.

## 5.2.2   Match Classification

There are many types of machine learning tasks that can effectively use vectors of features: indexing, match scoring, classification, etc. As mentioned above, in the context of a large-scale fragment matching application we anticipate two main uses. First, in the early stages of matching the goal is to *quickly* determine large sets of potentially-matching pairs of fragments. In the ideal case, this stage would run in $O(n)$ time for $n$ fragments, in contrast to the naive $O(n^2)$ strategy of checking every potential pair. Therefore, we anticipate that indexing and clustering methods are relevant, implying that we would like to determine which feature vectors are far apart, and which are nearby. Later in the pipeline, the rele-

vant task becomes separating matches from non-matches as effectively as possible. This may operate either via classification — predicting whether a proposed pair is likely to be a match or nonmatch — or via probabilistic regression — ordering proposed pairs from most to least likely. Either way, the most likely matches will, in the end, be presented to the user for ground-truth verification, meaning that all of the above strategies are amenable to incorporation in an "online learning" system that incrementally adjusts the importance of different features to adapt to the particular characteristics of each new database.

We therefore have four tasks — indexing, match classification, regression, and online adaptation to per-database feature importance — that all stem from the same set of features. In this theis we present results for match classification experiments, since it is likely that good performance on this task will lead directly to good performance on the others. We adopt an existing technique (decision trees) for producing trained classifiers, and explore classification performance using a cross-validation methodology. In most cases, the trees are simply trained on the absolute value of the difference between feature descriptor values for a pair, but a few cases require a more complex computation to convert the values of feature descriptors into a value likely to be predictive of a *match*. We also examine the typical variation in the different features.

## 5.3   Fresco Datasets

We evaluate our features using scanned frescoes from the archaeological excavation sites at Akrotiri and Kerkrade, as well as a modern-day "synthetic" fresco data set.

### 5.3.1   Akrotiri

As described in Section 1.4, the Theran frescos were discovered at the seaside town of Akrotiri on the island of Thera (modern-day Santorini, Greece). Though otherwise well preserved, the frescos exist today in thousands of fragments. A more complete description
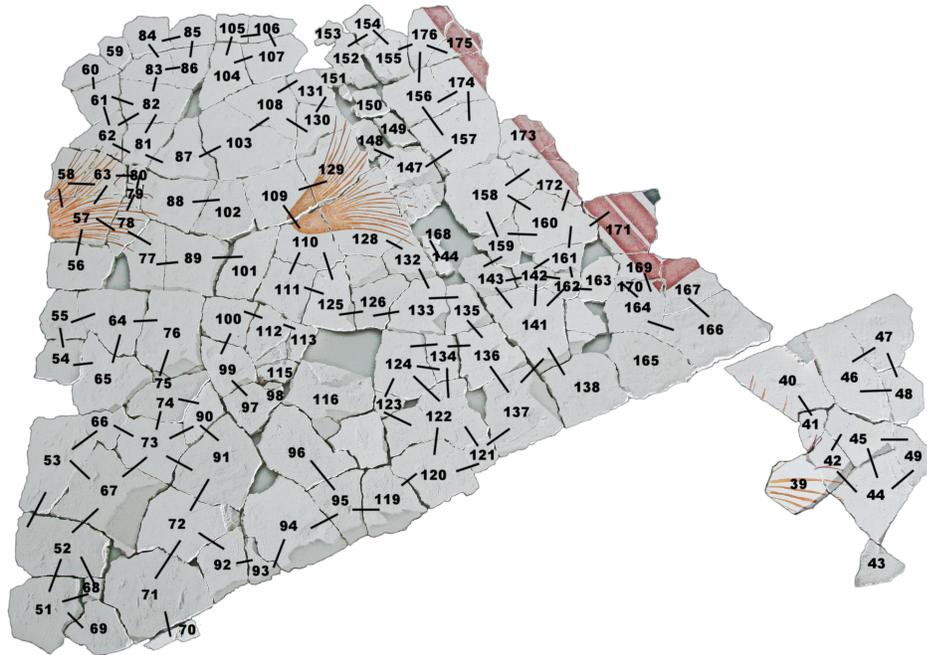
**Figure 5.3:** Synthetic Fresco: In Section 5.8 , we use the Synthetic Fresco from [Brown *et al.*, 2008] as ground-truth for our classification experiments.

of the Akrotiri Wall Paintings, their historical context and the 17th Century volcanic eruption that destroyed them is provided in Section 1.4. For this thesis, we work with a dataset of 1200 fragments taken from a fresco containing spiral motifs. There are visible string impressions and brushstrokes.

## 5.3.2 Kerkrade

The Kerkrade frescoes originate from a second-century Roman villa in Kerkrade, The Netherlands, near present-day Heerlen and Maastricht. They belong to the larger set of quality paintings from the Roman period found in the Netherlands, and are also a part of a select few that depict large-scale human figures. The Kerkrade fragments differ from those at Akrotri in two important ways: they are more eroded, and have visible brush strokes and texture resulting from the smoothing out of the plaster. We therefore, expect a different subset of the features to be important for matching. Our test set consists of 100 fragments.

### 5.3.3   Synthetic

This synthetic fresco, described by [Brown *et al.*, 2008], was created by conservators in a style similar to the one used at Akrotiri. The finished fresco was then broken into pieces to create fragments similar to the fragments found at that site. This fresco is characterized by large areas of white with smaller regions of color. Both string impressions and brush strokes are present on the fragment surfaces. Our ground truth set consists of 127 fragments (Figure 5.3).

## 5.4   Feature Descriptors

Our feature descriptors range from "generic" ones, found in fragment-matching or puzzle-assembly systems, to domain-specific. The latter, while still general, were inspired by visual cues used by conservators and archaeologists to perform manual reassembly of artifacts (such as frescos and pottery). These visual cues are time-tested and effective. The novelty of our normal-based features lies in their use of additional information (fine surface details) to detect these cues when they are not readily apparent.

As explained in Section 5.2.1, our descriptors are classified according to their type (per-fragment, per-patch, per-match) and the data from which they are computed (colors, normals, 3-D). In this section, we focus our attention on normal-based features. We claim that such features combine high classification performance with low acquisition cost and high matching efficiency (i.e., the features are largely per-fragment and per-patch, rather than per-match). These claims will be evaluated in subsequent sections. For completeness, we also describe the more "traditional" features used by our system.

### Average Color, Saturation, and Variance

**Type:** Per-Fragment and Per-Patch   **Data:** Color

We begin with features traditionally used in image-based matching systems, such as the

mean color (computed separately for each color channel) and color variance, both of which may be computed both per-fragment and per-patch. In addition, we use color saturation as a feature. This was inspired by the observation that two adjacent fragments will often exhibit a similar amount of deterioration in their pigments: either they are both faded, or both retain their original colors. We include this descriptor in the hope that it may combine with other features to boost classification performance.

## Contour Curvature

**Type:** Per-Patch   **Data:** Color

The curvature of the fragment's outline provides a *per-patch* descriptor that groups fragments of similar external shape, and is often used for (2-D) puzzle reconstruction. (Similar patches will have curvature of similar magnitude but *opposite* sign.) We have experimented with two alternative descriptors for 2-D curvature along the fragment contour. An *area-based* descriptor finds the fraction of the fragment covered by a circular patch centered on the contour:

$$CurvatureArea = \frac{Area(Fragment \cap Patch)}{Area(Patch)}. \tag{5.1}$$

As shown by [Manay *et al.*, 2004], this quantity is just a function of curvature, in the limit of small patch size: values of 0, $\frac{1}{2}$, and 1 correspond to curvatures of $+\infty$, 0, and $-\infty$, respectively.

The second curvature descriptor only looks locally at three adjacent points $A$, $B$, and $C$ on the contour:

$$CurvatureContour = 2 \frac{\angle(C-B) - \angle(B-A)}{\|C-B\| + \|B-A\|}, \tag{5.2}$$

where the numerator is the angle between the segments $\overline{AB}$ and $\overline{BC}$, and the denominator is their total length. This discrete approximation to curvature is accurate for low-curvature regions, as is generally the case in practice. The points $A$, $B$, and $C$ are picked at multiple scales: 2.5 mm, 5 mm, 8 mm, 10 mm, and 15 mm. Each one yields a separate descriptor,

providing even more information about the contour shape to be used in matching.

## Average Normal and Variance

**Type:** Per-Fragment and Per-Patch   **Data:** Normal

Many datasets that we examined exhibit significant variation in surface roughness from location to location: some regions are smooth while others are rough because of visible brush strokes, weathering, or the use of a different type of plaster. To characterize this, we look at the distribution of normals on the fragment or patch. However, we cannot simply consider the normal vectors themselves: one of their components is not known, in global coordinates, since the final orientation of the fragment is unknown. For this reason, we form a rotation-invariant quantity: the $z$ component of the normals (i.e., the component perpendicular to the fragment's "flat" surface). We use the mean and variance of these $z$ components as features.

## Color/Normal Variation

**Type:** Per-Fragment and Per-Patch   **Data:** Color and Normal

This descriptor captures the effect of *correlated* variation in color and normals, as frequently occurs when there are visible brush strokes or string impressions that were used as guides for painting. We begin by stacking the colors and normal $z$ components for pixels in a fragment or patch into an $n \times 4$ matrix, then perform a Singular Value Decomposition:

$$
\begin{pmatrix} c_{1,r} & c_{1,g} & c_{1,,b} & n_{1,z} \\ & \vdots & & \\ c_{n,r} & c_{n,g} & c_{n,b} & n_{n,z} \end{pmatrix} = U \begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & 0 & \sigma_4 \end{pmatrix} V^T. \tag{5.3}
$$

We use the sum of the $\sigma_i$ as a descriptor, yielding a compact, rotation-invariant, yet discriminative description of the nature of color and normal variation on the surface.
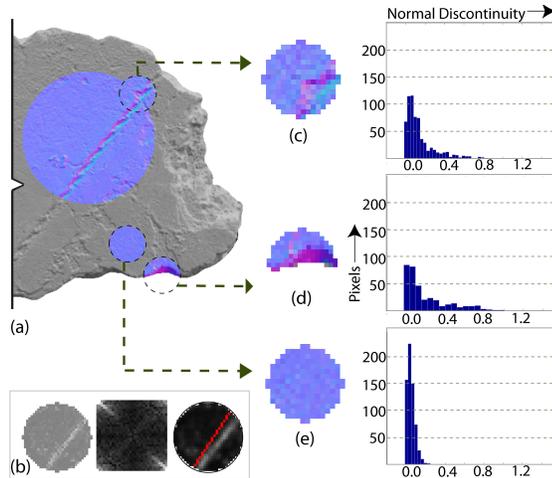
**Figure 5.4:** Normal Discontinuity - String Impressions: **(a)** Fragment with a visible string impression. We visualize the normal-discontinuity and dominant-orientation descriptors. **(b)** The dominant-orientation descriptor detects the orientation of the string impression, allowing it to be matched to an impression on an adjoining fragment. Visualized are the mean component of the normal, the power spectrum, and the detected dominant orientation. **(c-e)** Normal-$z$ histograms used by the normal-discontinuity descriptors. Histogram (c) has a long tail, while the smooth patch in (e) results in a histogram clustered around the origin. We inset patches from the edge to avoid incorrect long-tailed distributions, as shown in (d). We also experimented with larger patch sizes (a).

## Normal Discontinuity

**Type:** Per-Fragment and Per-Patch    **Data:** Normal

Many frescoes contain distinctive shapes on the front surface. For example, artists occasionally pressed a guide string on the wet plaster during fresco construction to provide an outline for straight bands of color. We posit that such features are strong matching cues and design descriptors that distinguish between relatively smooth patches (possibly with some noise in the normals due to erosion) and those with large discontinuities in the surface orientation.

Our analysis begins by computing a histogram of the differences $\{D_{ij}\} = \{\|n_{i,z} - n_{j,z}\|\}$ between neighboring normals in a patch. (We use only the $z$ components of the normal vectors because they are invariant to rotation in the plane.) We then compute a number of statistical measures on this distribution, to determine the degree to which it either is strongly peaked around zero or contains long "tails" of high curvature. We have

51

experimented with four measures to characterize the degree to which the normal difference distribution exhibits long tails:

1. The ratio between the normal differences' 80th and 50th percentiles;

2. The fraction of discontinuity values greater than a threshold (0.46), determined experimentally by looking at fragments containing string impressions;

3. The third moment of the distribution; and

4. The fourth moment of the distribution.

Figure 5.4, (c) and (e), shows normal discontinuity distributions for patches with and without string impressions. The presence of strings produces a distribution with significantly more large values.



Normal Map $\qquad \Omega_n = \arg\max_{\omega \neq 0} \|F_n(\omega)\|^2 \qquad \Omega_\theta$
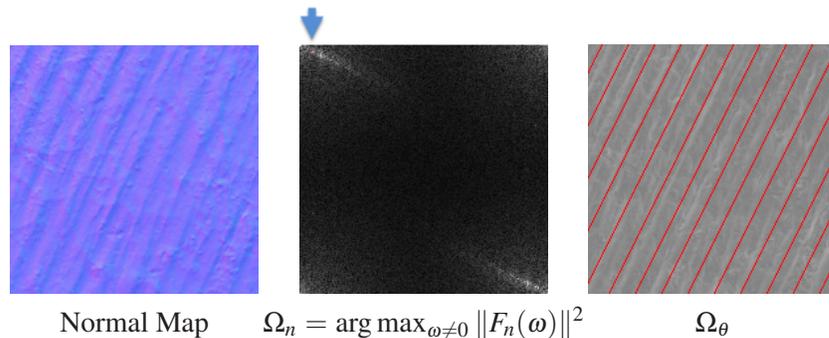
**Figure 5.5:** Brush Stroke Detection: Finding the dominant orientation on a sample taken from Kerkrade. We determine the strength, and direction of brush strokes by examining the frequency, amplitude and orientation of the dominant frequency (excluding the dc) in the power spectrum. **Left:** normal map. **Middle:** power spectrum. **Right:** red lines visualize the orientation of this dominant peak.

## Dominant Orientation

**Type:** Per-Fragment and Per-Patch   **Data:** Normal

This feature detects regular surface patterns, such as the brush strokes found on some fragments. We were inspired by examples like the ones in Figure 5.11, which illustrates the variation in strokes left by both paintbrushes and tools for smoothing out the underlying

plaster. We expect that the amplitude and frequency of these brush strokes extracted from normal maps, will strongly group them according to local variation in their characteristics. In addition, the orientation of these brush strokes must be continuous across fractures, producing a strong matching cue that essentially eliminates the search over orientation in possible matching fragments.

First we smooth the *z*-normal image of the fragment or patch, apply a Hanning window, and then find its 2-D Fourier transform $F_n(\omega)$. We then search for the frequency with the greatest energy. Because the patches are smooth, the highest-energy peak is usually the DC component, and low frequencies are generally stronger than high ones. For this reason, we apply a threshold to the frequency:

$$NormalDominantFrequency = \arg \max_{\omega > \omega_{min}} \|F_n(\omega)\|^2 \qquad (5.4)$$

where: $\omega_{min} = 3$ pixels. In addition to using the frequency and amplitude of this peak as features, we also use its orientation. To make this invariant to rotation, we compute the difference between the angle of the dominant peak and the normal to the fragment contour. In other words, we store the angle between the dominant directional variation (e.g. brush strokes) and the fragment edge — this quantity is expected to be the same for a matching fragment. Figure 5.5 demonstrates results of this process on a fragment from the Kerkrade dataset (see Section 5.3.2) containing strong brush strokes.

## Cracking and Erosion

**Type:** Per-Fragment and Per-Patch    **Data:** Color and Normal

Erosion of a plaster fresco frequently results in small pits in the surface, while the destruction of the original wall-painting produces an irregular pattern of cracks. The erosion descriptor uses morphological operators to quantify the degree of deterioration on the surface. Specifically, we extend versions of the black and white top-hat transforms [Serra, 1983]: where $\circ$ and $\bullet$ are image-morphological open and close operators and $\kappa$ is a struc-
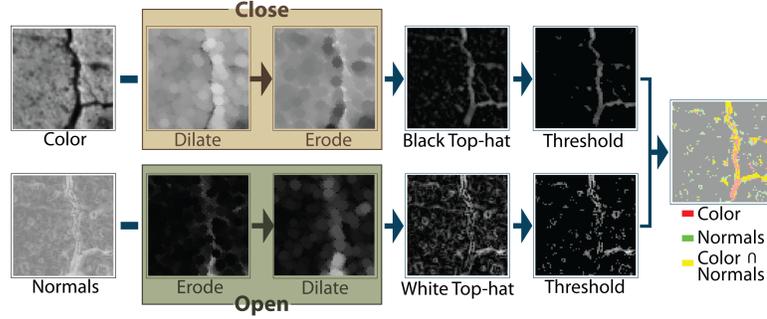
**Figure 5.6:** Erosion Detection: The black top-hat transform is generated by applying a closing operation to the color image and subtracting the result from the original color image. The white top-hat transform is generated by applying an opening operation to the normal map and then subtracting it from the original normal map. The final erosion map is the intersection of the two top-hat transforms and thresholding to reduce noise.
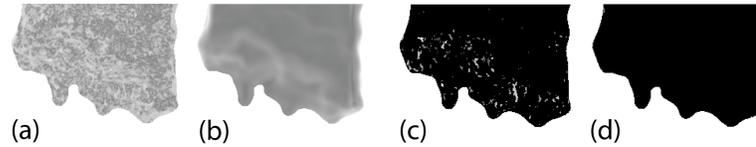


**Figure 5.7:** Comparing erosion detection on 2D normal maps and 3D geometry. Note: We only show the z component of the normals. (a) The 2D normal map acquired with a flat-bed scanner has more detail than the smoother normals in (b) acquired with a 3D scanner. (c) Areas of erosion, highlighted in grey, are detected on the 2D normal map, but no erosion is detected on the 3D geometry (d).

turing element.

Unlike previous applications, we improve discriminability by combining color and normals in the analysis. As shown in Figure 5.6, for patch-level erosion detection, we take the intersection of peaks in the threshholded black top-hat transform of the color map and the threshholded white top-hat transform of the normal map: where $I_c$ is a grayscale version of the color buffer, $I_n$ is the $z$ component of the surface normal, and $T'$ are the thresholded top-hat transforms. We found that an intensity threshold of 0.3 works well. For the structuring element $\kappa$, we use a circle (with a 3 pixel radius) to ensure our results are rotation invariant across fragments. We record two scores: the total number of pixels over all peaks (normalized by the number of visible pixels) and the average number of pixels per connected component. The first value records the density while the latter suggests the average size of each element. Figure 5.9 depicts the use of both color and normals for erosion detection.

At the fragment level, we use a multi-scale approach. In this case, we use a structuring element with ten different diameters and take the color black-top transform to be the sum of the transform taken across the ten scales. Similarly, the white top-hat transform is the sum of the white top-hat transform for the normal map over the ten scales. We chose this approach over a single large scale because features are less likely to expand beyond their boundaries. Multi-scale morphological scale spaces work locally, are good at separating features from uneven backgrounds and do not exhibit the blurring across features characteristic of Gaussian kernels. Figure 5.7 shows an example of fragment level erosion. Although there is some normal variation and noise over the entire surface, we combine color, normals and thresholding to ensure we only detect variations due to erosion (the cracks and pits near the edges).

## Thickness

**Type:** Per-Fragment and Per-Patch   **Data:** 3-D

In some datasets, the thickness of fragments varies considerably from location to location. Therefore, we use the 3-D scan to determine the fragment's thickness at every point. We use the average thickness, per fragment or per patch, as a descriptor. Note that in this case it is especially critical that we offset each patch away from the edge of the fragment, since the estimated thickness is likely to be unreliable near the edges.

## Ribbonmatcher Error and Volume Intersection

**Type:** Per-Match   **Data:** 3-D

To compare the performance of our per-fragment and per-patch descriptors to more descriptive per-match features, we look at two values computed via the brute force "ribbon-matching" approach of [Brown *et al.*, 2008]. For both values, the optimal alignment of the two fragments at the patch locations is computed using the ribbon matcher with a 12.5 mm strip width. The first value ("ribbon error") we consider is the mean-squared distance be-

tween fragments along this strip, as computed by the ribbon matcher. The second value ("volume intersection") measures the amount of interpenetration between fragments. A vertical plane is oriented along the matching edge, and the interpenetration of the two fragments is sampled on this plane. We compute the average of all squared lengths that exceed 1 mm. The intuition behind this descriptor is that correctly matching fragments may have some slight interpenetration due to sampling error, erosion (which affects the alignment), and accretions on the fragment edge when it was scanned. However, correctly matching fragments should not have any *substantial* interpenetration. Considering only interpenetrations greater than 1 mm accounts for "explainable" interpenetration, and squaring the distances penalizes deep intersections more than shallow ones. Note also that the volume descriptor considers the entire fragments, not just selected patches or ribbonmatcher strips.

## 5.5 Case Studies: New Features

Several of the features we consider are motivated specifically by the fresco-matching application, rather than being "generic" features applicable to a variety of shape matching problems. Here we consider a few of these, and present anecdotal evidence for their performance.

While we have found that color cues are meaningful for some datasets and perform no better than chance on others, we have uniformly observed that normal-based features provide reasonable performance. For example, Figure 5.8 shows the best match found
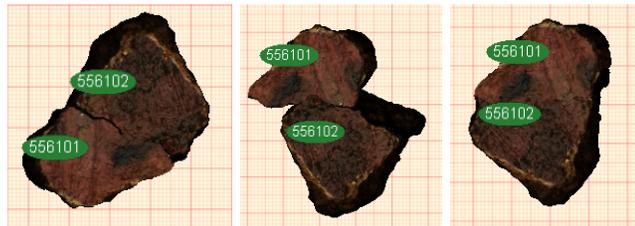


**Figure 5.8:** Kerkrade - Match Found Using Normal-Based Features: **Left:** ground-truth match between two fragments of the Kerkrade fresco. **Center:** best match, only considering this pair of fragments (at all orientations) and color-based features. **Right:** best match, considering normal-based features — notice that the correct match was found.

between a pair of fragments from the Kerkrade fresco using only color cues (center) and using normal-based features (right); only the latter is correct. We hypothesize that even in datasets that exhibit considerable color variation throughout the fresco, the variation within a single fragment is usually insufficient to yield the correct alignment.

Thus, we suggest that normal-based features combine the computational efficiency and ease of use of conventional 2-D features while improving upon classification performance in many cases.

We analyzed our erosion detection features on a number of fragments exhibiting strong cracking, strong color variation without erosion, and mild erosion (Figure 5.9). In each case we show the results of running our top-hat operators on the colors and normals, as well as their intersection. We found that the intersection of top-hat transforms applied to both the colors and normals gave good sensitivity to detection of cracking and erosion. At top, the normals detected all of the cracks, while the color served to limit sensitivity to additional normal variation. At center, the lack of variation in the normals successfully suppressed the detector in areas of color detail. At bottom, this white-colored fragment had only a few small pits, which were successfully detected in both colors and normals. Figure 5.10 shows a photograph of the Kerkrade Fresco fragments before they have been removed from the ground.

One of our frescoes — Kerkrade — exhibited strong variation in the types of brush strokes that were present. We observed a number of phenomena, including small strokes left by the artist's brush (Figure 5.11, left) and broader, deeper strokes left in the underlying plaster when smoothing it out (center and right). We also observed situations in which Brush strokes at different orientations were simultaneously visible. Though our current method does not detect these, returning only the strongest brush stroke direction present in a fragment or patch, we believe that it would be possible to extend the descriptor to handle these cases. In cases in which brush strokes are present, we informally observe the orientation of these strokes to be one of the strongest matching cues available. Fig-
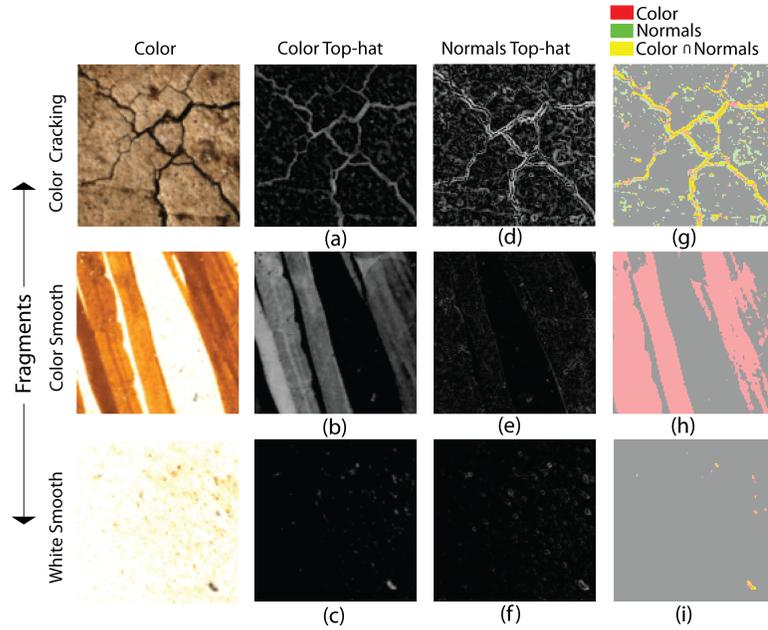
**Figure 5.9:** Erosion Detection - Combining Color and Normals: When computing erosion, we intersect the top-hat transforms of the color and normal maps to avoid capturing sharp variations in color due to dirt or stains, and high frequency noise in the normal maps. (a - c) Black top-hat transforms of color maps for three fragments (color and cracking, color and smooth, white and smooth). (d - f) The white top-hat transforms of the corresponding normal maps. (g - i) Intersection of black and white top-hat transforms. **Top:** Erosion is detected when there are cracks and pits in both the color and normal maps as shown by the yellow pixels in (g). **Middle:** No erosion is detected on the smooth fragment with color. There are no green or yellow pixels in (h) but several red pixels representing color detected by the black top-hat operator. **Bottom:** Only a few pits are detected on the smooth white fragment.

ure 5.12 depicts how the brush stroke orientation remains consistent across the boundary



**Figure 5.10:** Erosion in the Kerkrade Frescoes (photo by Lara Laken).
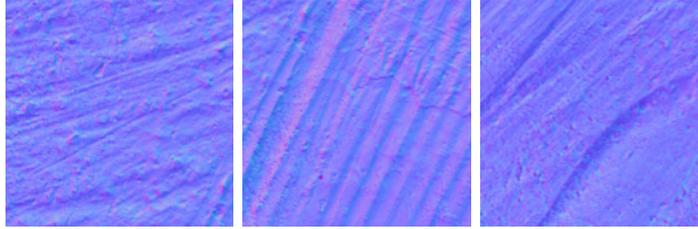
**Figure 5.11:** Kerkrade Brush Stroke Variations: Normal maps illustrating the wide variation in brush strokes on the Kerkrade fresco, including strokes left by a paintbrush (left) and strokes left when smoothing out the underlying plaster (center and right). The direction of these strokes (the angle between their dominant orientation and the fragment contour) is a strong matching cue. The amplitude and frequency also help distinguish between these three types of strokes.
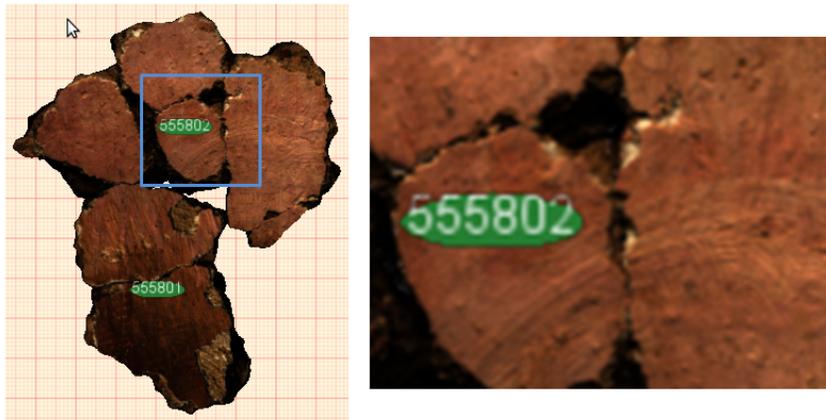


**Figure 5.12:** Brush Stroke Orientation: We observe that brush stroke orientation is consistent across the boundary contour between matching fragments. Our goal is to group fragments by brush stroke orientation, effectively reducing the space of possible matches.

of matching fragments.

## 5.6 Observations

In this section, we analyze the distributions of feature values across the three fresco databases introduced in Section 5.3. We observe that the importance of different features for discriminating matches from non-matches is different for each database, motivating the classifier-based evaluation methodology presented in Section 5.8.

Table 5.1 shows statistics for the feature correspondences computed on each fresco. In this analysis, we use 2,274 pairwise feature correspondences from each dataset, including

**Table 5.1:** Statistics of correspondence values (typically absolute values of differences of feature descriptor values) for random matches and non-matches in the "Synthetic," "Akrotiri," and "Kerkrade" datasets.

| Feature | Synthetic (min / mean / max / stdev) | Akrotiri (min / mean / max / stdev) | Kerkrade (min / mean / max / stdev) |
|---|---|---|---|
| **Color** | | | |
| ColorAvgR | 0.000 / 0.054 / 0.978 / 0.157 | 0.000 / 0.170 / 0.613 / 0.131 | 0.000 / 0.135 / 0.505 / 0.097 |
| ColorAvgG | 0.000 / 0.071 / 0.977 / 0.186 | 0.000 / 0.158 / 0.581 / 0.126 | 0.000 / 0.077 / 0.277 / 0.058 |
| ColorAvgB | 0.000 / 0.092 / 0.968 / 0.199 | 0.000 / 0.122 / 0.589 / 0.103 | 0.000 / 0.040 / 0.174 / 0.034 |
| ColorVariance | 0.000 / 0.040 / 0.628 / 0.083 | 0.000 / 0.026 / 0.167 / 0.022 | 0.000 / 0.008 / 0.085 / 0.009 |
| ColorSaturation | 0.000 / 0.074 / 0.883 / 0.160 | 0.000 / 0.131 / 0.589 / 0.096 | 0.000 / 0.084 / 0.410 / 0.078 |
| **Curv** | | | |
| CurvatureContour | 0.000 / 0.032 / 0.382 / 0.033 | 0.000 / 0.098 / 0.572 / 0.085 | 0 000 / 0.089 / 0.513 / 0.076 |
| CurvatureArea | 0.000 / 0.069 / 0.414 / 0.059 | 0.000 / 0.069 / 0.410 / 0.058 | 0.000 / 0.074 / 0.445 / 0.061 |
| **Normal** | | | |
| ColorNormalVariance | 0.000 / 0.051 / 0.896 / 0.098 | 0.000 / 0.036 / 0.227 / 0.029 | 0.000 / 0.036 / 0.242 / 0.035 |
| NormalMeanZ | 0.000 / 0.038 / 1.155 / 0.067 | 0.000 / 0.143 / 0.634 / 0.118 | 0.000 / 0.074 / 0.383 / 0.064 |
| NormalVariance | 0.000 / 0.056 / 0.483 / 0.073 | 0.000 / 0.064 / 0.712 / 0.064 | 0.000 / 0.059 / 0.301 / 0.049 |
| NormalDiscont8050Ratio | 0.000 / 0.704 / 8.000 / 1.198 | 0.000 / 0.406 / 2.000/ 0.361 | 0.000 / 0.590 / 4.000 / 0.606 |
| NormalDiscontThresholded | 0.000 / 0.053 / 0.833 / 0.109 | 0.000 / 0.201 / 0.893 / 0.149 | 0.000 / 1.086 / 0.584 / 0.098 |
| NormalDiscontThirdMoment | 0.001 / 1.665 / 9.649 / 1.504 | 0.001 / 0.389 / 5.085 / 0.575 | 0.000 / 1.171 / 3.926 / 1.029 |
| NormalDiscontFourthMoment | 0.004 / 15.647 / 147.952 / 18.466 | 0.000 / 1.746 / 50.572 / 4.222 | 0.001 / 3.546 / 61.146 / 4.760 |
| NormalDominantFrequency | 0.000 / 0.111/ 1.394/ 0.227 | 0.000 / 0.711 / 8.314 / 0.914 | 0.000 / 0.301 / 2.798 / 0.358 |
| NormalDominantOrientation | 0.000 / 0.727 / 2.931 / 0.549 | 0.002 / 0.826 / 3.028 / 0.649 | 0.002 / 0.767 / 3.033 / 0.593 |
| NormalDominantAmplitude | 0.000 / 0.007 / 0.096 / 0.001 | 0.000 / 0.015 / 0.083 / 0.012 | 0.000 / 0.012 / 0.007 / 0.010 |
| ColorNormalErosionDensity | 0.000 / 0.01 / 0.211 / 0.023 | 0.000 / 0.007 / 0.089 / 0.007 | 0.000 / 0.004 / 0.087 / 0.010 |
| ColorNormalErosionShape | 0.000 / 0.006 / 0.211 / 0.016 | 0.000 / 0.036 / 0.24 / 0.037 | 0.000 / 0.002 / 0.065 / 0.005 |
| **3D** | | | |
| Thickness | 0.001 / 1.392 / 8.432 / 1.399 | 0.002 / 2.953 / 12.821 / 2.221 | 0.000 / 4.700 / 22.199 / 4.022 |
| RibbonError | 0.003 / 8.412 / 282.209 / 16.806 | 0.000 / 11.701 / 136.566 / 14.536 | 0.082 / 15.742 / 112.204 / 16.765 |
| RibbonVolIntersection | 0.000 / 253.265 /5681.626 /648.324 | 0.000 / 293.093 / 14545.190 / 804.263 | 0.000 / 151.188 / 3183.708 / 308.539 |
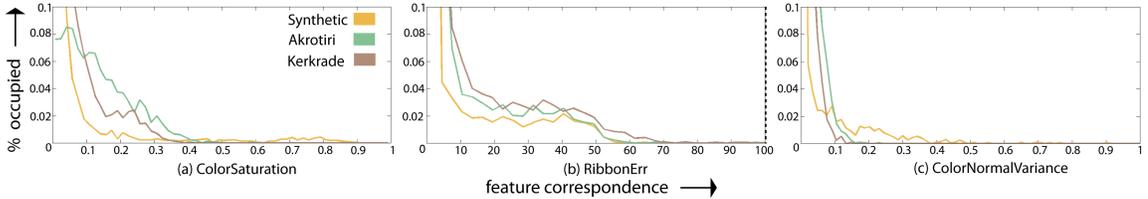


**Figure 5.13:** Comparing the distribution of feature correspondences across multiple datasets. Histogram plots show the percentage of correspondences with the specified feature correspondence ranges for the Synthetic, Akrotiri and Kerkrade frescos. 2,274 correspondences consisting of both matches and non-matches were taken from each dataset. (a) There is more variation in correspondences for ColorSaturation for Akrotiri and Kerkrade than the Synthetic dataset which is primarily composed of white fragments. (b) The RibbonErr correspondences are similar across all three datasets. (c) The shape of the curves show that ColorNormalVariance correspondences are more similar for Akrotiri and Kerkrade than the Synthetic dataset. The variation of correspondence ranges across the different datasets suggests that re-weighting the contribution of individual features based on observed statistics of the dataset would adaptively improve match retrieval for a specific dataset.

both ground-truth matches and randomly sampled non-matches. All patches have a 10mm diameter and each patch center is offset 8mm from the boundary contour (except for curvature descriptors, which are sampled along the boundary contour). The values shown are the minimum, mean, maximum, and standard deviation of "correspondence" values for each feature. For most features, this is just the absolute value of the difference between the

feature values computed on both fragments: these are expected to be near zero for correct matches. For a few features, however, the correspondence value is the absolute value of the sum for the two fragments. This is necessary for features such as curvature, which are expected to have opposite signs on corresponding fragments.

As shown in Figure 5.13, the distribution feature correspondences is different for each fresco. For example, almost all ColorSaturation correspondences (a) for the Synthetic fresco are near zero while only 8% of the samples are clustered near zero for Akrotiri. This is not surprising, as most Synthetic fragments are white. The distribution of the RibbonError feature (b) is similar for each dataset while the range of the ColorNormalVariance feature (c) is most similar for Akrotiri and Kerkrade but different for the Synthetic fresco.

We also observe that different features are more *discriminating* on different datasets. Figure 5.14 compares histogram plots of selected features for ground-truth matches (blue) vs. non-matches (red) for the Synthetic and Akrotiri datasets. There are equal numbers of matches in each example, with 2,274 total samples for the Synthetic dataset and 206 total samples for the Akrotiri dataset. The match and non-match curves for ColorSaturation are almost identical on the Synthetic dataset, suggesting there is little information for distinguishing between a match and a non-match using this feature. Conversely, the number of matches whose correspondence lies near zero is significantly greater than non-matches for Akrotiri. This is expected, as this dataset contains considerable pigmentation and hence color is a good matching cue. We also observe that there is a clear threshold above which correspondence values are more likely to apply to non-matches than matches. In many cases, 100% of correspondences above the threshold are non-matches.

Our analysis suggests that color and 3-D features work best for the Akrotiri and Kerkrade databases and that, in general, normal-based features will work well on all databases. Some normal features are significantly stronger than others, however, depending on the surface features of the database. For example, we observed that dominant-orientation features are especially important for the Kerkrade dataset, with color-based features performing no bet-
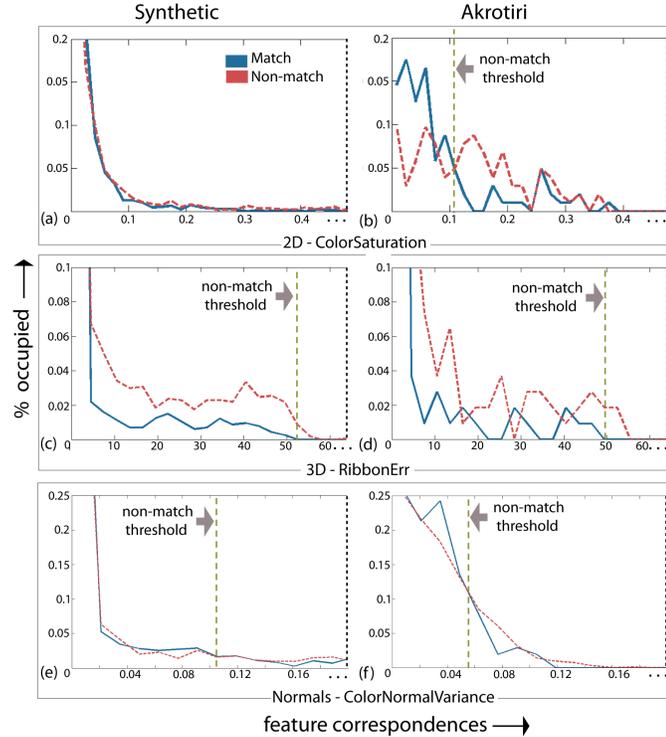
**Figure 5.14:** Feature discriminability. Feature correspondences for equal numbers of matches and non-matches for the Synthetic (2,274 samples) and Akrotiri (206 samples) datasets. (a) - (b) ColorSaturation is more discriminating on the Akrotiri dataset. The number of matches with correspondences near zero are greater than that of non-matches. Conversely, there is little distinction between the two curves for the Synthetic dataset. (c) - (d) RibbonErr correspondences greater than 50 belong to non-matches for both datasets. (e) - (f) ColorNormalVariance values greater than 0.05 on Akrotiri are more likely to be non-matches. Synthetic values greater than 0.10 are more likely to be non-matches (with less obvious distinction).

ter than chance. Curvature features are the least reliable.

We also compared classification results for each feature individually and in combination with other features. Individual features do not perform as well as combinations of features. We found that Patch level features are more robust than fragment level features. One exception to this rule is erosion, which was more effective when computed over the entire fragment. We anticipate that this feature is good at separating smooth fragments from eroded ones at the fragment level, but is too noisy at the patch level. In Section 5.8, we discuss our classification approach, results from combining features, and comparisons of patch level vs. fragment level features.

## 5.7 k-fold Cross Validation

Before we present our classification results, we provide background information on a basic machine learning concept, *k-fold cross validation*. Cross validation [Bishop, 2006] is a method of evaluating how well the results of a statistical analysis on one dataset will generalize to another independent dataset. In our case, we would like to determine how well a classifier trained on a set of observations (feature vectors) from one of the datasets in Section 5.3, will perform in practice. The *k-fold cross validation* process first partitions the dataset into parts - a training set and test sets. The classifier trains a predictive model by observing labeled samples in the training set, and then applies this knowledge to predict outcomes in the test set. In order to reduce variability, the process is iterated across multiple rounds, with the training and test partitions changing on each round. The final performance takes into consideration some measure of the performance across all rounds (for example the average). In our case, we use 10-fold cross validation. In this scenario, 90% of the data is used for training while 10% is held back for testing. This process is iterated over 10 rounds, with repartitioning of the training and test sets each time. This method of evaluation is particularly useful when there are only a small number of samples to consider.

There are several pitfalls of cross validation that must be avoided. First, this method is prone to overfitting. This occurs when the model performs well on the test cases but poorly on samples it has never seen. This occurs because the model is optimized well to its test sets but must be evaluated on new data that it is not optimized for. In addition, special care must be taken to ensure there is no leakage between the training and test sets. This results in an unfair advantage as the model has seen examples of the test set already in the training set. For the case of fresco fragments, we must also insure that we take samples uniformly and randomly across the entire fresco. This will prevent the classifier from becoming biased to the characteristics of any one region or cluster. We must also be careful to ensure that the classifier observes sufficient information to learn how to identify matches. This is difficult

when the amount of non-match correspondences considerably out-number those of match correspondences.

## 5.8 Classification Results

In Sections 5.2.2 and 5.7 we gave a general overview of the machine learning concepts that are applicable to our research. In this section, we explain the specifics of our evaluation methodology, including a description of our experiments, the algorithms we incorporate, and the characteristics of the training and test sets on which we operate.

### 5.8.1 Evaluation Methodology

In order to evaluate the performance of our features for matching, we work with manually labeled sets of matching fragments, and randomly sampled non-matches. Except where stated otherwise, we use 10-fold cross-validation, with manually separated training and test sets. In each of these sets, we ensure that each *pair* of fragments, whether matching or not, is placed entirely within either the training or test set. This is done because a single matching pair of fragments may result in multiple matching *patches*, so we wish to ensure that the classification algorithms do not gain advantage from training and testing on patches from the same pair of fragments. Using various subsets of the color, normal, and geometric features described above, we train classifiers to distinguish between matches and non-matches.

We explored four classification algorithms, as implemented by the "Weka" open-source data mining package.[1] The algorithms are:

- **J48 decision trees** implement the C4.5 algorithm of Quinlan [Quinlan, 1993]. This algorithm hierarchically subdivides the training set, at each node partitioning using the feature that results in the greatest difference in entropy among the subsets.

---

[1] http://www.cs.waikato.ac.nz/ml/weka/

- **Random forests** train decision trees on multiple subsets of features, combining the results into a single probabilistic classifier.

- **Support vector machines** compute a high-dimensional plane that separates data optimally between two categories.

- **Logistic regression** fits the data with a generalized linear model consisting of the logistic function $p = 1/(1 + e^{-z})$ applied to a linear combination of the input feature values.

To determine which classifier would generalize well to all of our datasets, we evaluated robustness to overfitting, computational efficiency, and the availability of a real-valued probability instead of merely a binary yes/no classification. The latter is important for our application, since it allows us to create a rank-ordered list of hypothesized matches, which is then presented to a human for verification. Since it will typically be impractical for a person to check all predicted matches, the availability of a ranking is crucial.

Table 5.2 shows the performance of each classifier using all of our per-fragment, per-patch and per-match features for 2,274 groundtruth samples (with equal numbers of matches and non-matches). The J48 decision trees had good matching performance but were more prone to over-fitting the data. In addition, they provide only a binary decision, not a probability. Random forests were less prone to over-fitting and provided probabilities, but the probability values (from combining multiple trees) were still strongly clustered. Support Vector Machines worked well and gave meaningful probabilities, but exhibited time, space and algorithmic complexities that are impractical for large datasets. They were also sensitive to parameter selection, which was often difficult due to the wide range of meaningful values for each feature. Logistic regression had lower computation costs, proved more resistant to over-fitting and required no manual parameter adjustments. It produced meaningful rankings for precision and recall and experimental analysis. Although logistic regression has a high false-positive rate, its collective advantages far outweigh other

**Table 5.2:** Comparison of machine learning algorithmns: We evaluate the performance of each algorithm using manual cross validation on 2,274 groundtruth samples containing an equal number of matches and non-matches. In this example, we combine all per-patch, per-fragment and per-match features.

| Classifier | Ground-truth Matches | | Ground-truth Nonmatches | |
|---|---|---|---|---|
| | Correct (TP) | Incorrect (FN) | Correct (TN) | Incorrect (FP) |
| J48 | 66% | 34% | 84% | 16% |
| RandomForest | 79% | 21% | 71% | 29% |
| SVM | 67% | 33% | 78% | 22% |
| LogisticRegression | 84% | 16% | 49% | 51% |

**Table 5.3:** Classification performance on synthetic fresco, on a test set of 220 samples using logistic regression. There are 110 matches (one pairwise match per matching fragment pair) and 110 non-matches (also unique and randomly sampled. We apply the best model from our manual cross validation training session.

| Features | Ground-truth Matches | | Ground-truth Nonmatches | |
|---|---|---|---|---|
| | Correct (TP) | Incorrect (FN) | Correct (TN) | Incorrect (FP) |
| AllColor | 79% | 21% | 31% | 69% |
| AllCurvature | 54% | 46% | 62% | 38% |
| AllNormal | 80% | 20% | 48% | 52% |
| Thickness | 80% | 20% | 37% | 63% |
| RibbonError | 86% | 14% | 68% | 32% |
| RibbonVolIntersect | 94% | 6% | 35% | 65% |
| AllCombined | 90% | 10% | 78% | 22% |

approaches. This method is used for the remaining results in this section.

## 5.8.2 Performance: Synthetic Fresco

We evaluate the performance of classifiers trained on different categories of features, on a set of ground-truth matches and non-matches from the Synthetic fresco. Because this is the dataset with the greatest number of known matches, we expect to learn the most meaningful results about feature performance by observing classification results on this fresco.

This test was conducted on 110 known matches and 110 known non-matches from this fresco. For maximum fairness, we only include one pair of matching or non-matching patches for each fragment pair. Table 5.3 shows the number of correctly and incorrectly classified instances among the matches (true positives and false negatives) and among the

non-matches (true negatives and false positives). The table rows represent classifiers trained on:

- All "color" features listed in Table 5.1. Color and curvature are considered by many traditional 2-D-only matching algorithms.

- All "curvature" features listed in Table 5.1, evaluated at all scales.

- All "normal" features listed in Table 5.1. These are the new per-fragment and per-patch features we propose.

- The fragment thickness.

- The RibbonError and RibbonVolIntersection features, which represent two outputs computed by the ribbon-matching algorithm of [Brown *et al.*, 2008] on the 3-D models.

- A combination of all features listed in Table 5.1.

The table demonstrates that each type of feature has its strengths and weaknesses when it comes to both finding matches and rejecting non-matches. Curvature and color features, considering both true-positive and true-negative numbers, are barely performing above chance. The normal-based features perform better on non-matches, confirming our hypothesis that such features, while remaining easy to acquire, easy to compute, and easy to incorporate into a fast pruning stage based on per-fragment and per-patch information, incorporate substantially more information about matching fragments than does color.

Turning to 3-D information, thickness performs moderately well, but not as well as normal-based features. The two ribbon-matcher features have substantially better performance, but note that these also have substantially higher computational cost: they are per-*match* features, not per-fragment or per-patch.

Finally, the combination of all features has the best overall performance, demonstrating that the classifier is successfully taking advantage of the best performance of each.
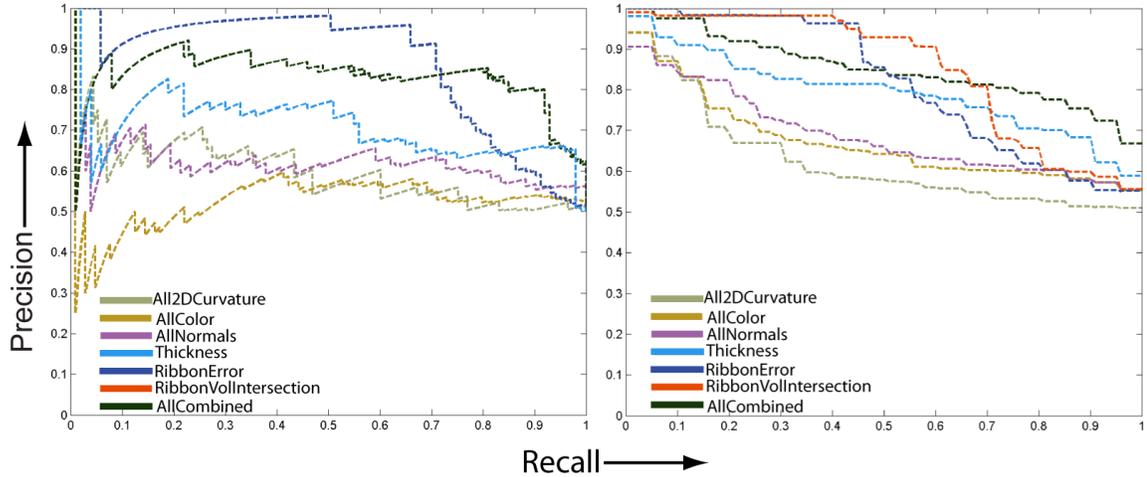
**Figure 5.15: Left:** Precision-recall for a classification experiment on the Synthetic fresco using the classes of features shown in Table 5.3. **Right:** A re-weighting experiment. 203 locations of potential matches, predicted by the ribbon matcher, are ranked by a trained classifier. Averages over 10-fold cross validation are presented.

### 5.8.3 Precision-recall

Because the logistic regression classifier outputs not only a prediction but also a probability, we are able to evaluate our results on a "ranking" task that provides more insight than is available with simple confusion matrices. We present our results using precision-recall curves, in which points represent predicted matches in probability-ranked order, with the $x$ coordinate (recall) representing the fraction of total matches found so far while the $y$ axis (precision) indicates the fraction of all predictions so far that have corresponded to true matches. Higher curves therefore represent better results.

Figure 5.15, left, shows results on an experiment similar to the one in Table 5.3, using the same sets of features. At right, we show a different way of using classifiers, namely a re-weighting experiment in which 203 locations of matches predicted by the existing "ribbon matcher" are ranked according to a classifier trained on different subsets of features. Average results for 10-fold cross validation are presented. In both cases, the results show that combining features leads to better precision than most individual features, especially at higher recall.
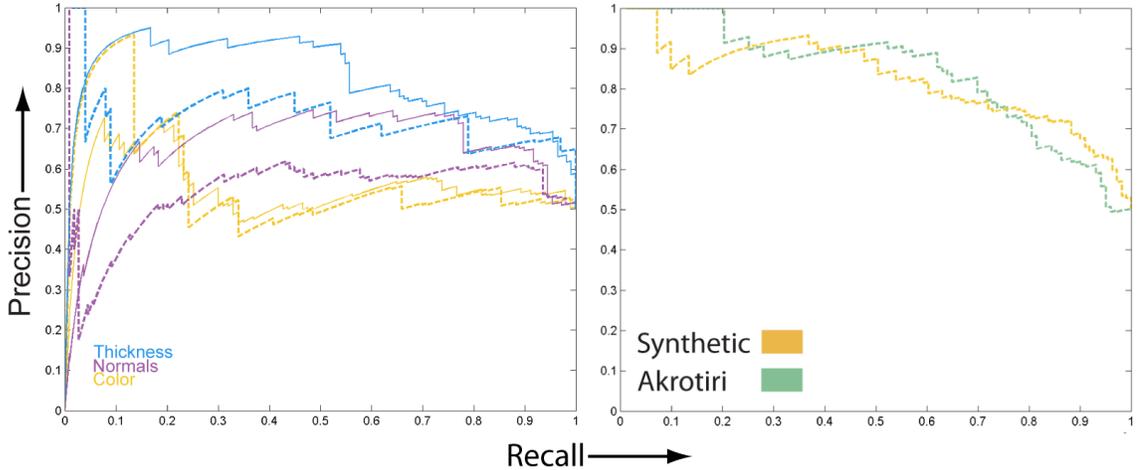
**Figure 5.16: Left:** Comparison of per-fragment features (dashed lines) with a combination of per-fragment and per-patch features (solid lines). Adding per-patch features always improves classification performance, but increases computation time. **Right:** Evaluating classifiers on multiple datasets. A model was trained on the Synthetic dataset using a combination of features: ColorSaturation, RibbonError and ColorNormalVariance (features used in Figure 5.14). We show results of the trained classifier on the original (Synthetic) dataset, and the same classifier on the Akrotiri dataset.

## 5.8.4 Per-Fragment vs. Per-Patch Features

To further examine the potential performance of the pipeline in Figure 5.2, we investigated the performance of per-fragment and per-patch features. Figure 5.16 left shows precision-recall curves where dashed lines include only per-fragment features, while solid lines include both per-fragment and per-patch features. For color-based features, both sets perform relatively poorly, but for normal and thickness features, there is more information available from per-patch features. In these cases, however, per-fragment features alone are still performing some degree of classification, suggesting that the pipeline of Figure 5.2 may, with appropriate thresholds, provide efficiency and accuracy.

## 5.8.5 Generalization Across Datasets

To examine the extent to which classifier performance generalizes across different datasets, we trained a classifier on the Synthetic fresco, using the three features of Figure 5.14. We then compared the performance of the classifier on the original Synthetic fresco, as well

as the Akrotiri fresco (Figure 5.16 right). We observe that the performance is reasonable, with the classifier sometimes performing slightly better and sometimes slightly worse. In general, we expect better performance with custom-trained classifiers for each dataset, but these preliminary results suggest that adapting classifiers from one dataset to another may still lead to reasonable results. In particular, the results are usually sufficient to perform a "bootstrapping": finding enough ground-truth matches to enable a new, custom classifier to be trained. In the future, we expect to use the results of the analysis of variance of each feature across each dataset to be able to adapt classifiers even more directly, by re-weighting the contribution of each feature to the classification without a full re-training step.

## 5.9   Discussion

Manual fragment assembly rarely occurs based on a single cue. Even where there is an obvious and essential cue, such as the edge geometry of fresco fragments, a good assembler relies on judiciously combining every available cue. Working in the context of fresco fragments, we have introduced several new feature descriptors based on normals and color that encapsulate cues such as fragment erosion and surface impressions. We have also shown how to use machine-learning techniques to combine descriptors within a multi-cue framework, including using our new per-fragment and per-patch descriptors to complement existing per-match features.

There is some danger to relying on surface-based features. While matching fragments often erode in similar ways, that is not always the case. When only one fragment has eroded or discolored, we may not identify the match. This is of course inherent in relying on any kind of cue: edge- and contour-based matchers will fail if too much of the side (as opposed to the front) has eroded or broken off, whereas matching on surface properties might still succeed. We believe the best bet is to support many different cues so we can identify as many matches as possible.

Although we have presented our work in the context of fresco fragments, we believe the ideas translate to many other matching problems such as distinguishing the brush strokes of different artists on oil paintings, classifying chisel marks on sculptures, or matching textured objects to their impressions for forensic identification. Different types of objects will naturally require different features, but we expect the normal-based descriptors we have presented will be valuable for many types of material with an exterior surface containing relief or erosion. Our features are simple functions that can be computed on any dataset of normal maps.

# Chapter 6

# Background: Non-Photorealistic Rendering

We have seen the application of multi-channel images for computer-assisted matching of fresco fragments (Chapter 5). Now we address another important challenge faced by cultural heritage professionals; producing accurate and meaningful digital representations of artifacts. These illustrations are important for archaeological study and archival documentation. Although digital cameras and other image capture devices are becoming more prevalent in museums and at excavation sites, our goal is to generate images that convey more information than is readily available in traditional color images. Thus, we focus our attention on the class of non-photorealistic (NPR) [Gooch and Gooch, 2001] rendering algorithms commonly used to create abstract representations of complex 3-D models. As demonstrated by *The Digital Michelangelo Project* (Section 1.3.2) and *The Pietà Project* (Section 1.3.1), full 3-D models are difficult to acquire accurately with high resolution (particularly when the surface geometry is complex with significant occlusion). Conversely, color images lack the surface orientation information required by most NPR functions. Because our RGBN images are easy to capture and encode some 3-D information in the form of surface normals, they provide a practical alternative. In this chapter, we review previous work in

non-photorealistic rendering, with an emphasis on stylization techniques best suited for scientific illustration with multi-channel images.

## 6.1   3D Models

Our research considers two classes of NPR effects previously explored for 3D models: enhanced shading (lighting) models and shape-conveying lines. NPR shading models are often simple functions of the surface normal and light direction that result in effects such as toon shading [Decaudin, 1996], warm-to-cool transitions [Gooch *et al.*, 1998], cartographic hill shading [Horn, 1981], or other artist-specified effects [Sloan *et al.*, 2001]. More complex models are also possible, with some systems employing curvature-based shading to emphasize creases [Kindlmann *et al.*, 2003] and others using light positions carefully selected to increase local contrast [Lee *et al.*, 2006]. The "exaggerated shading" technique increases contrast across scales and for all surface orientations [Rusinkiewicz *et al.*, 2006].

A second class of NPR effects focuses on "sparse," shape-conveying linear features instead of tone-conveying lines or hatch strokes. There are three line styles that occur more frequently. The first set of lines are silhouettes (both interior and exterior) and occluding contours as shown by [Dooley and Cohen, 1990; Elber and Cohen, 1990; Winkenbach and Salesin, 1994; Markosian *et al.*, 1997; Hertzmann and Zorin, 2000]. The second category belongs to ridge and valley lines. These are drawn where there are local maxima of principal curvature magnitude in a principal direction as demonstrated by [Interrante *et al.*, 1995; Thirion and Gourdon, 1996; Pauly *et al.*, 2003; Ohtake *et al.*, 2004]. The third class of lines we consider is suggestive contours ([DeCarlo *et al.*, 2003, 2004]). These lines complement contours to better depict shape and are characteristic of the lines an artist might draw.

All of the above NPR algorithms have been investigated only for 3D models or, in a few exceptions ([Interrante *et al.*, 1995; Ebert and Rheingans, 2000; Kindlmann *et al.*, 2003]), volumetric data. Because of the difficulty of 3-D modeling and acquisition, they have

rarely been applied for the task of illustrating real-world objects. In Chaper 7 we derive the signal processing tools necessary to adapt many of these methods for RGBN images, and demonstrate the algorithms on complex real-world objects.

## 6.2 Images with Discontinuities

One class of NPR algorithms that operate on more than simply colors includes techniques that depict depth discontinuities. These may be located in synthetic renderings [Saito and Takahashi, 1990] or in real-world datasets [Raskar *et al.*, 2004]. Typically, these methods produce "haloing" or "shadowing" effects around depth discontinuities [Rheingans and Ebert, 2001; Raskar *et al.*, 2004; Tan *et al.*, 2004; Luft *et al.*, 2006], leading to a better perception of relative depth. We develop a fast discontinuity shadowing effect in Section 7.3.7, using contours detected by the method of [Raskar *et al.*, 2004].

## 6.3 Images with Normals

There have been several applications developed for normal maps recovered from real-world objects. In some examples, the acquired normals are integrated to obtain positions. An approach by [Nehab *et al.*, 2005] introduced a hybrid algorithm that combines the depth information from a triangulation scanner with the normal information from photometric stereo. However, these approaches are typically unstable without the additional availability of coarse geometry to constrain the reconstruction. Therefore, several applications have retained the normals in the form of an image, using them for rendering. One class involves relighting: an image under arbitrary new lighting is formed by computing a (local) lighting model per-pixel. This approach was used by [Debevec *et al.*, 2000] for cinematic re-lighting. Another technique presented by [Malzbender *et al.*, 2001] fits a polynomial to measurements. The rendering need not be photorealistic. Some work has explored enhancing subtle detail by *reflectance transformation*, i.e. using a (usually highly-specular) ma-

terial that emphasizes deviations in the normals ([Malzbender *et al.*, 2001; Wenger *et al.*, 2005; Malzbender *et al.*, 2006]). This approach has been successful in domains such as archaeology and art history [Mudge *et al.*, 2005, 2006]. [Bartesaghi *et al.*, 2005] have explored a style based on hatching, with the density of hatch marks controlled by image intensity while their direction follows principal direction fields extracted from stereo or photometric stereo. We build upon these algorithms to develop a suite of geometric processing tools for RGBN images, then applying them to a variety of NPR styles.

# Chapter 7

# Non-Photorealistic Rendering Pipeline

In this chapter we investigate the creation of non-photorealistic illustrations from a type of data lying between simple 2D images and full 3D models: images with both a color (albedo) and a surface normal stored at each pixel. Images with normals combine an acquisition process only mildly more complex than that for digital photographs (and significantly easier than 3D scanning) with the power and flexibility of tools similar to those originally developed for full 3D models. We investigate methods for signal processing on images with normals, developing algorithms for scale-space analysis, derivative (i.e., curvature) estimation, and segmentation. These are used to implement analogues of stylized rendering techniques such as toon shading, line drawing, curvature shading, and exaggerated shading. We also introduce new stylization effects based on multiscale mean curvature shading, as well as fast discontinuity shadows. We show that our rendering pipeline can produce detailed yet understandable illustrations in medical, technical, and archaeological domains.

## 7.1 Motivation

Scientific illustrators use a variety of non-photorealistic techniques, including stylized shading and line drawing, to produce detailed yet comprehensible images [Hodges, 2003]. Such

**Figure 7.1:** Left: a photograph of a pinecone. Right: an illustration utilizing exaggerated shading and discontinuity shadows. The latter reveals both subtle details and depth complexity more clearly, as is common in hand-drawn scientific illustrations. Our pipeline, based on acquired images with per-pixel normals, produces such illustrations easily and with high fidelity; it would be difficult to generate equivalent results starting with either a single 2D image or a 3D scan.

illustrations are common in domains such as documenting archaeological excavations, or in textbooks on medicine or botany. However, consider how one might produce an illustration of the pinecone shown in Figure 7.1, right. In contrast to the photograph at left, this illustration reveals detail by emphasizing the small ridges present on each plate (visible when zooming in on the electronic version of this document), and improves overall comprehensibility by introducing consistent soft shadows. While it would be possible to produce this type of illustration by hand, it would require many hours of meticulous drawing and shading. The illustration could not be produced automatically from a photograph, because the effects require more information than is present in a 2D color image. Existing non-photorealistic rendering (NPR) methods would require starting with a 3D model, but capturing a model of the pinecone using a 3D scanner is difficult. This is both because of the significant occlusion present in the pinecone (which would require combining dozens of scans to obtain a hole-free model, even from one viewpoint) and because existing scanners have neither the resolution nor the low noise necessary to capture subtle details such as the ridges (the resolution of the images in Figure 7.1 is approximately 20 pixels per millimeter).

We argue that the most practical method for producing illustrations such as those shown in Figure 7.1, right, is to acquire a type of data lying between simple 2D images and full 3D models: images with a color albedo and a surface normal stored at each pixel ("RGBN

images"). Such datasets may be captured using *photometric stereo* [Woodham, 1980], in which normals are inferred from several images (captured from a single camera position) of an object illuminated from different directions. Moreover, we demonstrate how to derive signal processing (smoothing and derivative estimation) and segmentation algorithms for RGBN images, enabling a broad spectrum of stylized rendering methods. In short, RGBN images combine an acquisition process only mildly more complex than that for digital photographs with the power and flexibility of tools similar to those originally developed for full 3D models (with the obvious and important limitations that RGBN images do not allow for easy change of viewpoint or realistic cast shadows).

RGBN datasets have been acquired in previous projects, such as IBM's digitization of Michelangelo's Florentine Pietà [Bernardini *et al.*, 2002], and researchers including [Debevec *et al.*, 2000] have captured images with normals at sufficiently high quality to be used for cinematic relighting. In this paper we build upon previous work on *realistic* relighting to investigate the generation of *stylized* renderings from RGBN images. Our contributions include:

- tools for segmentation and signal processing on RGBN images, including smoothing and derivative estimation (Section 7.2);

- adaptation of (existing) shading and line drawing styles to RGBN images (Sections 7.3.1 - 7.3.4);

- new non-photorealistic rendering styles motivated by the limitations of RGBN images, namely multiscale mean curvature shading and fast discontinuity shadows (Sections 7.3.5 - 7.3.7);

- demonstration of our pipeline for producing detailed yet understandable illustrations (Section 5.8).

While much of the research in NPR has focused on elision of extraneous detail, this project also addresses the ability to enhance fine detail in an illustration. Thus several of the fig-

ures( 7.1, 7.12, 7.16, 7.18, and 7.19) reveal texture and structure at a fine scale largely unseen in previous work in NPR, yet critical for illustrations in, for example, archeology or botany whose goals are to communicate such detail. We believe that it would be difficult to achieve such effects starting from either traditional photographs or acquired 3D models.

## 7.2 Tools For RGBN Processing

While specific rendering techniques are explored in the following section, many of these algorithms rely on two fundamental signal processing tools — smoothing and derivative estimation — as well as on the ability to effectively segment objects. Therefore, we begin by describing a toolbox of methods analogous to those available in image processing and digital geometry processing. These tools tend to offer the power and control of geometry-based methods, while retaining the simplicity and efficiency of image-based methods.

### 7.2.1 Filtering



Original     Original (close-up)     Large $\sigma_c$, large $\sigma_n$     Small $\sigma_c$, large $\sigma_n$     Large $\sigma_c$, small $\sigma_n$     Small $\sigma_c$, small $\sigma_n$
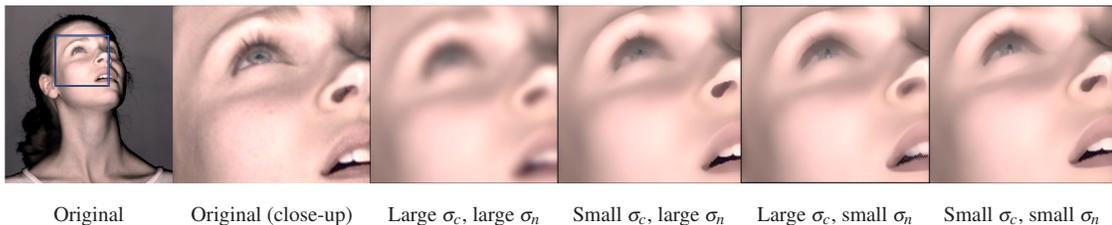
**Figure 7.2:** The RGBN bilateral filter is capable of producing different results, depending on the settings of the domain and range filter widths. For large $\sigma_c$ and $\sigma_n$, there is little edge preservation, and the filter resembles a simple Gaussian. Making $\sigma_c$ small preserves color detail, such as that around the eye, while making $\sigma_n$ small as well preserves both color and geometric edges.

**Gaussian Filtering:** A smoothing operator is a basic building block for many frequency-based methods, including denoising and scale-space analysis (i.e., multi-scale pyramids). The naive method for smoothing RGBN images would be to treat them as plain images with a 6D (color and normal) vector, in place of the conventional 3D color vector at each

pixel. One could then perform smoothing by convolving the image with a Gaussian and finally adjusting the normals to have unit length.

One problem with naive smoothing is due to foreshortening: regions with normals tilted away from the view direction (Figure 7.3 right) will be smoothed more than they should be. With the formulation of smoothing as convolution, the naive method underestimates the "area" allocated to each pixel by $\cos \theta$, where $\theta$ is the angle between the normal and the view direction. Alternatively, in the linear diffusion formulation of smoothing, the distance between adjacent samples is underestimated by a factor of $\cos \theta$, hence the rate of diffusion is over-
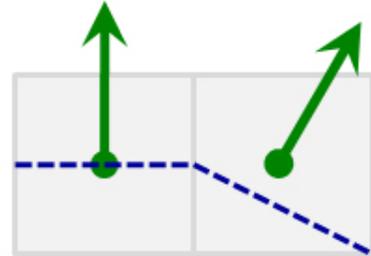


**Figure 7.3:** Foreshortening.

estimated by this amount. In either case, correcting for foreshortening involves scaling the weight of each RGBN pixel, or decreasing the diffusion rate, by a factor of $\sec \theta$. This factor changes as smoothing progresses, leading to a nonlinear problem.

We therefore approximate the smoothing process with a linear one, by assuming that the view direction is constant across the image. For example, assume that the viewer is in the $z$ direction. In this case, scaling the contribution of each normal by $\sec \theta$ transforms the vector $(n_x, n_y, n_z)$ into $(n_x/n_z, n_y/n_z, 1)$. (In practice we use $n_z + \varepsilon$ in the denominator to avoid dividing by zero.) Performing convolution or isotropic diffusion on these vectors will leave the third component as 1, meaning that smoothing is now linear. This remapping also removes the need for explicit normalization at the end of the process, since there is a unique mapping from the first two components of the new vector to a unit-length vector in 3D.

**Bilateral Filtering:**   While simple smoothing is sufficient for many applications, greater control and higher quality are often obtained with a filter that explicitly preserves edges.

80

The bilateral filter [Tomasi and Manduchi, 1998] is a non-iterative edge-preserving filter that bases the contribution of each pixel to the result on a *domain filter*, analogous to the Gaussian weight in standard smoothing, and a *range filter* that prevents the influence of pixels of significantly different intensities. The resultant color $c'_i$ of pixel $i$ is:

$$c'_i = \frac{\sum c_j \; g(|x_i - x_j|, \sigma_x) \; g(|c_i - c_j|, \sigma_c)}{\sum g(|x_i - x_j|, \sigma_x) \; g(|c_i - c_j|, \sigma_c)}, \tag{7.1}$$

where $c_i$ and $x_i$ are the color and location of pixel $i$, $g$ is a Gaussian, and the sum is over all pixels $j$ in the image. In this equation, $\sigma_x$ and $\sigma_c$ are the widths of the domain and range filters, respectively; decreasing $\sigma_c$ leads to better preservation of edges. The color differences may be computed in either a linear space such as RGB, or a perceptually-uniform space such as CIE-lab.

While we could apply a 6D bilateral filter to RGBN images directly, or apply separate filters to the colors and normals, we have observed situations in which it is profitable to apply a joint filter with separate control over color and normal similarity. For example, we may wish to avoid smoothing the colors across an edge that is visible only in the normal map (i.e., respecting discontinuities in shape rather than simply discontinuities in color). We achieve this effect by augmenting the bilateral filter with a term that reduces the influence of samples on the basis of differences in normals:

$$c'_i = \frac{\sum c_j \; g(|x_i - x_j|, \sigma_x) \; g(|c_i - c_j|, \sigma_c) \; g(|n_i - n_j|, \sigma_n)}{\sum g(|x_i - x_j|, \sigma_x) \; g(|c_i - c_j|, \sigma_c) \; g(|n_i - n_j|, \sigma_n)}. \tag{7.2}$$

The normal differences $|n_i - n_j|$ are computed using the $1/n_z$ foreshortening correction, as above. An analogous equation is used to compute the filtered normal maps.

By varying the bilateral filter parameters, we obtain a variety of effects appropriate for different rendering styles. Figure 7.2 compares several parameter settings: note that using a small $\sigma_c$ preserves sharp color boundaries, such as the pupil, while additionally making $\sigma_n$ small emphasizes creases, such as the eyebrow.

## 7.2.2 Curvature Estimation

Many algorithms for nonphotorealistic shading and line extraction make use of the curvatures of the surface and, in some cases, higher-order derivatives. As a brief review, let us recall that the *normal curvature* $\kappa_n$ of a surface in some direction is the reciprocal of the radius of the circle that best approximates a normal slice of surface in that direction. The normal curvature varies with direction, but for a smooth surface it satisfies

$$\kappa_n = \frac{\begin{pmatrix} s & t \end{pmatrix} \begin{pmatrix} e & f \\ f & g \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix}}{\begin{pmatrix} s & t \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix}} = \frac{\begin{pmatrix} s & t \end{pmatrix} \mathbf{II} \begin{pmatrix} s \\ t \end{pmatrix}}{\begin{pmatrix} s & t \end{pmatrix} \mathbf{I} \begin{pmatrix} s \\ t \end{pmatrix}} \tag{7.3}$$

for any vector $(s, t)$ expressed in terms of a tangent-plane coordinate system centered at the point. The symmetric matrices $\mathbf{I}$ and $\mathbf{II}$ appearing here, known as the first and second fundamental tensors, are therefore the basic quantities we wish to compute: knowing them, we may find the curvature in any direction or compute quantities such as mean curvature (half the trace of $\mathbf{I}^{-1}\mathbf{II}$), Gaussian curvature (the determinant), or the principal curvatures and directions.

The first fundamental tensor is defined as

$$\mathbf{I} = \begin{pmatrix} u \cdot u & v \cdot u \\ u \cdot v & v \cdot v \end{pmatrix}, \tag{7.4}$$

while the second fundamental tensor $\mathbf{II}$ is defined in terms of the directional derivatives of the surface normal:

$$\mathbf{II} = \begin{pmatrix} D_u n & D_v n \end{pmatrix} = \begin{pmatrix} \frac{\partial n}{\partial u} \cdot u & \frac{\partial n}{\partial v} \cdot u \\ \frac{\partial n}{\partial u} \cdot v & \frac{\partial n}{\partial v} \cdot v \end{pmatrix}, \tag{7.5}$$

where $(u, v)$ are the axes of an arbitrary coordinate system in the tangent frame. This

suggests that the first and second fundamental tensors may be estimated using finite differences, applying analogues of image edge detection kernels to the normal map. As with smoothing, however, we must be careful to account for the effect of foreshortening on the estimated curvatures. To do this, we take

$$u = \left(1, 0, -\tfrac{n_x}{n_z}\right), \quad v = \left(0, 1, -\tfrac{n_y}{n_z}\right) \tag{7.6}$$

in the above equations. These are vectors in the local tangent plane, and have the property that they project to the (unit-length) $\hat{x}$ and $\hat{y}$ directions in the image. Therefore, we may approximate the derivatives of the surface normal using finite differences:

$$\left(\tfrac{\partial n}{\partial u}\right)_{i,j} = \tfrac{1}{2}\left(n_{i+1,j} - n_{i-1,j}\right), \quad \left(\tfrac{\partial n}{\partial v}\right)_{i,j} = \tfrac{1}{2}\left(n_{i,j+1} - n_{i,j-1}\right). \tag{7.7}$$

This formulation corresponds to the simplest (smallest-support) symmetric discrete derivative kernel; larger kernels (such as Sobel) may be used to provide smoother estimates.

### 7.2.3 Segmentation

Though segmentation is a fundamental part of image editing and compositing, we observe that segmenting images based purely on pixel intensities is inherently a hard problem: the same object may have dramatic variations in color, or lighting may cause two adjacent objects to not have a visible color discontinuity. With RGBN images, on the other hand, there is a second channel of information available: the normal map. For this reason, RGBN segmentation in fact typically produces better results than color-only segmentation.

While there are may classes of image segmentation algorithms available today, we choose to begin with the graph-partitioning algorithm described by [Felzenszwalb and Huttenlocher, 2004]. This algorithm is efficient ($O(n \log n)$ time), easily adaptable to incorporate dissimilarity functions based on both color and normal differences, and captures perceptually important regions while maintaining a global effect. Informally, this algo-

rithm begins by building a graph corresponding to the image, with pixels at vertices and adjacent pixels connected by edges weighted by color dissimilarity. The algorithm partitions the graph such that between-segment dissimilarity (defined as the minimum edge weight between components) is greater than within-segment dissimilarity (defined as the largest weight of its minimum spanning tree, normalized by segment size and scaled by a user-selected constant).

In order to adapt this scheme for RGBN images, we modify the algorithm to include a dissimilarity function based on both normal and color differences, rather than on color only. In addition, we filter the image as a preprocess (using the bilateral filter) in order to remove artifacts due to noise. Incorporating these changes, we are able to produce accurate segmentations even when there are no visible edges in color. For example, in Figure 7.4 we are able to segment the hammer and shears into meaningful facets despite the lack of color information. Figure 7.5 shows an example where variations in the background (due to lighting) may cause artifacts if only color dissimilarities are considered (left). By combining normal and color segmentation, we can successfully segment the silhouette of



**Figure 7.4:** RGBN segmentation can produce accurate results even when there are no visible color edges. For example, in the segmentation result at bottom, note that the hammer has been segmented into multiple facets.

the face from the background. Figure 7.6 demonstrates that we have control over a variety
of segmentation effects.



**Figure 7.5:** RGBN Segmentation Using Color and Normals: **Left:** RGBN Segmentation
using color-only may lead to artifacts due color variations and non-uniform lighting. **Right:**
Segmentation based on color and normals results in more meaningful applications. In this
case we separate the silhouette of the face from the background.
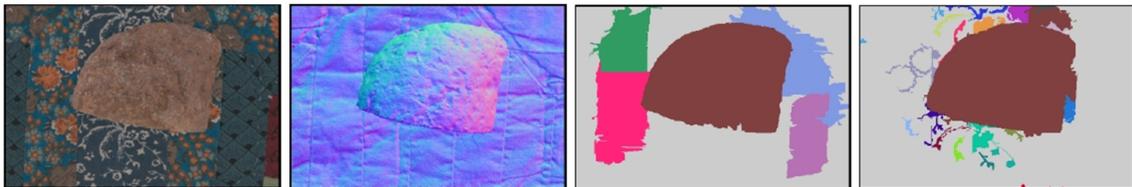


**Figure 7.6:** The Effects of Adjusting RGBN Segmentation Parameters: We use a combi-
nation of RGBN bilateral filtering and manual adjustments of the proportion of color and
normal dissimilarities to produce different segmentation results. When the normal con-
tribution dominates, we are able to segment the rock from the background. Conversely,
a higher contribution of color permits the segmentation of the patterns in the background
quilt.

## 7.3   Stylized Depiction

Now that we have a signal processing framework for manipulating RGBN images, we can
apply these tools for stylized depiction. In this section, we adapt several existing algorithms
for RGBN processing. We also present new algorithms for simulating shadows that convey
surface curvature. In Section 7.4, we will present applications of these techniques.
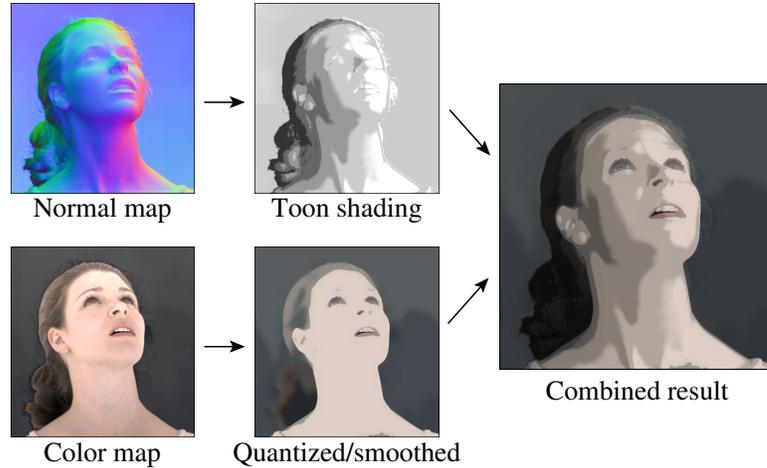
**Figure 7.7:** Combining toon shading with quantized color.

## 7.3.1 Toon Shading

Cartoon shading, consisting of quantizing the amount of diffuse shading (i.e., $n \cdot l$) and mapping each discrete value to a different color, is a popular ingredient in nonphotorealistic renderings [Decaudin, 1996]. While its roots lie in technical limitations of the cartoon and comics media, it remains popular because it abstracts shading while conveying information about geometry (the boundaries between toon shading regions are *isophotes* — curves of constant illumination — which have been shown to convey shape). Because toon shading is an effect that only depends on the surface normals, it easily extends to RGBN images. We use a bilateral filter to de-noise the normal maps, and further reduce the presence of jagged region boundaries by using a smooth-step function, instead of hard quantization.

In order to incorporate the color (albedo) information into the toon shading, we have investigated simply multiplying the shading calculation by the color (albedo) values, but this results in an un-natural image: it is difficult to interpret the toon shading in the presence of the smooth variations in color. Instead, we have found that it is beneficial to combine the toon shading with a *quantized* version of the color data, obtained by running a clustering algorithm. One result is shown in Figure 7.7: we have combined toon shading based on the normal map with a version of the color map that has been smoothed (using the color-only bilateral filter), then quantized to 16 colors.
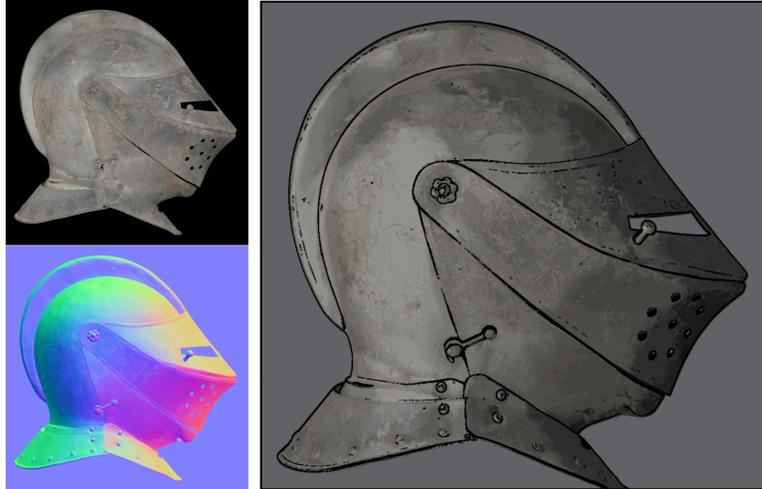
**Figure 7.8:** Locations of depth discontinuities overlaid on toon shading (right). Color only (top left). Normals (bottom left).

## 7.3.2 Discontinuity Lines

Silhouettes and occluding contours (i.e., locations of depth discontinuities) may be extracted either during the initial data analysis (as described in Section 3.2), or from the normal maps themselves. In cases in which extraction of discontinuities during acquisition is unreliable, we look for locations at which

- two adjacent pixels have very different normals (their dot product is below a threshold), and

- one of those normals is nearly orthogonal to the view (its dot product with the view is within a threshold of zero).

Figure 7.8 shows discontinuity lines together with toon shading: note that this highlights *geometric* discontinuities as opposed to *color* discontinuities.

## 7.3.3 Suggestive Contours

A second type of linear feature is the suggestive contour, introduced recently by [DeCarlo *et al.*, 2003, 2004]. Intuitively, these are "almost contours" — locations at which contours first appear with minimal change in viewpoint. Alternatively, they may be thought of as

locations of intensity minima in a head-lit image (i.e., locations at which $n \cdot v$ is not zero, as it would be for contours, but is a local minimum). DeCarlo et al. describe two general algorithms for finding suggestive contours. Their object-space algorithm extracts zero-crossings of normal curvature in the projected view direction, whereas an image space version looks for intensity valleys in a head-lit image. We have explored both approaches for suggestive-contour extraction in RGBN images. Figures 7.17 and 7.9 show the results of applying the image-space algorithm. We observe that this algorithm works well for bumpy surfaces, since it allows for flexibility in cleaning up short lines (the figures use bilateral filtering and a connected-components pass to prioritize lines by length and prune short lines). For smoother objects, such as the dish in Figure 7.10, the object-space algorithm (using the curvatures and curvature derivatives computed above) yields good results.

We also extract crest lines (ridges and valleys), which are defined as local maxima of the greatest principal curvature, in the corresponding principal direction. This is implemented by finding locations at which one of the components of the derivative-of-curvature tensor is zero, together with additional conditions on principal curvature [Ohtake *et al.*, 2004]. A sample result is shown in Figure 7.10, right.



**Figure 7.9:** Suggestive Contours - Image Space Algorithm: The image space algorithm is sensitive to noise in the normal map (left). Smoothing and pruning short lines produce better results (right).
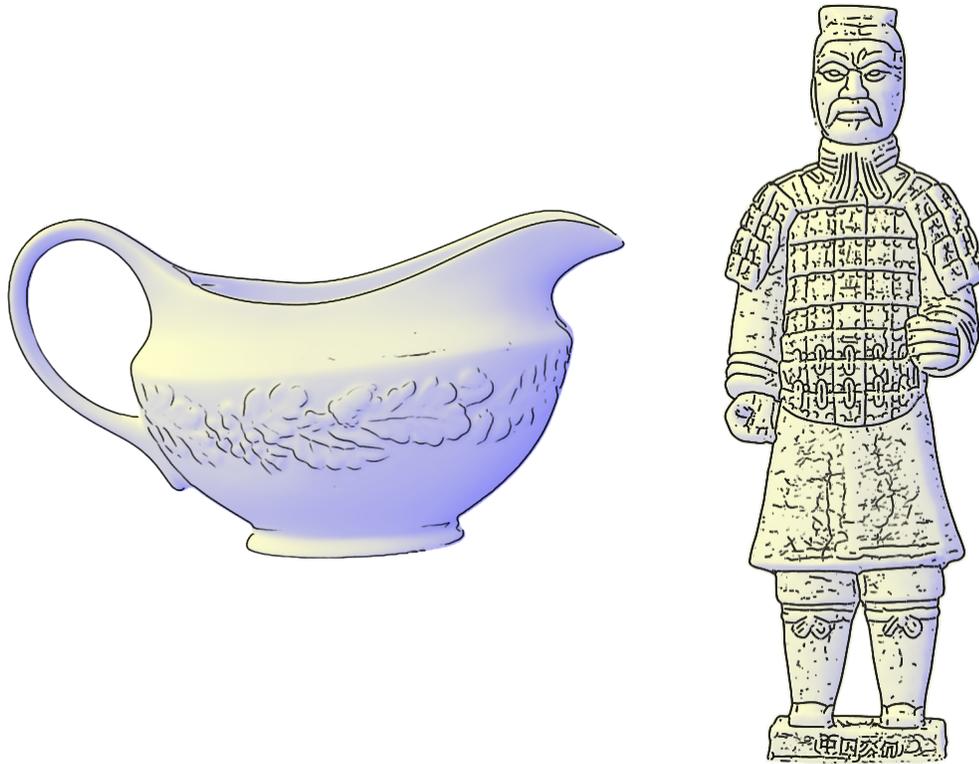
**Figure 7.10:** Suggestive Contours - Object Space Algorithm and Valley Lines: Renderings with suggestive contours (left) and valley lines (right).

### 7.3.4 Exaggerated Shading

A final effect that we adapt to RGBN images is "exaggerated shading," which increases local contrast at all surface orientations and across all scales [Rusinkiewicz *et al.*, 2006]. The computation uses several smoothed normal maps, adjusting the effective light source position at each point based on the smoothed normals and principal directions. The illustration of the terracotta soldier in Figure 8.1 reveals surface markings, such as the grooves on the individual buttons of the soldier's coat, at all levels of detail.

### 7.3.5 Curvature Shading and Shadows

Nonphotorealistic shading effects that convey cues about shadowing and indirect illumination are also frequently encountered, and include mean curvature shading [Kindlmann *et al.*, 2003], accessibility shading [Miller, 1994], ambient occlusion [Zhukov *et al.*, 1998], and depth shading [Cohen *et al.*, 2004]. While the details of these algorithms vary, broadly
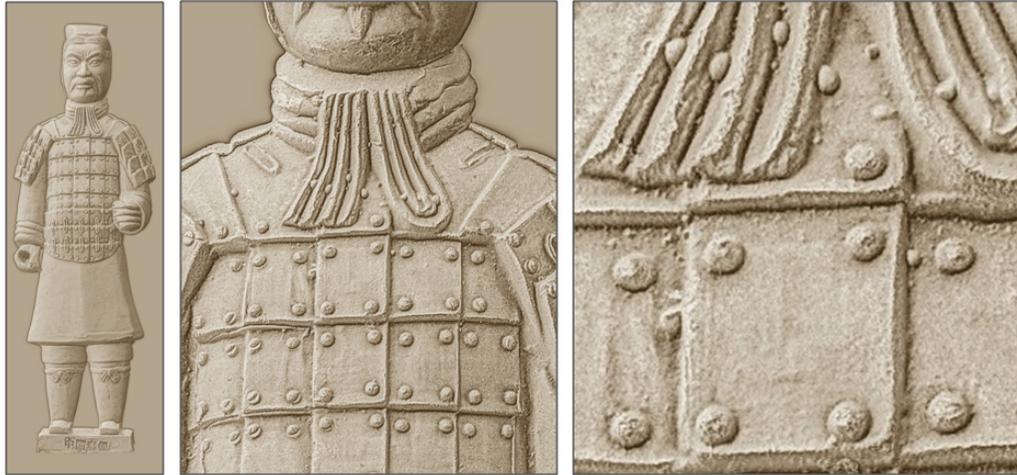
**Figure 7.11:** Exaggerated shading preserves subtle details at all levels of smoothing.

speaking they are all inspired by the intuitive observation that less light reaches valleys and folds on a surface. Hence, they darken indentations on the surface, and sometimes lighten bumps or ridges. Many of these algorithms, however, require knowing depth and hence are impractical to adapt to RGBN images. Therefore, we introduce two new algorithms that produce similar effects, yet may be used with RGBN images.

### 7.3.6 Multiscale Curvature Shading

Of the above algorithms, the most practical to adapt to RGBN images is mean curvature shading, in which areas of negative mean curvature (concavities) are darkened, while areas of positive mean curvature are optionally lightened (Figure 7.12b). However, one limitation of mean curvature shading as a method for conveying shape is that it only reveals high-frequency details. This is in contrast with methods such as ambient occlusion or accessibility shading, which are affected by features of many scales. Therefore, we introduce *multi-scale* mean curvature shading, in which the curvature is computed at multiple levels of smoothing and the final color is an average of shading computed at different scales:

$$c_H = \frac{1}{2} + \frac{1}{2} \sum_{i=0}^{n} \operatorname*{clamp}_{[-1...c_{max}]} a_i H_i, \tag{7.8}$$

(a) Diffuse shading  (b) Single-scale  (c) Multi-scale  (d) Multi-scale
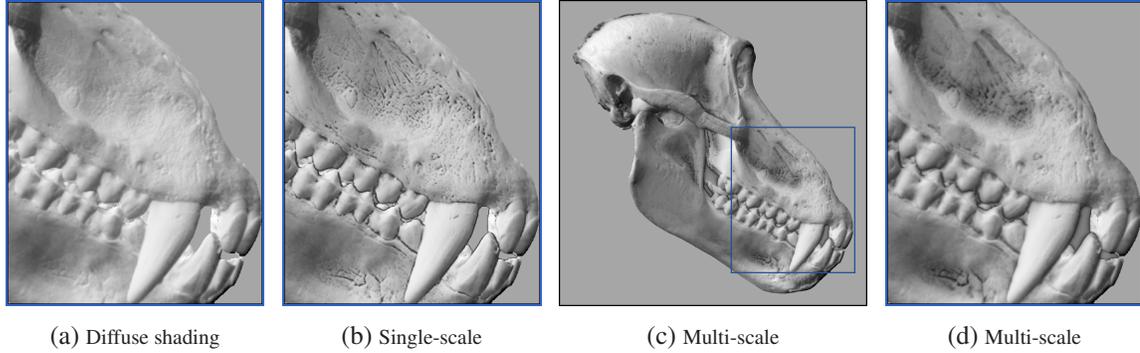
**Figure 7.12:** Mean curvature shading based on the original curvatures (b) reveals fine detail but does not convey a sense of overall shape. Multi-scale curvature shading (c and d) more closely resembles ambient occlusion, revealing shape over local neighborhoods.

where $H_i$ is the mean curvature at the $i$-th scale, $a_i$ are weights that may be equal across scales or chosen to emphasize high or low frequencies, and $c_{max}$ is a parameter that may be set to 1 to both lighten bumps and darken concavities, or to 0 to only darken concavities. As shown in Figure 7.12 c and d, multi-scale curvature shading brings out the detail of differently-sized features simultaneously and compensates somewhat for the difficulty of implementing more global methods, such as ambient occlusion, for RGBN images.

## 7.3.7 Fast Discontinuity Shadows

In order to provide a better sense of depth, we may also darken areas adjacent to the discontinuity lines computed during data acquisition. We simulate shadowing at these lines by using the fact that, as discussed by [Raskar *et al.*, 2004], we may infer not only the position but also the *direction* of the discontinuity. That is, we know on which side of the discontinuity line the occluding and occluded objects lie. Given this information, we may darken only the occluded side of the discontinuity line, achieving an effect similar to that demonstrated by the depth-buffer unsharp masking of [Luft *et al.*, 2006].

We begin with a *directional discontinuity map* $\big(d_x(x,y),\, d_y(x,y)\big)$, in which the value at each pixel has magnitude proportional to the strength of the discontinuity and direction pointing towards the occluder. We compute our shadowing by convolution: the goal is to sum up kernel functions placed at each pixel, scaled and oriented according to $(d_x, d_y)$. In

x, y shadow
kernels

Directional discontinuity map

Computed discontinuity shadows
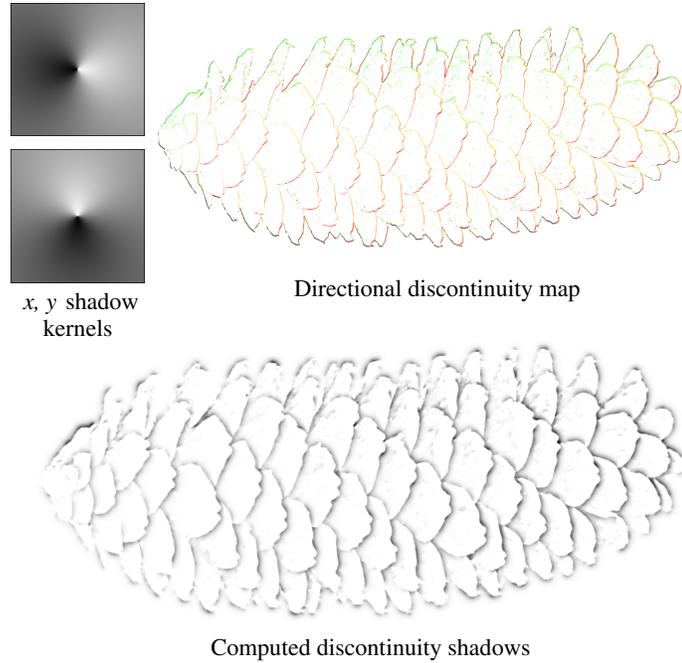
**Figure 7.13:** We convolve shadow kernels with our directional discontinuity map, then clamp to only negative values to yield our discontinuity shadows.

order to compute this efficiently, we observe that, for any 1D function $\rho(r)$, the functions

$$s_x(x, y) = x\,\rho(r), \quad s_y(x, y) = y\,\rho(r), \quad \text{where } r = \sqrt{x^2 + y^2}, \tag{7.9}$$

have the property that $(s_x, s_y) \cdot (d_x, d_y)$ is just a version of $s_x$ rotated according to the given direction. We use

$$\rho(r) = e^{-r}/r \tag{7.10}$$

and compute

$$Shadow = 1 \; + \; \underset{[-1..0]}{\operatorname{clamp}} \left[ d_x * s_x \; + \; d_y * s_y \right)\Big]. \tag{7.11}$$

Figure 7.13 shows the shadow kernels, discontinuity map, and computed shadows for the pinecone dataset, while Figure 7.1 shows a result combining the shadows with exaggerated shading.

## 7.4    Results and Real-World Applications

In this chapter, we described the advantages of using RGBN datasets for illustrating the complex surface structure of acquired objects. We explained how image processing algorithms can be expanded to properly process these datasets that store both color and normals at each pixel. Our system uses these operators in combination with non-photorealistic rendering equations to produce a variety of effects. In this section, we present the results of using our full non-photorealistic rendering pipeline on selected objects that would be impractical to acquire with full 3-D approaches in archeological settings. Our goal is to demonstrate the flexibility of the system as well as its effectiveness in a number of real-world applications within the cultural heritage domain.

The first applications we consider relate to the analysis of chisel marks and the interpretation of the imagery depicted on the Sennedjem Lintel. As shown in Figure 7.14 (e), by applying Lambertian shading to a grey scale intensity map, we separate color from shape and can focus our attention on the tool marks in the stone surface. These details may be
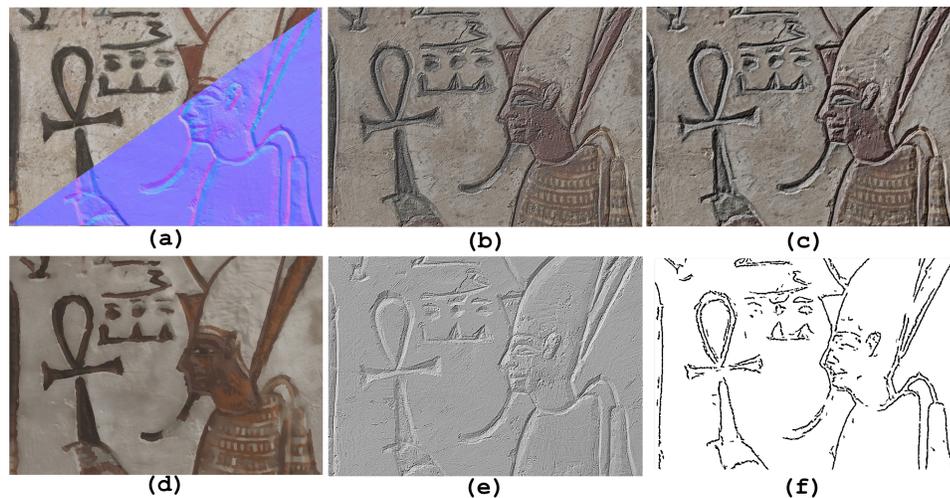


**Figure 7.14:** Analysing Chisel Marks on The Sennedjem Lintel from the Phoebe A. Hearst Museum of Anthropology: (a) RGBN image with color and normals at each pixel. (b) Exaggerated shading reveals fine surface details. (c) Details are further enhanced by darkening groves and emphasizing large features. (d) Toon shading depicting creases on the surface. (e) Lambertian shading computed on the greyscale image conveys surface shape. (f) Suggestive contours present another method for conveying shape.

further enhanced by darkening groves and emphasizing large features (c). Conservators can use this information to identify the tools used to sculpt the surface. In (b) exaggerated shading renders the lintel in a weathered style reminiscent of meticulous hand-shading, and revealing fine surface detail that is difficult to see in color images. This method is effective because it enhances subtle details on the surface that may provide clues about how the object's structure has changed over time. In addition, archaeologists often study such drawings like the ones on this lintel by replicating the imagery in the form of line drawings. Generating these illustrations digitally to compliment this process may be more expedient (especially for complex compositions). An example is shown in (f) where suggestive contours are used to extract lines of curvature.
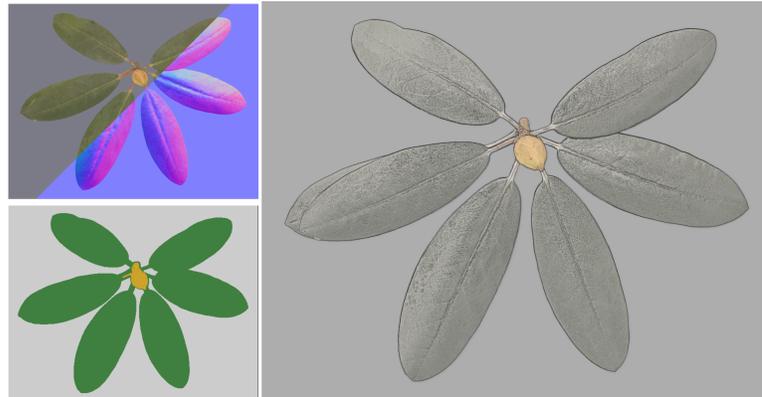


**Figure 7.15:** Textbook Illustration for Botany: A dataset of leaves with a bud (top left) is segmented on the basis of both colors and normals (bottom left) to separate the bud from the leaves. This enables one to apply a different stylization for each segment (right).

The second application we explore is textbook illustration. Figure 7.15 shows how RGBN segmentation can be used in combination with two different stylization techniques, toon shading and exaggerated shading, to diagram different components of the plant. The shape and volume of the bud is represented using toon shading while the patterns of veins of the leave are emphasized with exaggerated shading. Archeologists often work with botanists who examine and compare natural vegetation on a site with those illustrated on discovered artifacts. This process is important for determining how the natural environment has evolved over time, or to simply identify the depicted foliage. In Figure 7.16, suggestive
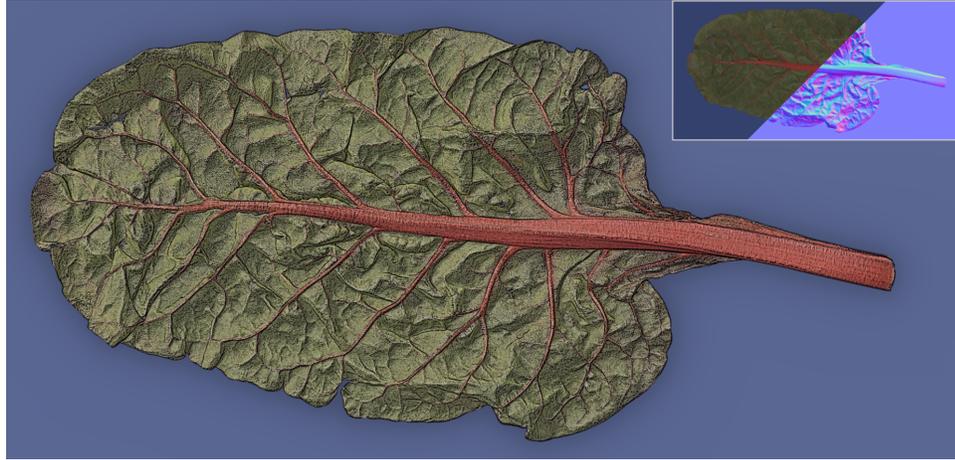
**Figure 7.16:** Analyzing Venation in A Leaf of Rainbow Chard: A leaf of rainbow chard, stylized using a combination of exaggerated shading and suggestive contour lines.

contours overlaid onto exaggerated shading reveal information about the venation on a leaf of chard that is not visible in the color-only representation.

Different rendering styles convey different characteristics of objects. In the hands of a skilled practitioner, they may be used to study the physical composition of objects. In Figure 7.17, the conch shell is rendered with a watercolor effect. This type of illustration is commonly used in biology textbooks. However, the image on the right uses exaggerated shading illustrated to reveal different layers that compose the shell. Combined with suggestive contours the groves and bumps on the surface of the shell are also emphasized.
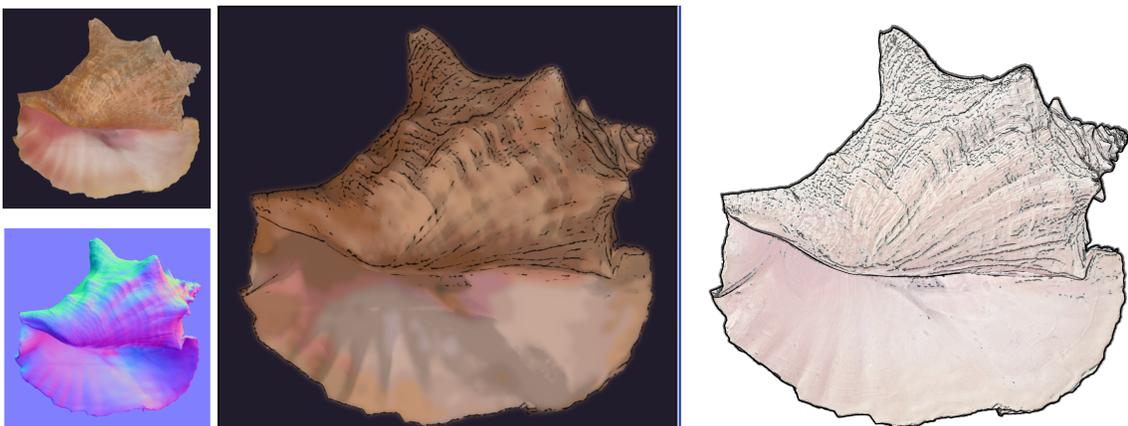


**Figure 7.17:** Analyzing The Surface Structure of A Conch Shell: Soft toon shading based on the normal map, combined with a smoothed version of the RGB color, yields a watercolor effect (middle). Suggestive contour lines are added to emphasize features (right). Color and normal maps (left).

**Figure 7.18:** Sub-Millimeter Analysis of Surface Markings: Illustration of tools reveals fine details, such as the maker's stamp on the shears.
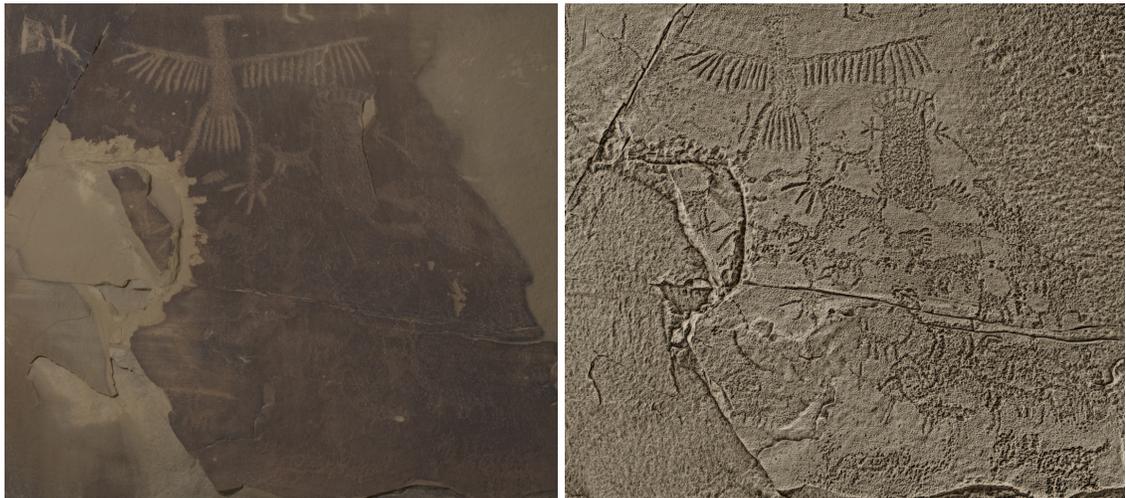


**Figure 7.19:** Studying The Iconographic Content of Petroglyphs: Color image and non-photorealistic rendering (with mean curvature shading and exaggerated shading) of the Legend Rock archaeological site.

Using the smoothing and curvature-estimation operators described in Section 7.2.2, we achieve results such as those demonstrated in Figures 7.18, and 7.19. Figure 7.19 presents an analysis of a petroglyph, from the Legend Rock site in Wyoming, that is thought to be up to 9,000 years old. On the right is a non-photorealistic visualization of the fine relief of the surface. When compared to the photograph on the left, it clearly reveals deep recesses and shallow inscriptions that are almost invisible in the color image. This is a

powerful tool for reading and interpreting the iconography of rock art. Curators may also use RGBN visualizations for sub-millimeter analysis of object surfaces. Close inspection of the tools rendered in Figure 7.18, reveals information about the manufacture and small graffiti markings, none of which are distinguishable in the color photograph.

## 7.5 Discussion

Now that we have presented the advantages of operating on RGBN images, we discuss some of the challenges of working with these datasets.

There are several limitations of our technique that are attributed to the acquisition process itself. It is difficult to acquire the surface normals of dark, shiny, translucent, or inter-reflecting objects, particularly when many different materials are simultaneously present in the scene. Hence it is difficult to find heuristics for robust outlier rejection. The acquired normals may also be noisy, leading to difficulty in producing clean illustrations, especially for styles such as suggestive contour rendering. Therefore, for acquired RGBN images it would be helpful to devise new algorithms for greater stability in extracted suggestive contours. One possible strategy is to look for such lines that persist across scales, in a fashion similar to that used by [Jeong *et al.*, 2005] for LOD control of line drawings. Figures 7.20 and 7.21 show the results of and experiment where suggestive contours were computed over multiple scales and only features that were stable across multiple scales where used in the final representation. Small features were also pruned to remove high frequency noise.

Color quantization artifacts may also appear as a jagged boundary between colors. One example of this is shown in Figure 7.22 (left) around the neck of the model. In some cases smoothing does not remove these defects. In Figure 7.22 we show how applying a Lambertian shader to the quantized color and normal map as a pre-processing step, and then combining the resulting image buffer (rather than the quantized color only) with the toon-lit intensity map, may eliminate such artifacts.
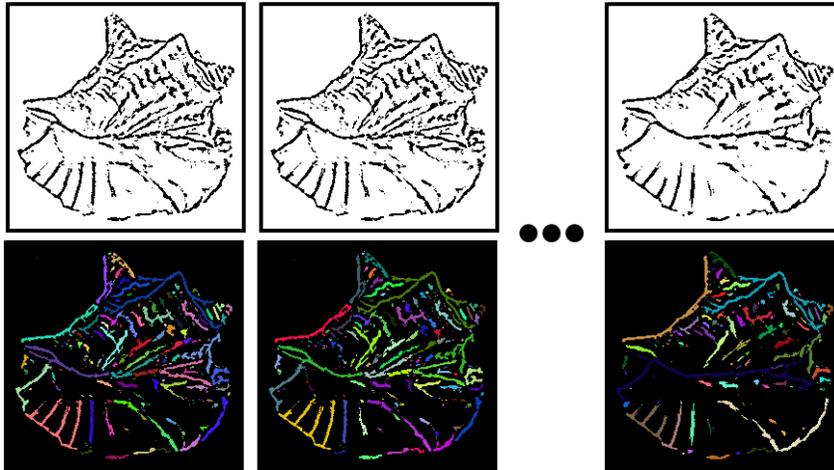
**Figure 7.20:** Suggestive Contours: Scale Space Analysis. In future work, we will explore approaches that use scale space analysis and connected components to prune lines that do not persist over multiple scales. In the top row, we show the result of the image space algorithm for suggestive contours as we vary the width of the smoothing kernel. In the bottom row, we depict the corresponding connected components at each level of smoothing. Each connected component is represented as a unique color.
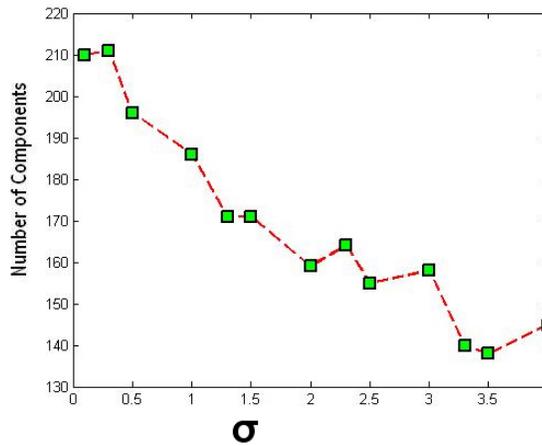


**Figure 7.21:** Suggestive Contours: Tracking Features Across Scales. This diagram illustrates how the number of connected components generated with the image space algorithm for suggestive contours decreases as $\sigma$ increases where $\sigma$ is the width of the smoothing kernel.

**Figure 7.22:** Applying lambertian shading in a preprocessing step may illuminate color quantization artifacts in the toon shading pipeline.

More fundamental to the RGBN datatype are limitations due to the lack of depth information. It is difficult or impossible to change the view, or to compute cast shadows except for very local ones. Finally, the RGBN images considered here only store a single (diffuse) color per pixel, though one may imagine extensions to support specularity or arbitrary BRDFs per pixel.

As shown in Section 7.4, we expect the techniques explored in our system to be suitable for easily creating illustrations of complex objects for which it is difficult to obtain full 3D scans, including domains such as historical documentation, botany, and medicine. Despite the drawbacks presented in this section, we believe that the variety of datasets and styles examined in this thesis demonstrates the wide applicability of RGBN acquisition and stylized rendering.

# Chapter 8

# Conclusion

In this thesis, we address research problems associated with three important aspects of cultural heritage preservation: (1) high-volume data acquisition in non-controlled environments, (2) digital reassembly from thousands of small fragments and (3) scientific illustration with high fidelity. Traditional 3-D acquisition methods are often complex, requiring a level of expertise not commonly available in real-world settings. In addition, the information recorded often lacks the level of detail required for accurate analysis of surface geometry. This limitation is even more apparent when objects have been eroded over time. Our unifying approach is to take advantage of the additional information available in non-traditional images. These multi-channel images are both easy to capture and more robust than color-only representations. Inspired by the latest advancements in 3-D acquisition technologies, we introduced algorithms for capturing artifacts in the field at an archaeological excavation site. We extend signal processing operators for these new datasets and then demonstrate how they improve discriminability in two applications; computer-assisted matching of fresco fragments and illustration of artifacts for archaeological study. In this chapter, we review our research contributions and discuss areas for future work.

## 8.1   Acquiring Fresco Fragments

We introduce a variant of *Shape from Shading* that uses the linear light source of a 2-D flatbed scanner (rather than a point source) to capture small flat objects (such as those at a fresco excavation site). Our approach leverages the slight distance between the scanner's sensor and the light. We scan fragments at multiple orientations and use an alignment algorithm to effectively rotate the light in relation to each pixel. We can then solve for the normal at each pixel. With a linear light source, however, the formulation The Lambertian Lighting Law is no longer applicable. To address this issue, we introduce a unique calibration step that allows us to represent the observed brightness ($l$) in relation to the normal direction ($n$) as parametric function that is easily invertible. Our system uses an 'off-the-shelf' flatbed scanner and avoids the known inefficiencies of traditional 3-D scanning technologies. Moreover, it is a cost-effective solution capable of documenting higher resolution surface details, even for large volumes of small objects. To validate the practicality of our approach, we deployed our system at the Akrotiri excavation site in Santorini, Greece (for use by non-experts).

## 8.2   Multi-Channel Images

The datasets we acquire are in the form of multi-channel images that store more than just a color at each pixel. We focus our work on RGBN images, multi-channel images that store a $(r, g, b)$ color triplet as well as a normal direction, $(x, y, z)$. Throughout our work we leverage the unique qualities of such datasets. In particular, they are easy to acquire at high-resolutions. Although they may be operated on in image space, the results exhibit a high level of accuracy and geometric complexity (typically assisted with a full 3-D models). We demonstrate how to process these datasets for use in applications that improve the preservation of cultural heritage artifacts.

## 8.3   Multi-Cue Machine Learning System

The first application we presented used multiple feature descriptors and machine learning algorithms to reassemble fresco fragments. The system considered different categories of features (ordered by computation time and discriminability) at different matching stages to ensure scalability for large datasets (thousands of fragments). The high resolution color and normals generated by our 2-D acquisition pipeline were incorporated to find similarities among possible matching pairs of fragments. These novel feature descriptors outperformed color and contour features in the presence of discoloration, cracking, and eroded edges. Finally, we demonstrated the effectiveness of machine learning approaches for predicting matches (a non-traditional application of these techniques).

## 8.4   Illustration of Complex Real-World Objects

Next, we developed a non-photorealistic rendering pipeline, capable of visualizing subtle geometric details on the surface of artifacts. We began by creating a palette of tools for processing RGBN datasets. We extended existing signal processing operators in order to provide a variety of effects such as filtering, segmentation and curvature estimation. We applied these effects in combination with several stylized rendering algorithms to produce a variety of illustrations for a diverse range of acquired objects. We addressed the limitations resulting from a lack of depth information by introducing new algorithms for simulating shadows. The resulting digital representations revealed fine surface detail often not available in color-only images. We provided several illustrative examples that demonstrate the effectiveness of our pipeline in real-world archeological settings.

## 8.5 Future Directions

**Acquisition** Researchers continue to investigate approaches that use flatbed scanners for acquiring per pixel normals. Work by [Pintus *et al.*, 2009] builds upon the ideas we present in Chapter 3 but combines traditional photometric stereo with a flatbed scanner that has two light sources. There are opportunities to explore other novel acquisition techniques. Handheld scanners, for example would permit access to recessed areas of complex objects and allow the scanning of fragments that are too large for the scanner's surface.

**Multi-Channel Image Datasets** In this work, we primarily focus on images that store both a color and a surface normal at each pixel. Study of other types of multi-channel images will be beneficial. For example, the concept of RGBN images may be extended further to include the video analog of RGBN (*RGBN/t*). This representation presents unique challenges in acquisition (requiring high-frame-rate cameras and synchronized illumination). Such datasets, however, would allow for stylization effects in the spirit of the work of *Video Tooning*. In this example [Wang *et al.*, 2004] developed a system for transforming an input video into a highly abstracted, spatio-temporally coherent cartoon animation capable of generating a variety of styles. Taking the idea of multi-channels even further, there are opportunities to combine color and normals with light of varying wavelengths. Some objects leave unique 'fingerprints' across the electromagnetic spectrum. These spectral signatures can be analyzed to identify properties like the material composition of an object.

Researchers have made great strides in the analysis and synthesis of texture in RGB images (e.g. Efros and Leung [1999]). Of course such algorithms should apply directly in the RGBN domain. However, a more interesting challenge would be to explore additional operations for these datasets. For example, new algorithms might consider the use of a known RGB channel to hallucinate or inpaint areas of missing normals, or vice versa.

There are opportunities for novel acquisition and visualization tools that explore these new data formats and the unique information they convey. If incorporated into conventional
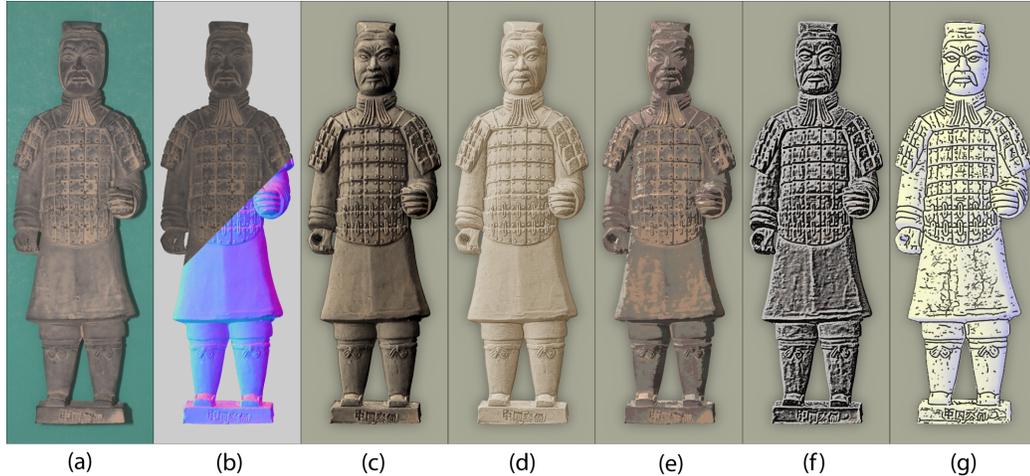
**Figure 8.1:** Combining Stylization Effects: A variety of stylization effects for depicting shape and fine surface details of real-world objects. (a) Color-only image. (b) RGBN image (with color and surface normal at each pixel). (c) Mean curvature shading and exaggerated shading. (d) Exaggerated shading. (e) Toon shading. (f) Exaggerated shading and toon shading. (g) Yellow/blue colormap and valley lines.

image editing programs, RGBN images, for example, would facilitate access to the combinatorial strength of the variety of algorithms described in this thesis. Work by [Pereira and Velho, 2009] presents tools for editing RGBN images while preserving the integrity of the normals. With the increased interest in these techniques in the domain of museum conservation, there is a need for tools that permit experts to combine a variety of styles in a museum settings (Figure 7.11) with intuitive interfaces.

**Matching Problems**  There are many areas of future work in the area of computer-assisted matching. In our work, we demonstrated that classifiers trained on one dataset of frescoes will still perform well on another. Future projects may explore how improved performance could be achieved with an online learning approach: a pre-trained classifier is used to generate an initial classification, with re-training occurring as instances are confirmed to be either correctly or incorrectly classified.

**General Applications**  The datasets and systems presented in this thesis are easily extended to more generalized matching problems in computer graphics. We have the flex-

ibility to operate on any dataset of normals. For example, machine learning techniques could be applied to analyzing patterns in brushstroke textures for brush stroke analysis and detecting forgeries. The additional forms of information would be beneficial for forensic analysis. Overall, we believe that our high-fidelity images will continue to be valuable clues for interpreting the history of fragile historic artifacts. Our framework and empirical analysis have demonstrated the discriminability of incorporating additional information while still maintaining efficiency. It is our hope that this work will be of interest to researchers in the areas of cultural heritage preservation, data acquisition, shape matching and image analysis. We hope to inspire the increased use of normal maps for recognition and machine learning methods for generalized graphics applications in the broader graphics community.

# Bibliography

BARTESAGHI, A., SAPIRO, G., MALZBENDER, T., AND GELB, D. Three-dimensional shape rendering from multiple images. *Graphical Models*, 67(4), 2005.

BERNARDINI, F., MARTIN, I., MITTLEMAN, J., RUSHMEIER, H., AND TAUBIN, G. Building a digital model of Michelangelo's Florentine Pietá. *IEEE Computer Graphics and Applications*, 22(1):59–67, 2002.

BESL, P. AND MCKAY, N. A method for registration of 3-d shapes. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2, pages 239–256. 1992.

BISHOP, C. M. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, 2006.

BROWN, B. J. *Registration and Matching of Large Geometric Datasets for Cultural Heritage Applications*. Ph.D. thesis, Princeton University, 2008.

BROWN, B. J., TOLER-FRANKLIN, C., NEHAB, D., BURNS, M., DOBKIN, D., VLACHOPOULOS, A., DOUMAS, C., RUSINKIEWICZ, S., AND WEYRICH, T. A system for high-volume acquisition and matching of fresco fragments: Reassembling Theran wall paintings. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 27(3), 2008.

COHEN, J., DUNCAN, D., SNYDER, D., COOPER, J., KUMAR, S., HAHN, D., CHEN, Y., PURNOMO, B., AND GRAETTINGER, J. iclay: Digitizing cuneiform. In *Proc. Symposium on Virtual Reality, Archaeology, and Cultural Heritage (VAST)*. 2004.

DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. Acquiring the reflectance field of a human face. In *Proc. ACM SIGGRAPH*. 2000.

DECARLO, D., FINKELSTEIN, A., AND RUSINKIEWICZ, S. Interactive rendering of suggestive contours with temporal coherence. In *Proc. NPAR*. 2004.

DECARLO, D., FINKELSTEIN, A., RUSINKIEWICZ, S., AND SANTELLA, A. Suggestive contours for conveying shape. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 22(3):848–855, 2003.

DECAUDIN, P. Cartoon-looking redering of 3d-scenes. Technical Report 2919, INRIA, 1996.

DOOLEY, D. AND COHEN, M. F. Automatic illustration of 3D geometric models: Lines. In *Proc. I3D*. 1990.

DOUMAS, C. *The Wall-Paintings of Thera*. Thera Foundation – P. M. Nomikos, Athens, 1992.

DOUMAS, C. In *Santorini: A Guide to the Island and its Archaeological Treasures*. Ektodike Athenon S.A., 1999.

EBERT, D. AND RHEINGANS, P. Volume illustration: Non-photorealistic rendering of volume models. In *Proc. IEEE Visualization*. 2000.

EFROS, A. AND LEUNG, T. Texture synthesis by non-parametric sampling. In *Proc. ICCV*. 1999.

ELBER, G. AND COHEN, E. Hidden curve removal for free form surfaces. In *Proc. SIGGRAPH*. 1990.

FELZENSZWALB, P. AND HUTTENLOCHER, D. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2), 2004.

FORNASIER, M. AND TONIOLO, D. Fast, robust and efficient 2d pattern recognition for re-assembling fragmented images. *Pattern Recognition*, 38(11), 2005.

FUNKHOUSER, T., SHIN, H., TOLER-FRANKLIN, C., CASTANEDA, A. G., BROWN, B., DOBKIN, D., RUSINKIEWICZ, S., AND WEYRICH, T. Learning how to match fresco fragments. In *Eurographics 2011 Special Area Track on Cultural Heritage*. 2011.

GARDNER, A., TCHOU, C., HAWKINS, T., AND DEBEVEC, P. Linear light source reflectometry. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 22(3):749–758, 2003.

GOOCH, A., GOOCH, B., SHIRLEY, P., AND COHEN, E. A non-photorealistic lighting model for automatic techical illustration. In *Proc. ACM SIGGRAPH*. 1998.

GOOCH, B. AND GOOCH, A. *Non-Photorealistic Rendering*. A. K. Peters Ltd., 2001.

HERTZMANN, A. AND ZORIN, D. Illustrating smooth surfaces. In *Proc. ACM SIGGRAPH*, pages 517–526. 2000.

HODGES, E. *The Guild Handbook of Scientific Illustration*. Wiley, 2nd edition, 2003.

HORN, B. K. P. *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*. Ph.D. thesis, Massachusetts Institute of Technology, 1970.

HORN, B. K. P. Hill shading and the reflectance map. *Proc. IEEE*, 69(1):14–47, 1981.

HUANG, Q.-X., FLÖRY, S., GELFAND, N., HOFER, M., AND POTTMANN, H. Reassembling fractured objects by geometric matching. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 25(3), 2006.

INTERRANTE, V., FUCHS, H., AND PIZER, S. Enhancing transparent skin surfaces with ridge and valley lines. In *Proc. IEEE Visualization*. 1995.

JEONG, K., NI, A., LEE, S., AND MARKOSIAN, L. Detail control in line drawings of 3D

meshes. *The Visual Computer*, 21(8–10):698–706, 2005.

KARASIK, A. AND SMILANSKY, U. 3d scanning technology as a standard archaeological tool for pottery analysis: practice and theory. *Journal of Archaeological Science*, 2007. doi:doi:10.1016/j.jas.2007.08.008.

KINDLMANN, G., WHITAKER, R., TASDIZEN, T., AND MÖLLER, T. Curvature-based transfer functions for direct volume rendering: Methods and applications. In *Proc. IEEE Visualization*. 2003.

KOLLER, D. AND LEVOY, M. Computer–Aided Reconstruction and New Matches in the Forma Urbis Romae. In *Formae Urbis Romae - Nuove Scoperte, Bullettino Della Commissione Archeologica Comunale di Roma*. 2005.

KOLLER, D., TRIMBLE, J., NAJBJERG, T., GELFAND, N., AND LEVOY, M. Fragments of the city: Stanford's digital forma urbis romae project. In *Proceedings of the Third Williams Symposium on Classical Architecture, Journal of Roman Archaeology*, volume Suppl. 61, pages 237–252. 2006.

KONG, W. AND KIMIA, B. On solving 2D and 3D puzzles under curve matching. In *Proc. CVPR*, volume 2, pages 583–590. 2001.

LEE, C. H., HAO, X., AND VARSHNEY, A. Geometry-dependent lighting. *IEEE Trans. Visualization and Computer Graphics*, 12(2), 2006.

LEITÃO, H. C. G. AND STOLFI, J. A multiscale method for the reassembly of two-dimensional fragmented objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1239–1251, 2002.

LEVOY, M., PULLI, K., CURLESS, B., RUSINKIEWICZ, S., KOLLER, D., PEREIRA, L., GINZTON, M., ANDERSON, S., DAVIS, J., GINSBERG, J., SHADE, J., AND FULK, D. The digital michelangelo project: 3D scanning of large statues. In *Proc. ACM*

*SIGGRAPH*, pages 131–144. 2000.

LUFT, T., COLDITZ, C., AND DEUSSEN, O. Image enhancement by unsharp masking the depth buffer. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 25(3), 2006.

MALZBENDER, T., GELB, D., AND WOLTERS, H. Polynomial texture maps. In *Proc. ACM SIGGRAPH*. 2001.

MALZBENDER, T., WILBURN, B., GELB, D., AND AMBRISCO, B. Surface enhancement using real-time photometric stereo and reflectance transformation. In *Proc. Eurographics Symposium on Rendering*. 2006.

MANAY, S., YEZZI, A. J., HONG, B. W., AND SOATTO, S. Integral invariant signatures. In *Proc. ECCV*, pages 87–99. 2004.

MARKOSIAN, L., KOWALSKI, M., TRYCHIN, S., BOURDEV, L., GOLDSTEIN, D., AND HUGHES, J. Real-time nonphotorealistic rendering. In *Proc. ACM SIGGRAPH*, pages 415–420. 1997.

MASSELUS, V., DUTRÉ, P., AND ANRYS, F. The free-form light stage. In *Proc. Eurographics Rendering Workshop*. 2002.

MILLER, G. Efficient algorithms for local and global accessibility shading. In *Proc. SIGGRAPH*. 1994.

MUDGE, M., MALZBENDER, T., SCHROER, C., AND LUM, M. New reflection transformation imaging methods for rock art and multiple-viewpoint display. In *Proc. Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)*. 2006.

MUDGE, M., VOUTAZ, J.-P., SCHROER, C., AND LUM, M. Reflection transformation imaging and virtual representations of coins from the Hospice of the Grand St. Bernard. In *Proc. Symposium on Virtual Reality, Archaeology and Cultural Heritage*

*(VAST)*. 2005.

NEHAB, D., RUSINKIEWICZ, S., DAVIS, J., AND RAMAMOORTHI, R. Efficiently combining positions and normals for precise 3D geometry. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 24(3), 2005.

OHTAKE, Y., BELYAEV, A., AND SEIDEL, H.-P. Ridge-valley lines on meshes via implicit surface fitting. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 23(3), 2004.

PAPAODYSSEUS, C., PANAGOPOULOS, T., EXARHOS, M., TRIANTAFILLOU, C., FRAGOULIS, D., AND DOUMAS, C. Contour-shape based reconstruction of fragmented, 1600 bc wallpaintings. *IEEE Trans. on Signal Processing*, 50(6), 2002.

PAULY, M., KEISER, R., AND GROSS, M. Multi-scale feature extraction on point-sampled models. In *Proc. Eurographics*. 2003.

PEREIRA, T. AND VELHO, L. Rgbn image editing. In I. Press, editor, *Proceedings of SIBGRAPI - XXII Brazilian Symposium on Computer Graphics and Image Processing*. 2009.

PINTUS, R., MALZBENDER, T., WANG, O., BERGMAN, R., NACHLIELI, H., AND RUCKENSTEIN, G. Photo repair and 3D structure from flatbed scanners. In *Proc. VIS-APP*. 2009.

QUINLAN, J. R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

RASKAR, R., TAN, K.-H., FERIS, R., YU, J., AND TURK, M. Non-photorealistic camera: Depth edge detection and stylized rendering using multi-flash imaging. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 23(3), 2004.

RHEINGANS, P. AND EBERT, D. Volume illustration: Nonphotorealistic rendering of volume models. *IEEE Trans. Vis. Comp. Gr.*, 7(3):253–264, 2001.

RUSHMEIER, H. AND BERNARDINI, F. Computing consistent normals and colors from photometric data. In *Proc. 3DIM*. 1999.

RUSINKIEWICZ, S., BURNS, M., AND DECARLO, D. Exaggerated shading for depicting shape and detail. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 25(3), 2006.

SAĞIROĞLU, M. Ş. AND ERÇIL, A. A texture based matching approach for automated assembly of puzzles. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 1036–1041. IEEE Computer Society, Washington, DC, USA, 2006. ISBN 0-7695-2521-0. doi:http://dx.doi.org/10.1109/ICPR.2006.184.

SAITO, T. AND TAKAHASHI, T. Comprehensible rendering of 3-d shapes. In *Proc. ACM SIGGRAPH*, pages 197–206. 1990.

SCHUBERT, R. Using a flatbed scanner as a stereoscopic near-field camera. In *IEEE Computer Graphics and Applications*, pages 38–45. 2000.

SERRA, J. *Image Analysis and Mathermatical Morphology*. Academic Press, 1983.

SLOAN, P.-P., MARTIN, W., GOOCH, A., AND GOOCH, B. The lit sphere: A model for capturing npr shading from art. In *Proc. Graphics Interface*. 2001.

TAN, K.-H., KOBLER, J., FERIS, R., DIETZ, P., AND RASKAR, R. Shape-enhanced surgical visualizations and medical illustrations with multi-flash imaging. In *Proc. MICCAI*. 2004.

THIRION, J.-P. AND GOURDON, A. The 3D marching lines algorithm. *Graphical Models and Image Processing*, 58(6):503–509, 1996.

TOLER-FRANKLIN, C., BROWN, B., WEYRICH, T., FUNKHOUSER, T., AND RUSINKIEWICZ, S. Multi-feature matching of fresco fragments. In *Proc. SIGGRAPH Asia*. 2010.

TOLER-FRANKLIN, C., FINKELSTEIN, A., AND RUSINKIEWICZ, S. Illustration of complex real-world objects using images with normals. In *Proc. NPAR*, pages 111–119. 2007.

TOMASI, C. AND MANDUCHI, R. Bilateral filtering for gray and color images. In *Proc. ICCV*. 1998.

VLACHOPOULOS, A. The wall paintings from the Xeste 3 building at Akrotiri. Towards an interpretation of the iconographic programme. In N. Brodie, J. Doole, G. Gavalas, and C. Renfrew, editors, *Horizons: A colloquium on the prehistory of the Cyclades, Cambridge*, pages 451–465. 2008.

WANG, J., XU, Y., SHUM, H.-Y., AND COHEN, M. F. Video tooning. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 23(3):574–583, 2004.

WENGER, A., GARDNER, A., TCHOU, C., UNGER, J., HAWKINS, T., AND DEBEVEC, P. Performace relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 24(3), 2005.

WILLIS, A. Stochastic 3d geometric models for classification, deformation, and estimation. *Ph.D. Thesis*, 2004.

WINKENBACH, G. AND SALESIN, D. Computer-generated pen-and-ink illustration. In *Proc. SIGGRAPH*. 1994.

WOODHAM, R. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.

ZHUKOV, S., IONES, A., AND KRONIN, G. An ambient light illumination model. In *Proc. Eurographics Rendering Workshop*. 1998.

ZICKLER, T., MALLICK, S. P., KRIEGMAN, D. J., AND BELHUMEUR, P. N. Color

subspaces as photometric invariants. *International Journal of Computer Vision*, 79:13–30, 2008.