The Vision Workbook:
A User's Guide to the NASA Vision Workbench v1.0

Matthew D. Hancher
Michael J. Broxton
Laurence J. Edwards

Intelligent Systems Division
NASA Ames Research Center

DRAFT

January 7, 2010

Acknowledgements

Many people have contributed to making this first Vision Workbench open-source release a reality. Thanks to Terry Fong, Kalmanje KrishnaKumar, and Dave Korsmeyer in the Intelligent Systems Division at NASA Ames for supporting us in this research and allowing us to pursue our crazy dreams. Thanks to Larry Barone, Martha Del Alto, Robin Orans, Diana Cox, Kim Chrestenson, and everyone else who helped us navigate the NASA open source release process. Thanks to Randy Sargent, Matt Deans, Liam Pedersen, and the rest of the Intelligent Robotics Group at Ames for lending their incredible expertise and being our guinea pigs time and again. Thanks to our interns—Kerri Cahoy, Ian Saxton, Patrick Mihelich, Joey Gannon, and Aaron Rolett—for bringing many exciting features of the Vision Workbench into being. Finally, thanks to all our users, past, present and future, for making software development enjoyable and worthwhile.

Portions of this work were funded by the Mars Critical Data Products Initiative and the Exploration Technology Development Program.

Contents

Chapter 1

Introduction

This document is designed to be a gentle introduction to programming with the NASA Vision Workbench, a C++ image processing and machine vision library. The Vision Workbench was developed through a joint effort of the Intelligent Robotics Group (IRG) and the Adaptive Control and Evolvable Systems Group (ACES) within the Intelligent Systems Division at the NASA Ames Research Center in Moffett Field, California. It is distributed under the NASA Open Source Agreement (NOSA) version 1.3, which has been certified by the Open Source Initiative (OSI). A copy of this agreement is included with every distribution of the Vision Workbench in a file called `COPYING`.

You can think of the Vision Workbench as a "second-generation" C/C++ image processing library. It draws on the authors' experiences over the past decade working with a number of "first-generation" libraries, such as OpenCV and VXL, as well as direct implementations of image processing algorithms in C. We have tried to select and improve upon the best features of each of these approaches to image processing, always with an eye toward our particular range of NASA research applications. The Vision Workbench has been used within NASA for a wide range of image processing tasks, including alignment and stitching of panoramic images, high-dynamic-range imaging, texture analysis and recognition, lunar and planetary map generation, and the production of 3D models from stereo image pairs. A few examples of image data that has been processed with the Vision Workbench are show in Figure 1.1.

The Vision Workbench was designed from the ground up to make it quick and easy to produce efficient implementations of a wide range of image processing algorithms. Consider this example:

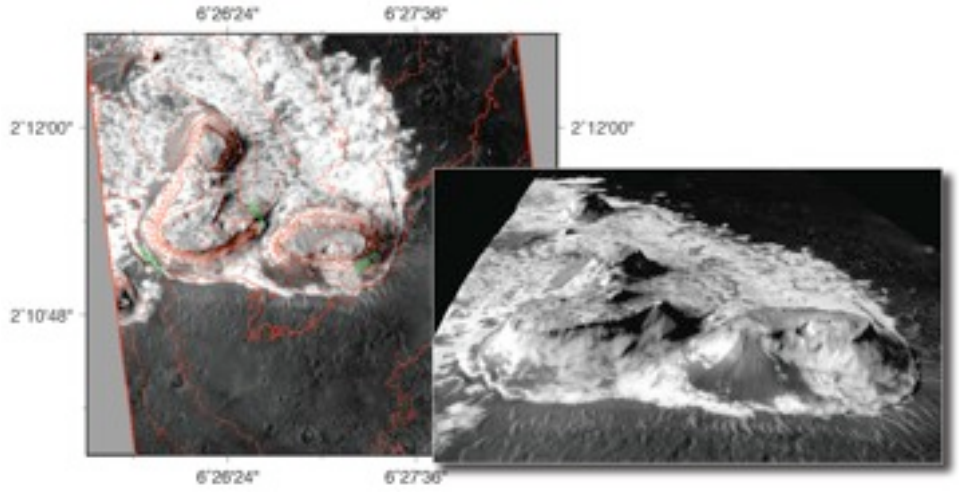
```
background_image += 0.1 * ( source_image - background_image );
```

Hopefully it is reasonably clear what this line of code does, even if you don't know what an IIR filter like this is good for. Higher level functions have similarly simple interfaces. For example, to apply a Gaussian filter to an image with a sigma of 3 pixels you can simply say:

```
image = gaussian_filter( image, 3 );
```

In many cases like these, code written using the Vision Workbench is significantly smaller and more readable than code written using more traditional approaches.

At the core of the Vision Workbench is a rich set of template-based image processing data types representing pixels, images, and operations on those images, as well as mathematical entities (like vectors and geometric transformations) and image file I/O. On top of this core the Vision Workbench also provides a number of higher-level image processing and machine vision modules,



(a)

(b)

(c)

(d)

Figure 1.1: Examples of image data processed with the help of the Vision Workbench. (a) A Martian terrain map generated from stereo satellite imagery. (b,c) Original and high-dynamic-range image mosaics from a NASA field test. (d) A lunar base map generated from the Clementine data set.

providing features including camera geometry modeling, high-dynamic-range imaging, interest point detection and matching, image mosaicing and blending, and geospatial data management.

That said, the Vision Workbench is not for everyone, and in particular it is not intended as a drop-in replacement for any existing image processing toolkit. It is specifically designed for image processing in the context of machine vision, so it lacks support for things like indexed color palettes that are more common in other areas. It also lacks a number of common features that the authors have simply not yet had a need for, such as morphological operations. If you encounter one of these holes while using the Vision Workbench please let us know: if it is an easy hole to fill we may be able to do so quickly. Finally, there are many application-level algorithms, such as face recognition, that have been implemented using other computer vision systems and are not currently provided by the Vision Workbench. If one of these meets your needs there is no compelling reason to re-implement it using the Vision Workbench instead. On the other hand, if no existing high-level tool solves your problem then you may well find that the Vision Workbench provides the most productive platform for developing something new.

Since this is the first public release of the Vision Workbench, we thought we should also provide you with some sense of the direction the project is headed. It is being actively developed by a small but growing team at the NASA Ames Research Center. A number of features are currently being developed internally and may released in the future, including improved mathematical optimization capabilities, a set of Python bindings, and stereo image processing tools. Due to peculiarities of the NASA open-source process we cannot provide snapshots of code that is under development and not yet approved for public release. If you have a specific use for features that are under development, or in general if you have suggestions or feature requests, please let us know. Knowing our users' needs will help us prioritize our development and, in particular, our open-source release schedule.

We hope that you enjoy using the Vision Workbench as much as we have enjoyed developing it! If you have any questions, suggestions, compliments or concerns, please let us know. Contact information is available at the bottom of the `README` file included with your distribution.

Chapter 2

Getting Started

This chapter describes how to set up and start using the Vision Workbench. It explains how to obtain the Vision Workbench and its prerequisite libraries, how to build and install it, and how to build a simple example program. This chapter does *not* discuss how to program using the Vision Workbench. If that's what you're looking for then skip ahead to Chapter 3.

2.1 Obtaining the Vision Workbench

Most likely if you are reading this document then you already know where to obtain a copy of the Vision Workbench sources if you haven't obtained them already. However, if not, a link to the most up-to-date distribution will always be available from the NASA Ames open-source software website, at `opensource.arc.nasa.gov`.

In addition to obtaining the Vision Workbench, you will also need to obtain and install whatever pre-requisite libraries you will need. The only strict requirement is the Boost C++ Libraries, a set of extensions to the standard C++ libraries that is available from `www.boost.org`. Many modern Linux systems come with some version of Boost already installed, generally in the directory `/usr/include/boost`. The Vision Workbench has been tested with Boost versions 1.32 and later.

Other libraries are required only if you want to use particular features of the Vision Workbench. A summary of the various libraries that the Vision Workbench will detect and use if present is given in Table 2.1. It lists the particular Vision Workbench module that uses the library, whether it is required or optional for that module, and where the library can be obtained. Details of each of the modules and the features that are enabled by each dependency are given in the corresponding sections of this book. If you are just starting out with the Vision Workbench, it is generally fine to begin only with Boost. You can always go back and rebuild the Vision Workbench with support for additional features later if you discover that you need them.

One dependency that is worth discussing briefly is LAPACK, which provides Vision Workbench with a computational linear algebra back end. LAPACK is a comprehensive and widely used linear algebra support library in the public domain. LAPACK also require the Basic Linear Algebra Subroutines (BLAS) library, which is usually bundled with LAPACK.

The basic matrix and vector algebra in the Math module does not depend on LAPACK and BLAS, however the routines in `<vw/Math/LinearAlgebra.h>` will only be built if LAPACK is detected by the build system. For your convenience, we provide a stand-alone LAPACK and BLAS distribution on the Vision Workbench web page. This distribution has been tested with the Vision Workbench, so we recommend its use if you are installing LAPACK for the first time. However, other versions of LAPACK and BLAS that come pre-installed on your system will probably work

Name	Used By	Source
Boost	All	`http://www.boost.org/`
LAPACK	Portions of Math, HDR	See note in Section 2.1
PNG	FileIO (opt.)	`http://www.libpng.org/`
JPEG	FileIO (opt.)	`http://www.ijg.org/`
TIFF	FileIO (opt.)	`http://www.libtiff.org/`
OpenEXR	FileIO (opt.)	`http://www.openexr.com/`
PROJ.4	Cartography (req.)	`http://www.remotesensing.org/proj/`
GDAL	Cartography (opt.)	`http://www.remotesensing.org/gdal/`

Table 2.1: A summary of Vision Workbench dependencies.

just as well. In particular, Mac OS X users *do not* need to install LAPACK; machine optimized linear algebra support is provided by Apple's `veclib` framework on Mac OS X. Remember to add the `-framework veclib` flag when linking your application against the Vision Workbench if you are using the functions in `<vw/Math/LinearAlgebra.h>` on the mac platform.

2.2 Building the Vision Workbench

If you are using a UNIX-like platform such as Linux or Mac OS it is generally straightforward to build the Vision Workbench once you have installed any necessary libraries. First unpack the distribution, go to the distribution's root directory, and configure the build system by running "`./configure`". This script will examine your machine to determine what build tools to use and what libraries are installed as well as where they are located. Near the end of its output it will list whether or not it was able to find each library and which Vision Workbench modules it is going to build. You should examine this output to confirm that it was able to find all the libraries that you had expected it to. If not then you may need to configure the build system to search in the right places, as discussed in Section 2.4.

Assuming the output of the `configure` script looks good, you can now proceed to build the Vision Workbench itself by running "`make`". Most of the Vision Workbench is header-only, so "building" the Vision Workbench should be relatively quick. Once the build is complete, confirm that things are working properly by building and running the unit tests by typing "`make check`". If there are no errors, the final step is to install the Vision Workbench headers, library, and sample programs using "`make install`". By default the installation location is the directory `/usr/local`, so you will need to obtain the necessary privileges to write to this directory using a command such as `su` or `sudo`. If you do not have administrator privileges on you computer then see Section 2.4 for information on how to specify an alternative installation directory.

Building the Vision Workbench under Windows is possible, but it is not currently automatically supported. The easiest thing to do is to include the `.cc` files from the Vision Workbench modules that you want to use directly in your own project file. You will of course still need to install the Boost libraries as well as any other libraries you want to use. Pre-built Windows versions of a number of libraries, such as the JPEG, PNG, and TIFF libraries, are available online from the GnuWin32 project at `gnuwin32.sourceforge.net`. You will need to configure your project's include file and library search paths appropriately. Also be sure to configure your project to define the preprocessor symbol `NOMINMAX` to disable the non-portable Windows definitions of `min()` and `max()` macros, which interfere with the standard C++ library functions of the same names.

```
1  // __BEGIN_LICENSE__
2  // Copyright (C) 2006, 2007 United States Government as represented by
3  // the Administrator of the National Aeronautics and Space Administration.
4  // All Rights Reserved.
5  // __END_LICENSE__
6
7
8  #include <iostream>
9  #include <vw/Image.h>
10 #include <vw/FileIO.h>
11
12 int main( int argc, char *argv[] ) {
13   try {
14     VW_ASSERT( argc==3, vw::ArgumentErr() << "Invalid command-line args." );
15     vw::ImageView<vw::PixelRGBA<float> > image;
16     read_image( image, argv[1] );
17     write_image( argv[2], image );
18   }
19   catch( vw::Exception& e ) {
20     std::cerr << "Error: " << e.what() << std::endl;
21     std::cerr << "Usage: vwconvert <source> <destination>" << std::endl;
22     return 1;
23   }
24   return 0;
25 }
```

Listing 1: [vwconvert.cc] A simple demonstration program that can copy image files and convert them from one file format to another.

2.3 A Trivial Example Program

Now that you've built and installed the Vision Workbench let's start off with a simple but fully-functional example program to test things out. The full source code is shown in Listing 1. You should be able to obtain an electronic copy of this source file (as well as all the others listed in this book) from wherever you obtained this document. For now don't worry about how this program works, though we hope it is fairly self-explanatory. Instead, just make sure that you can build and run it successfully. This will ensure that you have installed the Vision Workbench properly on your computer and that you have correctly configured your programming environment to use it.

The program reads in an image from a source file on disk and writes it back out to a destination file, possibly using a different file format. When reading and writing images, the Vision Workbench infers the file format from the file extension of the filename. This example program takes the source and destination filenames as two command-line arguments. For example, to convert a JPEG image called `image.jpg` in the current directory into a PNG image you might say:

```
vwconvert image.jpg image.png
```

Note that exactly what image file formats are support will depend on what file format libraries you have installed on your system.

In order to to build this program you will need to configure your compiler to find the Vision Workbench headers and then configure your linker to find not only the Vision Workbench libraries but also all of the libraries that the Vision Workbench in turn requires.

Some Vision Workbench header files include boost headers, and the compiler needs to be able to find these files when you build your application. No additional configuration is necessary if boost is installed in a stardard system directory, however for non-standard installations, you will need to direct the compiler (usually using the `-I` flag) to the right directory. Note that the Vision Workbench's dependency on boost is unique in this regard; you do not normally need to configure the compiler to find header files for Vision Workbench third party library dependencies.

Keeping track of nested library dependencies like this can be difficult. The Vision Workbench addresses this problem using the GNU `libtool` utility, and we suggest that you use it too. All Vision Workbench libraries are built with an accompanying `libvw<module_name>.la` file that encodes dependency information that `libtool` later uses to pull in all required library dependencies automatically. It's easy to use, and it lets you take advantage of the work that the Vision Workbench build system does to locate your libraries and sort out their dependencies.

Listing 2 shows a sample `Makefile` that demonstrates how to build a Vision Workbench application using `libtool`, among other things. If you already have your own `Makefile` or other build system, the important section to look at is the section titled "Linking rule". It demonstrates how to invoke `libtool` to build a program: invoke the compiler as you usually would, but prefix the command with "`libtool --mode=link`". This will make `libtool` interpret the command line it has been given as a linking command, filling in all the specifics about library dependencies. In this case it will recognize the `-lvw` option, and will expand it to include references to all the libraries upon which the Vision Workbench depends.

You can test this by creating an empty directory and copying the files `vwconvert.cc` and `Makefile.example` into it, renaming the latter as simply `Makefile`. (Both of these files are included in the Vision Workbench source distribution in the directory `docs/workbook`.) You should then be able to build the program by running "`make`". This assumes that you have `libtool` installed on your computer. If not, don't worry: the Vision Workbench includes a copy of the `libtool` script in the base directory of the source distribution. If you see an error message suggesting that `libtool` cannot be found you may need to modify your `Makefile` so that the `LIBTOOL` variable explicitly points to this file.

If you choose not to use `libtool` then you will need to manually ensure that all the necessary dependencies are linked in to your program. The easiest way to be sure that you aren't missing any is to look inside the same files that `libtool` would use to generate the list, the `.la` files. For example, the `vw` library that is included by the `-lvw` option points to the file `lib/libvw.la` underneath whatever directory you installed the Vision Workbench in. This is a human-readable file that lists this library's dependencies, among other things. If any of these dependency libraries are themselves `.la` files then you will need to examine them in turn to find all the recursive dependencies. As you can imagine, this is a cumbersome process, and we suspect that in then end you'll be much happier using `libtool` directly instead.

Using `libtool` on Mac OS X

Users of Mac OS X should be aware that the `libtool` command available in this environment is different than the GNU `libtool` we are discussing here. On these systems, you will need to use the `glibtool` command or use the `libtool` script in the root of the Vision Workbench source distribution directory.

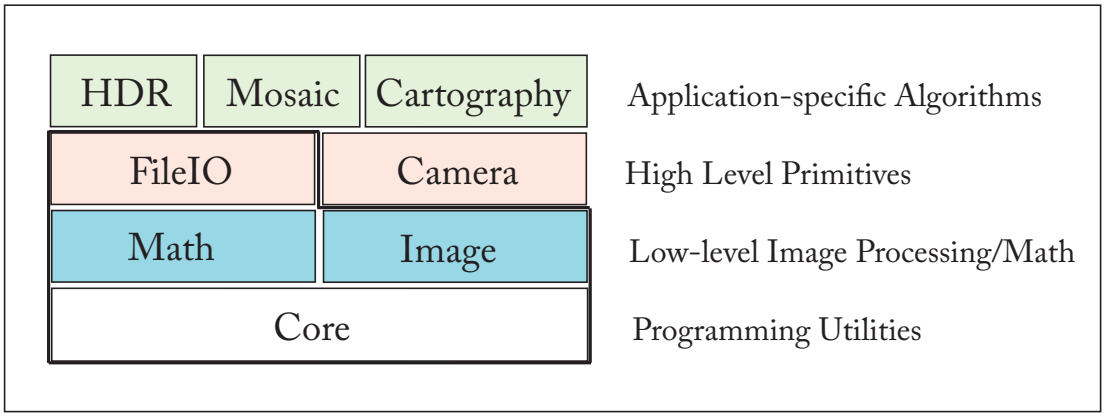


Figure 2.1: *Vision Workbench inter-module dependencies.* Module in this figure depend on those beneath them. These dependencies split the modules into four general classes of increasing complexity and sophistication. The modules surrounded by the bold outline are considered the "foundation" modules that are part of the most basic Vision Workbench distribution.

2.4 Configuring the Build System

The Vision Workbench build system offers a variety of configuration options that you provide as command-line flags to the `configure` script. We'll discuss a few of the most important options here, but for a complete list you can run "`./configure --help`". As an alternative to specifying command-line flags every time, you may instead create a file called `config.options` with your preferences in the base directory of the Vision Workbench repository. A file called `config.options.example` is provided that you can copy and edit to your liking. Note that none of this has any impact on Visual Studio users, who must instead configure their projects by hand.

The single most important option is the `--with-paths=PATHS` flag, where you replace `PATHS` with a whitespace-separated list of paths that the build system should search when looking for installed libraries. For example if you specify the option `--with-paths=/foo/bar` then it will search for header files in `/foo/bar/include`, library files in `/foo/bar/lib`, and so on. The default search path includes a number of common locations for user-installed libraries, such as `/usr/local`, `$(HOME)/local`, and `/sw`. The `PKG_PATHS` configuration file variable has the same effect as this option.

The next most important options have the form `--enable-module-foo[=no]`, where `foo` is replaced by the lower-case name of a module such as `mosaic` or `hdr`. This allows you to control whether or not certain modules are built. Disabling modules that you do not use can speed up compilation and testing time, which is especially useful if you are making changes to the Vision Workbench source and need to recompile often. The corresponding configuration file variables have the form `ENABLE_MODULE_FOO`, in all-caps, and are set to either `yes` or `no`.

It is worth mentioning that the Vision Workbench has several inter-module dependencies that you should take into account when enabling and disabling modules. These are shown in Figure 2.4.

Two handy options, `--enable-optimize` and `--enable-debug`, determine the compiler options used when building the few library files. You can again specify an optional argument of the form `=no` to disable the corresponding feature, and you can also specify a particular optimization level in the same manner. For example, if you want to make it as easy as possible to debug Vision Work-

bench code using a debugger you might use `--enable-optimize=no --enable-debug` to disable all optimizations and include debugging symbols. The corresponding configuration file variables are `ENABLE_OPTIMIZE` and `ENABLE_DEUBG`. Keep in mind that since most Vision Workbench code is header-only you should remember to configure your own project similarly or you may not notice any difference. For normal non-debugging use, we strongly recommend that you enable moderate compiler optimization; much of the heavily templatized and generic Vision Workbench code requires basic optimizations such as function inlining to achieve a reasonable level of performance.

Finally, to specify that the build system should install the Vision Workbench someplace other than `/usr/local`, specify the path using the `--prefix=PATH` option. The corresponding configuration file variable is, of course, called `PREFIX`.

```
1  # The Vision Workbench installation prefix (/usr/local by default)
2  VWPREFIX = /usr/local
3
4  # If you don't have libtool installed, you can specify the full
5  # path to the libtool script in the base directory of your Vision
6  # Workbench source tree, e.g. $(HOME)/VisionWorkbench-1.0/libtool
7  LIBTOOL = libtool
8
9  # Compilation flags:
10 # -O3   turns on optimization, which you should almost always do
11 # -g    enables debugging support
12 # -Wall turns on all compiler warnings
13 CXXFLAGS = -I$(VWPREFIX)/include -O3 -g -Wall
14
15 # Boost:
16 # The Vision Workbench header files require the boost headers. If
17 # boost is installed in a non-standord location, you may need
18 # to uncomment this line and insert the path to the boost headers.
19 # CXXFLAGS += -I<path to boost include dir>
20
21 # Linking flags:
22 # -lvw includes the Vision Workbench core libraries
23 LDFLAGS = -L$(VWPREFIX)/lib -lvw
24
25 # Object files:
26 # List the object files needed to build your program here.
27 OBJECTS = vwconvert.o
28
29 # Linking rule:
30 # Duplicate and modify this rule to build multiple programs.
31 vwconvert: $(OBJECTS)
32         $(LIBTOOL) --mode=link $(CXX) $(LDFLAGS) -o $@ $^
33
34 # Clean-up rule:
35 clean:
36         rm -f *.o *~ \#*
```

Listing 2: [`Makefile.example`] An example `Makefile` that shows how to build a Vision Workbench program using `libtool`.

Chapter 3

Working with Images

This chapter is designed to be a first introduction to programming using the Vision Workbench. It describes images, pixels, color spaces, image file I/O, and basic image manipulation, setting the stage for the fundamental image processing operations described in Chapter 4.

3.1 The `ImageView` Class

The `ImageView` class is the centerpiece of the Vision Workbench in most applications. Simply put, it represents an image in memory. The class is similar to related classes that appear in other C++ computer vision libraries, including VXL, GIL, and VIGRA, so if you are already familiar with one of those libraries you should find nothing too foreign here.

3.1.1 The Basics

An `ImageView` represents a two-dimensional rectangular array of data, such as an image of pixels. It is actually a class *template*, and when you declare an `ImageView` object you specify the particular kind of data that it should contain. For example, you can make an `ImageView` of RGB (red/green/blue) pixels to represent a full-color image or an `ImageView` of vectors to represent a vector field. You specify the pixel type as a template parameter to the `ImageView` class like this:

```
ImageView<PixelRGB<float32> > my_image;
```

In this case we've made a full-color RGB image. Notice that `PixelRGB` is itself a template: here we've specified that we want each channel of each RGB pixel to be stored as a 32-bit floating-point number. All of the core pixel types in the Vision Workbench are themselves templates like this.

The `ImageView` class is defined in the C++ header file `<vw/Image/ImageView.h>`, and the standard pixel types are defined in the header `<vw/Image/PixelTypes.h>`. Thus, for the above line of code to compile you must include those two headers at the top of your program. (Alternatively, all of the header files relating to basic image manipulation are collected together in the convenience header `<vw/Image.h>`.) Furthermore, all of the core classes and functions of the Vision Workbench are defined in the C++ namespace `vw`. One way to use them is to be fully specific:

```
vw::ImageView<vw::PixelRGB<vw::float32> > my_image;
```

The other way, which may be simpler for new users, is to bring the entire `vw` namespace into the global scope by saying

```
using namespace vw;
```

at the top of your program after you've included the necessary headers. For brevity, in the examples in this book we will often assume that you have included the necessary headers and we will omit explicit references to namespace `vw`. The exception to this is the complete programs, such as `vwconvert.cc` (Listing 1, above), which are intended to be fully self-contained.

By default the dimensions of an `ImageView` are zero, which may not be what you want. One option is to specify an image's dimensions when we construct it:

```
ImageView<PixelRGB<float> > my_image( 320, 240 );
```

This creates an image with 320 columns and 240 rows. If we ever want to set or change the size of an image later on in the code we can use the `set_size()` method:

```
my_image.set_size( 640, 480 );
```

You can also find out how many columns or rows an image has using the `cols()` and `rows()` methods, respectively:

```
int width = my_image.cols();
int height = my_image.rows();
```

Note that when you call `set_size()` with new image dimensions the Vision Workbench allocates a new chunk of memory of the appropriate size. This is a destructive operation: any old data is not copied into the new buffer, and the old buffer will be automatically deallocated if no other objects are using it.

Once you've made an `ImageView`, the simplest way to access a particular pixel is by indexing directly into it:

```
PixelRGB<float> some_pixel = my_image( x, y );
```

In this example we've assumed that `x` and `y` are integer variables with the desired pixel's coordinates. For a less trivial example, one way to fill our image with the color red would be to loop over all the rows and columns, setting each pixel at a time:

```
PixelRGB<float> red(1.0, 0.0, 0.0);
for ( int y=0; y<my_image.rows(); ++y )
  for ( int x=0; x<my_image.cols(); ++x )
    my_image(x,y) = red;
```

This is not the fastest way to access the pixels of an image, but it is arguably the most flexible. (Later we will learn about much simpler ways to fill an image with a single color.)

3.1.2 The Standard Pixel Types

The Vision Workbench provides a number of standard pixel types that you can use to manipulate the most common sorts of images. We've already encountered `PixelRGB`, the standard RGB pixel type. As we mentioned earlier, this is a template class whose template parameter specifies the underlying numeric data type used to store each channel of the pixel. This is called the pixel's *channel type*. The Vision Workbench defines convenient platform-independent names for the standard channel types, so that you never have to worry about whether `int` or `short` is 16 bits wide on your platform.

Type	Description	Notes
int8	Signed 8-bit integer	
uint8	Unsigned 8-bit integer	Common for low-dynamic-range imaging
int16	Signed 16-bit integer	
uint16	Unsigned 16-bit integer	
int32	Signed 32-bit integer	
uint32	Unsigned 32-bit integer	
int64	Signed 64-bit integer	
uint64	Unsigned 64-bit integer	
float32	32-bit floating point	Common for high-dynamic-range imaging
float64	64-bit floating point	

Table 3.1: The standard Vision Workbench channel types.

Type	Description	Channels
PixelGray<T>	Grayscale	Grayscale value (v)
PixelGrayA<T>	Grayscale w/ alpha	Grayscale value (v), alpha (a)
PixelRGB<T>	RGB	Red (r), green (g), blue (b)
PixelRGBA<T>	RGB w/ alpha	Red (r), green (g), blue (b), alpha (a)
PixelHSV<T>	HSV	Hue (h), saturation (s), value (v)
PixelXYZ<T>	XYZ	CIE 1931 X (x), Y (y), and Z (z) channels
Vector<T,N>	An N-dimensional vector	N vector components
T	A unitless scalar	N/A

Table 3.2: The standard Vision Workbench pixel types. The channel type T should generally be one of the types from Table 3.1.

These Vision Workbench channel types are listed in Table 3.1. These are the only channel types with which the Vision Workbench has been tested, so it is best to stick to these unless you have a compelling reason not to.

The standard pixel types are listed in Table 3.2. The first four, used for grayscale and RGB images with and without alpha channels, are the most common. (For those of you who are unfamiliar with the term, an *alpha* channel is used to represent the opacity of a pixel. For the rest of you, note that the Vision Workbench generally stores alpha pixels in pre-multiplied form.)

Each of the channels in a pixel can be accessed by indexing into it directly, as in `my_pixel(i)` or `my_pixel[i]`. The order of the channels is the same as the order in which they appear in the name of the type. If you know a particular pixel's type you can also access it's channels by name, so for example `my_rgb_pixel.r()` access an RGB pixel's red channel. (Note that grayscale values are accessed via `v()`, for "value".)

When you are writing Vision Workbench programs you may often find yourself working with only one pixel type at a time. In this case it can be convenient to place a `typedef` near the top of your file defining a convenient shorthand:

```
typedef vw::ImageView<vw::PixelRGB<float32> > Image;
```

This way you can refer to your RGB image type by the much shorter identifier `Image`. In the remainder of this book when we say `Image` you may assume that you may substitute the `ImageView` class type that is most appropriate for your application.

Standard conversions are provided among all the RGB and grayscale pixel types, and also between `PixelRGB` and the special color types `PixelHSV` and `PixelXYZ`. The `ImageView` class can take advantage of these pixel conversions to perform color space conversion on entire images. For example, images are generally stored on disk in an RGB color space but it is sometimes helpful to convert them to HSV for processing. This is easy with the Vision Workbench:

```
ImageView<PixelRGB<float> > rgb_image;
read_image( rgb_image, filename );
// Convert the RGB image to HSV:
ImageView<PixelHSV<float> > hsv_image = rgb_image;
```

(We'll have more to say about `read_image()` shortly, but it does what you'd expect.) Later you could assign the HSV image back to an RGB image prior to saving it to disk.

3.1.3 Copying `ImageViews`

In the Vision Workbench, `ImageView` objects have *shallow copy semantics*. That is, when you copy an `ImageView` you're making a new `ImageView` that points to the *same* data, rather than a new copy of the data. This is a relatively inexpensive operation, which makes it perfectly reasonable to do things like construct a `std::vector` of `ImageViews`. The underlying image data is reference-counted, and when the last `ImageView` stops using a block of image data it is deallocated.

Though this behavior can be quite powerful, it may not always be what you want. If you ever need to make a duplicate of an `ImageView`, so that you can modify one without affecting the other, you should use the `copy()` function found in `<vw/Image/Algorithms.h>`.

```
// This makes a shallow copy, pointing to the same:
Image new_image_1 = my_image;
// This makes a deep copy, pointing to new, identical data:
Image new_image_2 = copy( my_image );
```

It is important to understand that this shallow copy behavior only applies when the source and destination image types—and in particular the source and destination pixel types—are *identical*. If the pixel types are different then you are not actually making a copy in the C++ sense of the word but are instead assigning one image view to another. In the above example involving RGB and HSV images, even though the source and destination objects are both `ImageViews` they in fact have different types because they have different template parameters. Therefore the data is copied deeply while being converted to the new pixel type. This holds even if the source and destination pixel types differ only in their underlying channel type.

3.1.4 `ImageView` as a STL-Compatible Container

An `ImageView` can be thought of as a container of pixels, and in fact you can use it as a standard C++ container class. The iterator type is, as expected, called `ImageView<T>::iterator`, and it allows you to access each of the pixels of an image one at a time. The `begin()` and `end()` methods return iterators pointing to the first and one-past-the-last pixels, respectively. The first pixel is located at position $(0, 0)$, and incrementing the iterator advances to the next column. After it passes through the last column, the iterator wraps around to the beginning of the next row.

This C++ Standard Template Library (STL) compliant iterator exists mainly to allow you to take advantage of the many *algorithms* provided by the STL that operate on containers. For example, you can use `sort()` to sort all of the pixel values in an image.

```
std::sort( my_image.begin(), my_image.end() );
```

That particular example may be more cute than it is useful, but others occur more frequently. For instance, you can use `std::count()` to count the number of pixels with a particular value, or `std::replace()` to replace all pixels that have one value with another.

3.1.5 Image Planes

The `ImageView` class also supports another feature found in many other image processing libraries: image planes. Like duct tape, planes are the wrong solution to almost every problem, and we discourage their use. Basically, planes allow you to store some number of two-dimensional pixel arrays of the same size ("planes") together in a single object. Planes are different from channels in that the number and meaning the planes is not specified at compile time. This means that the Vision Workbench can not take advantage of that information as readily: for example, it has no way to know whether a three-plane image is RGB, HSV, or something altogether different, and it cannot optimize operations by unrolling inner loops as it is able to with channels. (It may not be readily apparent, but the sample program shown in Listing 1 demonstrates one of the very few possibly-legitimate uses of planes; this will be discussed more in the following section on File I/O.)

To create a multi-plane image, pass the desired number of planes as a third argument to the `ImageView` constructor or to the `set_size()` method. You can query the number of planes in an image with the `planes()` method. To access a pixel in particular plane of an image, pass the plane as a third argument when indexing into the image.

```
Image my_image(320,240,3);      // A 3-plane image
my_image.set_size(320,240,3);   // Same here
int planes = my_image.planes(); // Now planes == 3
Pixel pix = my_image(x,y,p);    // Access a pixel
```

Once again, if you are thinking about using planes we encourage you to first consider these alternatives. If you want a way to store a collection of related images, consider using a `std::vector` of `ImageView`s instead. If you just want to store a bunch of numbers at each pixel location, consider using `Vector<T,N>` as a pixel type.

3.2 Image File I/O

The most common way to get image data into and out of the Vision Workbench is by loading and saving images using file I/O. There are several mechanisms for doing this, varying in complexity, flexibility and (for the time being) completeness of implementation.

3.2.1 Reading and Writing Image Files

The simplest method for file I/O is to use the `read_image()` and `write_image()` functions, passing them an `ImageView` and the filename of the image file on disk that you would like to read from or write to.

```
read_image( image, filename );
write_image( filename, image );
```

Name	Extension(s)	Description
PNG	`.png`	Standard for loss-less compression
JFIF/JPEG	`.jpg`, `.jpeg`	Standard for lossy compression, no alpha
TIFF	`.tif`, `.tiff`	Highly flexible, complicated
OpenEXR	`.exr`	High dynamic range
PDS	`.img`	Planetary Data System images

Table 3.3: The standard Vision Workbench image file formats. Which formats your installation supports depends on what supporting libraries you have installed. Adding support for additional file formats is discussed in Chapter 14.

Notice that the order of arguments to these two functions is reversed: in both cases the destination is first and the source second.

Both functions determine the image file type by looking at the extension of the filename that you provide them. The exact set of file formats that are supported depends on which file format libraries the Vision Workbench found on your system when you build it. For example JPEG support depends on `libjpeg`, and so forth. The file formats that the Vision Workbench is designed to support are listed in Table 3.3. Note that the file extensions are case-insensitive.

Image data on disk is generally stored with one of the four standard pixel types: grayscale or RGB with or without alpha. The image reading and writing routines will freely convert between these formats. You should generally create an `ImageView` with the pixel type that you would like to work with and let the file I/O system take care of the rest.

```
ImageView<PixelGrayA<float> > image;
read_image( image, "some_file.jpg" );
```

In this example we loaded in a JPEG image file (which has an RGB pixel format) and then converted the data grayscale and padded it with a constant alpha value of 1.0, corresponding to fully opaque. Attempting to save this image back as a JPEG file would reverse the conversion. (Any transparency is composited on to a black background whenever the alpha channel is removed.)

3.2.2 More Sophisticated File I/O

We will only provide an overview of the more advanced file I/O techniques here. Many of them are partially (in some cases barely) implemented. If you want to use any of these features you can learn more about them in Chapter 14.

Images on disk are handled via an abstract image resource class, called `DiskImageResource` and defined in `<vw/FileIO/DiskImageResource.h>`. You can create one directly using the same file-extension-based file type deduction mechanism discussed above.

```
DiskImageResource *dir1 = DiskImageResource::open( filename );
DiskImageResource *dir2 = DiskImageResource::create( filename, format );
```

In the first case we are opening an existing file, and in the second case we are creating a new file. Creating a new file resource requires providing some hints about the underlying image format, such as its dimensions and pixel type, which are supplied by a `GenericImageFormat` object.

Once you have a resource you can query it for information about its dimensions, pixel format and channel type. For example, you can choose to process different pixel formats differently.

```
switch( dir1->pixel_format() ) {
  case VW_PIXEL_GRAY: /* process grayscale file */ break;
  case VW_PIXEL_RGB:  /* process RGB file */       break;
  /* ... */
}
```

You can use the `DiskImageResource`'s `read()` and `write()` methods to read the data into or write the data out of an `ImageView`, respectively.

If you wish to force a particular file format, you can create a resource object of the appropriate type directly.

```
DiskImageResourcePNG *dirp1 = new DiskImageResourcePNG( filename );
DiskImageResourcePNG *dirp2 = new DiskImageResourcePNG( filename, format );
```

In this case we show how to create PNG image resources. If you do this then you can take advantage of any special services provided by the particular file format's resource type, such as the ability to read or write special file header information.

Finally, you can make a read-only `ImageView`-like object that corresponds to an image on disk. This is called a `DiskImageView` and is defined in the header of the same name. This can be used to process images that are too large to be loaded into memory all at once.

3.3 Manipulating Images

We have seen how images are represented via the `ImageView` class, how to save and load them to and from disk, and how to manipulate their pixels individually. Now it is time to begin discussing how to perform slightly higher-level operations on images.

3.3.1 Simple Image Manipulation

We begin with the simple image manipulation functions listed in Table 3.4 and defined in the header file `<vw/Image/Manipulation.h>`. Many of these should be self-explanatory. The results of applying several of these transforms to an image are shown in Figures 3.1(b)–3.1(i). The 90-degree rotation functions are one of the few places where the Vision Workbench makes any kind of assumption about the interpretation of the x, y coordinate system. When it is necessary to make a distinction we assume that the origin $(0, 0)$ is the top-left corner of the image. If you have been interpreting the origin as the top-right or bottom-left you will need to invert your notion of clockwise vs. counter-clockwise when calling these two functions.

None of these functions, by themselves, modify image data or produce new images. Instead, each function returns a special *view* on to the same image data. In most cases you will assign the result to another `ImageView`, causing the data to be processed and the resulting image to be stored in the new buffer:

```
image2 = flip_vertical( image1 );
```

It's worth taking a moment to study exactly what goes on behind the scenes when you perform an operation like this. First the Vision Workbench resizes the destination image (`image2` in the above example) if necessary so that its dimensions are the same as those of the source image (a flipped version of `image1`). Second it computes the result of the operation, storing the result in the destination image as it goes. The important point is that if the destination image already has the

Function	Description
`rotate_180(im)`	Rotate the image 180 degrees
`rotate_90_cw(im)`	Rotate the image 90 degrees clockwise
`rotate_90_ccw(im)`	Rotate the image 90 degrees counter-clockwise
`flip_vertical(im)`	Flip the image vertically
`flip_horizontal(im)`	Flip the image horizontally
`transpose(im)`	Transpose the `x` and `y` coordinates of the image
`crop(im,x,y,c,r)`	Crop the image, specifying (x, y) and $(cols, rows)$
`crop(im,bbox)`	Crop the image, specifying a bounding box
`subsample(im,factor)`	Subsample the image by an integer factor
`subsample(im,xfac,yfac)`	Subsample the image by integer factors in x and y
`select_col(im,col)`	Refers to an individual column of an image
`select_row(im,row)`	Refers to an individual row of an image
`select_plane(im,plane)`	Refers to an individual plane of an image
`select_channel(im,channel)`	Refers to an individual channel of an image
`channels_to_planes(im)`	Interprets a multi-channel image as a multi-plane image
`pixel_cast<PixelT>(im)`	Casts an image to a new pixel type
`pixel_cast_rescale<PixelT>(im)`	Casts an image to a new pixel type, with rescaling
`channel_cast<ChanT>(im)`	Casts an image to a new channel type
`channel_cast_rescale<ChanT>(im)`	Casts an image to a new channel type, with rescaling
`planes_to_channels<PixelT>(im)`	Interprets a multi-plane image as a multi-channel image
`weighted_rgb_to_gray(im)`	Converts RGB to grayscale with default weights
`weighted_rgb_to_gray(im,r,g,b)`	Converts RGB grayscale with the given weights

Table 3.4: The simple image manipulation functions, defined in the header file `<vw/Image/Manipulation.h>`. The functions in the top section return writable views.

same dimensions as the source image then it is *not* resized or reallocated. This avoids unnecessary memory allocations in common situations, such as when you are processing many identically-sized images in a loop. However, it also means that you must be careful when processing an image and assigning it back to itself:

```
image = flip_vertical( image );   // Bad idea: self-assignment
```

In this example, the destination image clearly has the same dimensions as the source (since they are the same image) and so no new image buffer is allocated. As a result the `flip_vertical` operation will clobber the source image with partial results, producing garbage. One solution to this problem is to force the creation of a temporary buffer using the `copy` function:

```
image = copy( flip_vertical( image ) );   // Much better
```

The functions listed in the upper section of Table 3.4 all provide new ways of accessing the same data without doing any additional processing. As a result, these functions are all able to return *writable* views of their image argument. That is, you can use them to modify an image by placing them on the left side of an equals sign. For example, suppose you want to add a small inset to a larger image, by copying a small image into the larger one at a particular position. One easy way is to specify the destination region using the `crop()` function:

Function	Description
`copy(im)`	Produce a deep copy of an image
`fill(im,value)`	Fill an image with a pixel value *in-place*
`clamp(im,[low],[high])`	Clamp values to the given range
`normalize(im,[low],[high])`	Normalize values to the given range
`threshold(im,[thresh],[low],[high])`	Threshold an image to two values
`grassfire(im)`	Compute the grassfire image of an image
`blob_index(im)`	Apply index numbers to valid regions of an image
`bounding_box(im)`	Return the bounding box of an image
`nonzero_data_bounding_box(im)`	Compute the bounding box of nonzero data
`image_blocks(im,width,height)`	Tile an image with bounding boxes

Table 3.5: The simple image algorithms defined in the header file `<vw/Image/Algorithms.h>`.

```
int cols = small_image.cols(), rows = small_image.rows();
crop( large_image, xpos, ypos, cols, rows ) = small_image;
```

Here we've cropped a region of the large image and used it for writing instead of reading. Note that the assignment proceeds just as before: first the destination image dimensions are checked, and then the data is copied. However in this case the Vision Workbench will throw an exception if the dimensions differ, since it is not meaningful to "resize" a cropped region in the same sense that you can freely resize an `ImageView`. This approach can also be used, for example, to replace one channel of a multi-channel image using `select_channel()`.

The functions listed in the lower section of Table 3.4, on the other hand, all do a small amount of processing of pixel values. The `pixel_cast()` function converts all the pixels in an image to the given new pixel type. The `pixel_cast_rescale()` variants rescale the values if the channel type has changed, e.g. mapping the 0–255 range of `uint8` on to the 0.0–1.0 nominal range of `float32`. The `channel_*` variants cast the pixels to have the given new channel type, leaving the overall pixel format unchanged. The `pixels_to_channels()` function takes a multi-plane image and reinterprets it as a multi-channel image with the given pixel type. Finally, `weighted_rgb_to_gray` converts RGB pixels to the corresponding grayscale pixel type using an arbitrary weighting of the red, green, and blue channels. The default weights are based on a human perceptual model that weights green most strongly, followed by red and then blue.

3.3.2 Image Algorithms

We will now introduce a number of additional simple image operations that are defined in the header file `<vw/Image/Algorithms.h>`. You have already seen one of them, `copy()`, which forces the creation of a deep copy of an image in a new buffer. The rest are listed in Table 3.5. The result of two of these functions can be seen in Figures 3.1(j) and 3.1(k). We hope to implement a number of additional image algorithms, mirroring the STL container algorithms but optimized for images, at some point in the future.

The `fill()` function is noteworthy because it is currently the only core Vision Workbench function that modifies image data *in-place*. It is especially useful for filling a single channel of an image. For example, you can use it to make an RGBA image fully opaque.

```
fill( select_channel( rgba_image, 3 ), 1.0 );
```

(a) `mural` (Original)

(b) `rotate_180(mural)`

(c) `rotate_90_cw(mural)`

(d) `rotate_90_ccw(mural)`

(e) `transpose(mural)`

(f) `flip_vertical(mural)`

(g) `flip_horizontal(mural)`

(h) `crop(mural,80,60,160,120)`

(i) `subsample(mural,2)`

(j) `threshold(mural,0.5)`

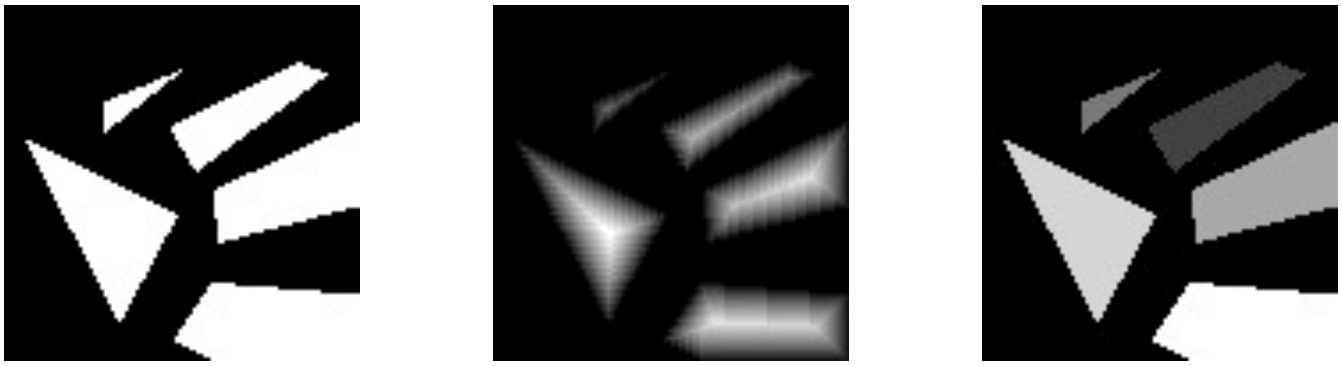
(k) `clamp(mural,0.25,0.75)`

Figure 3.1: Sample output from the simple image operations discussed in this section.

(Note that 1.0 represents fully-opaque if the image has a floating-point channel type.)

The `clamp()`, `normalize()`, and `threshold()` functions return modified versions of their image arguments. You can assign the result back to the original image, or you can save it in a different image instead and keep the original. The `clamp()` function clamps the values in the image to the given range. The `normalize` function scales and shifts the values of an image so that the values span the specified range. The default range is from zero to the nominal maximum value for the channel type, e.g. 1.0 for floating-point images. This is particularly useful for saving intermediate results of your algorithms to disk for debugging. Finally, the `threshold` function returns a two-valued image based on whether the pixels in the source image is greater than or less than the given threshold value. The default high and low output values are the same as for `norm`, and the default threshold is zero. For example, this line will convert a floating-point grayscale image to pure black-and-white:

```
image = threshold( image, 0.5 );
```



(a) `pattern` (Original)

(b) `normalize( channel_cast<float>( grassfire(pattern)))`

(c) `normalize( channel_cast<float>( blob_index( create_mask(pattern))))`

Figure 3.2: Sample output from more complex operations.

The `grassfire()` algorithm, named for the algorithm that it implements, is more specialized. It takes an image and efficiently computes how far each pixel is from from a pixel whose value is zero, assuming that pixels outside the image boundaries all have zero value. It measures distance in the four-connected Manhattan sense, i.e. as the sum of the horizontal and vertical distances. This algorithm is used in a variety of applications, such as avoiding obstacles and unknown terrain in path planning.

The `blob_index()` algorithm, applies an index value to isolated sections of images label valid. The determination of a pixel's validity is from a special pixel called PixelMask. PixelMask is discribed in the *Pixels Types* chapter. `blob_index` is useful algorithm for segmenting an image.

Chapter 4

Image Processing

Now that we've covered all the basics of how to manipulate images, it's time to move on to some more interesting image processing tasks. We begin with an introduction to image filtering, followed by a discussion of image math. We then take a brief detour to introduce the Vision Workbench's `Vector` and `Matrix` classes before describing image transformation and warping.

By the end of this chapter you will have encountered all of the core building blocks that comprise the heart of the Vision Workbench. There are a number of directions that you can go from here, depending on what you are hoping to accomplish. We conclude this chapter with an overview of the many more specialized features of the Vision Workbench and a discussion of where to look (in this book and elsewhere) in order to learn more about them.

4.1 Image Filtering

Image filtering has traditionally been the bread and butter of image processing software packages. The Vision Workbench includes a number of functions to perform the most common filtering operations. We will first describe the special-purpose filters, and then we will discuss the more general convolution-based linear filtering functions. All of the filter functions discussed in this section are defined in the header file `<vw/Image/Filter.h>`. We will not discuss frequency-domain filtering in this chapter; that is covered later in Section 14.7.

4.1.1 The Special-Purpose Filters

At the moment only three special-purpose filters are fully supported. The first is a Gaussian smoothing or blurring filter, which convolves the image with a discrete Gaussian kernel that has a user-specified standard deviation (a.k.a. "sigma") and user-specified size in each axis. In order for the filter to accurately approximate a Gaussian, the size of the kernel should be at least a few times the standard deviation. However, unnecessary computation is performed if the size is much larger than that. You can omit the size arguments, in which case the function will pick a kernel size based on your standard deviation that is reasonable for most applications. In the most common case the two standard deviations are equal, in which case you need only specify a single value for sigma.

```
result = gaussian_filter( image, sigma );
result = gaussian_filter( image, xsigma, ysigma );
result = gaussian_filter( image, xsigma, ysigma, xsize, ysize );
```

Function	Description
`gaussian_filter(im,...)`	Apply a Gaussian smoothing filter to an image
`derivative_filter(im,...)`	Apply a discrete differentiation filter to an image
`laplacian_filter(im,...)`	Apply a discrete Laplacian filter to an image
`convolution_filter(im,...)`	Apply a general 2D convolution filter to an image
`separable_convolution_filter(im,...)`	Apply a separable convolution filter to an image

Table 4.1: The Vision Workbench image filtering functions, defined in `<vw/Image/Filter.h>`.

Type	Description
`ConstantEdgeExtension`	Extends an image with constant (i.e. nearest-neighbor) values
`ZeroEdgeExtension`	Extends an image with a zero value in all directions
`ReflectEdgeExtension`	Extends an image by reflecting across its edges
`PeriodicEdgeExtension`	Extends an image by repeating it periodically

Table 4.2: The edge extension modes.

In these examples, the `sigma` arguments are generally floating-point whereas the `size` variables are integers.

The next filter is the derivative filter, which performs a discrete spatial differentiation of your image. Here again, you can specify the order of differentiation in the two axes as well as the filter kernel size.

```
result = derivative_filter( image, xderiv, yderiv );
result = derivative_filter( image, xderiv, yderiv, xsize, ysize );
```

There is a minimum filter size below which it is not possible compute any given derivative, and these functions will throw an exception if you try. For the most part it is a good idea to just let the Vision Workbench pick the kernel size.

The final special-purpose filter is the Laplacian filter, which performs a discrete approximation to the Laplacian operation $\nabla^2 = \frac{d^2}{dx^2} + \frac{d^2}{dy^2}$.

```
result = laplacian_filter( image );
```

This filter does not take any special parameters. Note that if you are accustomed to using a "larger" derivative or Laplacian filter to reduce the effect of noise, you are probably better off applying a smoothing operation (e.g. via `gaussian_filter()`) first.

4.1.2 Edge Extension Modes

To filter the regions near the edges of an image properly, filters like these need to make some sort of assumption about the contents of the source image *beyond* the image boundaries. This is generally referred to as "edge extension". The default assumption made by the filters discussed in this section is that in each direction the image is extended with a constant value equal to the value of the nearest edge pixel. However, you can specify an alternative edge extension mode if you wish, by passing an extra argument to the filters. The C++ type of the argument determines the edge extension mode used.

```
result = gaussian_filter( image, 3.0, ConstantEdgeExtension() );
result = gaussian_filter( image, 3.0, ZeroEdgeExtension() );
```

Both of these examples filter the source image using a standard deviation of three pixels and an automatically-chosen kernel size. However, the first explicitly requests the default edge extension behavior, while the second requests that the source image be assumed to be zero outside the image boundaries.

Notice the "extra" set of parentheses after the names of the edge extension modes. Remember that those names are C++ *types*, and you can only pass an *object* as an argument to a function. Those parentheses invoke the edge extension type's constructor, returning a dummy object that you pass as the final argument to the filtering function. If you find this confusing, don't worry too much about it right now. Just keep in mind that when you're using a type as an argument to a function to change its behavior you need the extra parentheses. The types that are currently supported as edge extension modes are listed in Table 4.2.

4.1.3 General Convolution Filtering

Most of the filters used in image processing are convolution filters, which express each output pixel as a fixed weighted sum of neighboring input pixels. An image convolution filter is usually described by a rectangular array of weights called the *kernel*. The easiest way to think about an image kernel is as the result that you would desire from the filter if the input image had the value 1 at the origin and zero everywhere else. (This is also known as the "impulse response" of the filter.) For example, a first-order derivative filter in the x direction might have the kernel $[\ 1 \quad 0 \quad -1\ ]$. In this case we also need to know that the middle number of the kernel (the zero in this case) is the kernel's origin.

In the Vision Workbench, convolution kernels—which as we've said are nothing more than rectangular arrays of numbers—are represented by images. The pixel type for a kernel should generally be a scalar type such as `float`. Once you've put the kernel that you'd like into an image it is straightforward to use it to filter another image.

```
ImageView<float> kernel;
/* set up your kernel here */
result = convolution_filter( image, kernel );
```

In this case the Vision Workbench assumes that the center pixel of the kernel is the kernel's origin. If this is not what you want then you can specify the coordinates of the kernel's origin explicitly instead.

```
result = convolution_filter( image, kernel, ox, oy );
```

In either case you can also optionally specify an edge extension mode, just like you could for the special-purpose filters.

Convolution filtering can be computationally expensive if the kernel is large. Fortunately, many useful kernels have a special form that makes it possible to improve the performance considerably. These are called *separable* kernels, and are themselves the result of convolving a single-column image with a single-row image. In other words, the kernel K must satisfy $K(x, y) = K_x(x)K_y(y)$ for some functions K_x and K_y. The Gaussian and derivative filters are both of this form, for example, though the Laplacian filter is not.

The Vision Workbench provides special support for efficient convolution filtering with separable kernels. You must supply the *separated* kernel, i.e. two one-dimensional kernels.

```
result = separable_convolution_filter( image, xkernel, ykernel );
result = separable_convolution_filter( image, xkernel, ykernel, ox, oy );
```

Per-pixel Sum	Per-pixel Difference	Per-pixel Product	Per-pixel Quotient
`image + image`	`image - image`	`image * image`	`image / image`
`image += image`	`image -= image`	`image *= image`	`image /= image`
`image + value`	`image - value`	`image * value`	`image / value`
`image += value`	`image -= value`	`image *= value`	`image /= value`
`value + image`	`value - image`	`value * image`	`value / image`

Table 4.3: The Vision Workbench image operators are included automatically when you include `<vw/Image/ImageMath.h>`).

As in the general 2D convolution case, the origin of the kernel is assumed to be in the middle if you do not specify otherwise and in either case you can add an optional argument specifying the edge extension mode. You can still supply the one-dimensional kernels as images, just as you did in the general 2D convolution case, but here you can also provide them in another STL-compliant container, such as a `std::vector` or (as we shall introduce later this chapter) a `vw::Vector`. If you do chose to represent the kernels as images, remember that each should have one of the dimensions set to 1.

4.2 Doing Math with Images

In image processing it is often desirable to perform some mathematical operation on every pixel of an image, or to corresponding pixels from several images. For example gamma correction involves applying a mathematical function to each pixel, and background subtraction involves subtracting the corresponding pixels from two images. In the Vision Workbench, these operations and others like them fall under the rubric of "image math", and the functions to support them are defined in the header `<vw/Image/ImageMath.h>`.

4.2.1 Image Operators

In most cases writing code to perform image math is trivial. The mathematical expressions that you would normally write for individual pixels work just as well for whole images of pixels. For example, consider the background subtraction problem mentioned above.

```
result_image = input_image - background_image;
```

That's all there is to it. Setting up an IIR low-pass filter to estimate the background image is just as easy.

```
background_image = alpha*input_image + (1-alpha)*background_image;
```

(Here we're assuming that `alpha` is a small positive floating-point number.) The important point is that there is no need for you to write a loop that performs an operation like this on each pixel. Just write the mathematical expression, replacing pixels with images, and you're all set.

This works, of course, because the Vision Workbench has overloaded the standard C++ mathematical operators to work on images. These operators are listed in Table 4.3. Operation with scalars is treated identically to per-pixel operation with constant-value images. In order to simplify division with large images, the image division operators have been designed so that division by zero returns zero instead of throwing an exception.

Function	Description	Function	Description
`sin`	Sine, $\sin x$	`asin`	Inverse sine, $\sin^{-1} x$
`cos`	Cosine, $\cos x$	`acos`	Inverse cosine, $\cos^{-1} x$
`tan`	Tangent, $\tan x$	`atan`	Inverse tangent, $\tan^{-1} x$
`atan2`	Two-argument form of inverse tangent, $\tan^{-1} x/y$		
`sinh`	Hyperbolic sine, $\sinh x$	`cosh`	Hyperbolic cosine, $\cosh x$
`tanh`	Hyperbolic tangent, $\tanh x$	`exp`	Exponential, e^x
`log`	Natural logarithm, $\ln x$	`log10`	Base-10 logarithm, $\log_{10} x$
`ceil`	Ceiling function, $\lceil x \rceil$	`floor`	Floor function, $\lfloor x \rfloor$
`sqrt`	Square root, $\sqrt{x}$	`pow`	Power function, x^y
`asinh`	Inverse hyperbolic sine, $\sinh^{-1} x$	`acosh`	Inverse hyperbolic cosine, $\cosh^{-1} x$
`atanh`	Inverse hyberbolic tangent, $\tanh^{-1} x$	`cbrt`	Cube root, $\sqrt[3]{x}$
`exp2`	Base-2 exponential, 2^x	`expm1`	Exponential minus 1, $e^x - 1$
`log2`	Base-2 logarithm, $\log_2 x$	`log1p`	Lograithm of one-plus, $\ln(1 + x)$
`tgamma`	Gamma function, $\Gamma(x)$	`lgamma`	Log of Gamma function, $\ln \lvert\Gamma(x)\rvert$
`hypot`	Hypotenuse, $\sqrt{x^2 + y^2}$	`copysign`	Sign-copying function
`round`	Rounding function	`trunc`	Floating-point truncation
`fdim`	Positive difference, $\max(x - y, 0)$		

Table 4.4: The Vision Workbench image math functions, as defined in `<vw/Image/ImageMath.h>`. The functions in the bottom section are not available under the Windows operating system.

There is one important issue to bear in mind when using image operators: the underlying per-pixel operations must themselves be meaningful. For example, multiplying an image whose pixel type is `PixelGray` by an image whose pixel type is `PixelRGB` is not well-defined, and attempting to do so will result in a compiler error. The Vision Workbench will not automatically "promote" the grayscale image to RGB.

This raises the question of what happens when you multiply two images both of whose pixel type is, for example, `PixelRGB`. What does it mean to multiply two RGB colors? Multiplication is defined for numbers, not colors. The answer is that in this situation the Vision Workbench will actually perform the mathematical operation on a per-*channel* basis rather than just a per-pixel basis.

A good rule of thumb when working with image operators is to restrict yourself to operating on images of the same type, or combinations of images of one type and images of scalars. As long as you obey this rule you should find that the image operators always do what you expect.

4.2.2 Mathematical Functions

Of course, C++ provides a range of mathematical functions, too, such as exponentials and logarithms, trigonometric functions, and so forth. The Vision Workbench extends these functions to operate on images as well. The supported functions are listed in Table 4.4. Note that these image functions are built on top of the standard C++ functions that operate on regular numbers. Therefore, the Vision Workbench only supports those functions that are provided by your platform. In

particular, the bottom half of Table 4.4 lists functions that are *not* currently available under the Microsoft Windows operating system.

You can use these functions just like you use the mathematical operators: write the same expression that you would write for individual pixels, but substitute whole images instead.

```
float gamma = 1.8;
result_image = pow( input_image, gamma );
```

This example demonstrates how to use the `pow()` function to gamma-correct an image. Here the variable `gamma` is a floating-point number representing the desired gamma correction factor for the entire image. However, if instead we wanted to apply a variable gamma correction factor on a per-pixel basis, the following code would do the trick.

```
ImageView<float> gamma_image; // Initialize with different gamma values
result_image = pow( input_image, gamma_image );
```

This example demonstrates that the arguments of a two-argument mathematical function can be either scalar or image values. Just as with the operators, scalar arguments are treated the just like a constant-value image.

Note that unlike the normal mathematical functions that C++ inherited from C, it is not necessary (or correct) to use a different function name when you are working with `float` image data than you would use to work with `double` image data. The function names listed in Table 4.4 are correct for image math in all cases. Those in turn use the proper underlying mathematical functions as appropriate—for example, `sin()` invokes `sinf()` on each pixel if it is applied to a `float`-based image.

4.3 Vectors and Matrices

Before introducing the next image processing topic, image transformation and warping, we must first take a brief detour to introduce the Vision Workbench vector and matrix classes. We will assume in this chapter that you have a good familiarity with the underlying mathematical entities that these classes represent. Note that our mathematical usage of the word "vector" here is somewhat different from the C++ standard library's use of the word to mean a dynamically-resizable array.

4.3.1 Vectors and Vector Operations

The Vision workbench vector class is called, appropriately enough, `Vector`. Like `ImageView`, `Vector` is a template class whose first template parameter is required and specifies the underlying numeric type. However, while the dimensions of an image are always specified at run-time via the image's constructor or the `set_size()` method, `Vector` comes in two variants. The first form behaves in just the same way, but the second form has a fixed size that is specified at compile time. This eliminates the need for frequent dynamic allocation when working with vectors in the common case when the vector dimension is known.

Declaring either type of vector is straightforward:

```
Vector<float> vector1(3);
Vector<float,3> vector2;
```

Both of those statements declare three-dimensional vectors of floating-point numbers. In the first case the vector is allocated dynamically on the heap and the size could have been chosen at run-time. In the second case the vector is allocated statically on the stack, but the dimension can *not* vary at run time. The first form is generally useful when, say, reading a large vector of data in from a file, while the second form is more useful when performing geometric computations.

The second, fixed-dimension form also has special constructors that you can use to initialize the vector contents:

```
Vector<float,3> vector2(1,2,3);
```

These constructors are available with up to four arguments. Alternatively, you can construct both fixed-size and dynamically-sized vector with data copied from a block of memory that you point them to:

```
float *some_data;
Vector<float> vector1(3, some_data);
Vector<float,3> vector2(some_data);
```

Remember that this copies the data, so it can be inefficient; see the discussion of `VectorProxy` below for an alternative. Three of the most commonly used vector types have special aliases, for convenience:

```
typedef Vector<double,2> Vector2;
typedef Vector<double,3> Vector3;
typedef Vector<double,4> Vector4;
```

These types are used throughout the Vision Workbench as the standard geometric vector types.

You can query a vector about its size (i.e. dimension or length) with the `size()` method, and you can index into a vector to access individual elements:

```
for( unsigned i=0; i<vector1.size(); ++i ) vector1(i) = 0;
```

This example loops over all the elements of a vector, setting them to zero. You can also into a vector with square brackets instead of parentheses if you prefer. For fixed-length vectors there is one more way to access up to the first three elements, via methods called `x()`, `y()`, and `z()`.

```
vector2.x() = 0;  // Set the first element to zero
```

These methods are only available if the vector has sufficient length. For example, attempting to use the `z()` method of a vector of type `Vector<float,2>` will result in a compile-time error. Remember, these methods are only available for fixed-size vectors, *not* dynamically-sized ones. Dynamically-sized vectors, however, can be resized:

```
vector1.set_size(10);
```

The `set_size()` function takes an optional second argument that specifies whether or not the vector contents should be preserved. This argument defaults to `false`, so in the above example the old contents (if any) are lost.

The `Vector` classes support the standard mathematical operations of vector addition and subtraction and scalar multiplication and division via the usual C++ operators. They also support the a range of elementwise mathematical operations, such as adding a scalar to each element or multiplying the corresponding elements of two vectors, via functions of the form `elem_*`. There are

Function	Description
`- vector`	Vector negation
`vector + vector`	Vector sum
`vector - vector`	Vector difference
`vector * scalar`	Scalar product
`scalar * vector`	Scalar product
`vector / scalar`	Scalar quotient
`vector += vector`	Vector sum assignment
`vector -= vector`	Vector difference assignment
`vector *= scalar`	Scalar product assignment
`vector /= scalar`	Scalar quotient assignment
`elem_sum(vector,vector)`	Elementwise vector sum (same as `+` operator)
`elem_sum(vector,scalar)`	Elementwise sum of a vector and a scalar
`elem_sum(scalar,vector)`	Elementwise sum of a scalar and a vector
`elem_diff(vector,vector)`	Elementwise vector difference (same as `-` operator)
`elem_diff(vector,scalar)`	Elementwise difference of a vector and a scalar
`elem_diff(scalar,vector)`	Elementwise difference of a scalar and a vector
`elem_prod(vector,vector)`	Elementwise product of two vectors
`elem_prod(vector,scalar)`	Elementwise vector product (same as `*` operator)
`elem_prod(scalar,vector)`	Elementwise vector product (same as `*` operator)
`elem_quot(vector,vector)`	Elementwise quotient of two vectors
`elem_quot(vector,scalar)`	Elementwise quotient (same as `/` operator)
`elem_quot(scalar,vector)`	Elementwise quotient of a scalar and a vector
`norm_1(vector)`	1-norm of a vector, i.e. $\sum \lvert v_i \rvert$
`norm_2(vector)`	Euclidean 2-norm of a vector, i.e. $\sqrt{\sum v_i^2}$
`norm_2_sqr(vector)`	Squared 2-norm of a vector, i.e. $\sum v_i^2$
`norm_inf(vector)`	Infinity-norm of a vector, i.e. $\max \lvert v_i \rvert$
`sum(vector)`	Sum of elements, i.e. $\sum v_i$
`prod(vector)`	Product of elements, i.e. $\prod v_i$
`normalize(vector)`	The normalized form of a vector, i.e. $v/\lvert v \rvert$
`dot_prod(vector,vector)`	Vector dot product, i.e. $u \cdot v$
`cross_prod(vector,vector)`	Vector cross product, i.e. $u \times v$

Table 4.5: The vector math functions defined in `<vw/Math/Vector.h>`.

a number of vector norms and related functions, as well as a vector dot product and cross product. (The cross product is, of course, only valid for three-dimensional vectors.) The complete list of vector math functions defined in `<vw/Math/Vector.h>` is given in Table 4.5.

A `Vector` object is also a *container* in the C++ Standard Template Library sense of the word. There is a `Vector<...>::iterator` type that serves as the vector's iterator, and there are `begin()` and `end()` methods that return iterators to the first and one-past-the-last elements, as usual. This can be an extremely convenient way to load data into and out of `Vector`s.

You can extract a portion of a vector using the `subvector()` function, which takes three arguments: the original vector, the position of the first element to extract, and the number of elements in the resulting vector:

```
Vector<float,3> vector2 = subvector(vector1,5,3);
```

This example copies the fifth, sixth, and seventh elements of `vector1` into a new three-element vector.

The streaming operator `<<` is also defined for writing vectors to C++ output streams, which you can use to dump vector contents for debugging:

```
Vector<float,3> vector2(1,2,3);
std::cout << vector2 << std::endl;
// The output is: [3](1,2,3)
```

Note that the size of the vector is printed first, followed by the vector's contents.

Sometimes it can be useful to work with data that is already stored in memory as though it were stored in a `Vector` object. As long as the data is stored in the usual packed format this is easy to do using the special `VectorProxy` type, which also comes in fixed-size and dynamically-sized variants:

```
float some_data[10] = {0,1,2,3,4,5,6,7,8,9};
VectorProxy<float> proxy1(10, some_data);
VectorProxy<float,10> proxy2(some_data);
```

The constructor arguments are the same as are used in `Vector` to initialize a vector with data from a block of memory, except the data is not copied. You can now treat these proxy objects just like the were regular `Vector`s, except the contents will be stored in the region of memory that you pointed them to. In some situations this can be considerably more efficient than copying the data unnecessarily. (It is of course not possible to resize a `VectorProxy`, since the proxy does not have any control over the memory that it is using.)

4.3.2 Matrices and Matrix Operations

The Vision Workbench `Matrix` class is the matrix counterpart to the `Vector` class, and behaves quite similarly. Once again, there are fixed-dimension and dynamically-sized versions:

```
Matrix<float> matrix1(3,3);
Matrix<float,3,3> matrix2;
```

Note that the arguments to matrix-related functions such as these constructors are given in i, j order, i.e. row followed by column. This is *different* from images, where arguments are given in x, y order, i.e. column followed by row. You may find this confusing at first if you are moving to the Vision Workbench from an environment like Matlab where there is no distinction between images and matrices. However, it is in keeping with the standard index ordering seen in the bulk of the image processing and mathematics literatures, respectively.

You can initialize the matrix with data already stored in memory, as long as the data is stored in a packed row-major format:

```
float some_data[4] = {1,2,3,4};
Matrix<float> matrix1(2,2,some_data);
Matrix<float,2,2> matrix2(some_data);
```

As in the case of `Vector`, the initialization data is *copied* into the matrix in this case, but there is also a proxy form that allows you treat in-memory data like an ordinary matrix:

```
float some_data[4] = {1,2,3,4};
MatrixProxy<float> matrix1(2,2,some_data);
MatrixProxy<float,2,2> matrix2(some_data);
```

The three most common matrix types have been given convenient aliases:

```
typedef Matrix<double,2,2> Matrix2x2;
typedef Matrix<double,3,3> Matrix3x3;
typedef Matrix<double,4,4> Matrix4x4;
```

These types are again the standard types used throughout the Vision Workbench in geometric applications.

You can query a matrix's dimensions using the `rows()` and `cols()` methods, and can index into the matrix to access individual elements. There are two ways to do this:

```
matrix(row,col) = 1;    // "New"-style indexing
matrix[row][col] = 1;   // "Old"-style indexing
```

A dynamically-sized matrix can be resized using the `set_size()` method:

```
matrix.set_size(rows,cols);
```

As in the case of resizing vectors, the default behavior is that any old data is not saved. The `set_size()` method takes an optional third boolean parameter that can be set to `true` to request that it preserve the overlapping entries.

Once you've made one or more matrices you can use a wide range of mathematical operator and functions to manipulate them. The standard C++ operators, elementwise math functions, and a number of other functions similar to those for vectors are supported. A list of the matrix math functions is given in Table 4.6. Notice that some of these functions also operate with vectors: all vector functions that involve matrices are defined in `<vw/Math/Matrix.h>` instead of `<vw/Math/Vector.h>`.

There is a special method, `set_identity()`, that can be used to set a square matrix to the identity matrix of that size.

```
Matrix<float> id(3,3);
id.set_identity();
```

If you want to treat a single row or column of a matrix as though it were a vector, you can do so using the `select_row()` and `select_col()` function:

```
Vector<float> first_row = select_row(matrix,1);
select_col(matrix,2) = Vector3(1,2,3);
```

The second of these examples illustrates that you can use the `select_*` functions to write into matrix rows and columns as well as read them out. Finally, you can treat a block of a matrix as a smaller matrix in its own right using the `submatrix()` function:

```
Matrix<float> block = submatrix(matrix,row,col,rows,cols);
```

Function	Description
`- matrix`	Matrix negation
`matrix + matrix`	Matrix sum
`matrix - matrix`	Matrix difference
`matrix * scalar`	Scalar product
`scalar * matrix`	Scalar product
`matrix / scalar`	Scalar quotient
`matrix += matrix`	Matrix sum assignment
`matrix -= matrix`	Matrix difference assignment
`matrix *= scalar`	Scalar product assignment
`matrix /= scalar`	Scalar quotient assignment
`matrix * matrix`	Matrix product
`matrix * vector`	Matrix-vector product
`vector * matrix`	Vector-matrix product
`elem_sum(matrix,matrix)`	Elementwise matrix sum (same as `+` operator)
`elem_sum(matrix,scalar)`	Elementwise sum of a matrix and a scalar
`elem_sum(scalar,matrix)`	Elementwise sum of a scalar and a matrix
`elem_diff(matrix,matrix)`	Elementwise matrix difference (same as `-` operator)
`elem_diff(matrix,scalar)`	Elementwise difference of a matrix and a scalar
`elem_diff(scalar,matrix)`	Elementwise difference of a scalar and a matrix
`elem_prod(matrix,matrix)`	Elementwise product of two matrices
`elem_prod(matrix,scalar)`	Elementwise matrix product (same as `*` operator)
`elem_prod(scalar,matrix)`	Elementwise matrix product (same as `*` operator)
`elem_quot(matrix,matrix)`	Elementwise quotient of two matrixs
`elem_quot(matrix,scalar)`	Elementwise quotient (same as `/` operator)
`elem_quot(scalar,matrix)`	Elementwise quotient of a scalar and a matrix
`norm_1(matrix)`	Matrix 1-norm
`norm_2(matrix)`	Matrix 2-norm
`norm_frobenius(matrix)`	Matrix Frobenius norm
`sum(matrix)`	Sum of elements, i.e. $\sum v_i$
`prod(matrix)`	Product of elements, i.e. $\prod v_i$
`trace(matrix)`	Matrix trace, i.e. $\sum M_{ii}$
`transpose(matrix)`	Matrix transpose, i.e. M^T
`inverse(matrix)`	Matrix inverse, i.e. M^{-1}
`null(matrix)`	Matrix nullspace, i.e. a $\Re x$ where $Mx = 0$

Table 4.6: The matrix math functions defined in `<vw/Math/Matrix.h>`.

You can also use this function to write into a region of a matrix, much as in the previous example using `select_col()`.

Like `Vector`, `Matrix` is a C++ STL-compatible container class. The `Matrix<...>::iterator` iterates over the elements of a matrix in the same order that the `ImageView`'s iterator does: across each row, moving down the matrix from each row to the next. This is again a good method for loading or extracting matrix data from other containers. To extract the matrix data to a stream for debugging output you can use the `<<` stream output operator:

```
double data[4] = {1,2,3,4};
Matrix2x2 matrix(data);
std::cout << matrix << std::endl;
// The output is: [2,2]((1,2)(3,4))
```

Again, the output includes the matrix dimensions (rows followed by cols), followed by the matrix data.

4.4 Transforming or Warping Images

We return now to our discussion of image processing by introducing a new concept: image transformation. Most of the image processing operations we have dealt with so far (with the exception of the simple transforms in Section 3.3.1) have operated on pixel values. Image transformation, or warping, is a common image processing operation that operates instead on a pixels *location*.

4.4.1 Transform Basics

Let's start with a basic example of image transformation. First, include the `<vw/Image/Transform.h>` header file. Now, imagine you would like to translate all of the pixels in the image 100 pixel positions to to right. This operation does nothing to the pixel values except to relocate them in the image. The Vision Workbench provides a convenient method for performing this operation.

```
double u_translation = 100;
double v_translation = 0;
result_image = translate(input_image, u_translation, v_translation);
```

This simple example already raises some interesting questions. How big is the output image? What happens to the pixels that are translated off the right edge of the image? What value is used to fill in pixels where the original image has no data?

The answer to the first question is straight-forward. By default, the transformed image will have the same dimensions as the input image. However, you can easily override this behavior by selecting a different region from the output image using the `crop()` function. For example, you could grow the right side of the output image to include the shifted pixels.

```
result_image = crop(translate(input_image, u_translation, v_translation),
                    0, 0, input_image.cols() + x_translation, input_image.rows());
```

If `input_image` was 320x240, `result_image` will be 420x240 pixels and it will have a 100x240 black band on its left side.

This is a good time to stop to consider what is really happening here, because the ability to arbitrarily crop the output of a transformed image is extremely useful. Under the hood, our call to translate is returning an object that behaves like an image (so it can be cropped), but it is actually presenting an image-like interface to some processed, edge-extended data. Thus, you can use `crop()` to select a region of pixels anywhere in the pixel space that contains the resulting image. It is not until you *assign* the cropped image to `result_image` that this data is once again rasterized and stored as a contiguous block in memory.

Note that the Vision Workbench adopts a consistent coordinate system when working with pixels in the transformed image space. The origin is at the upper left hand corner of the original

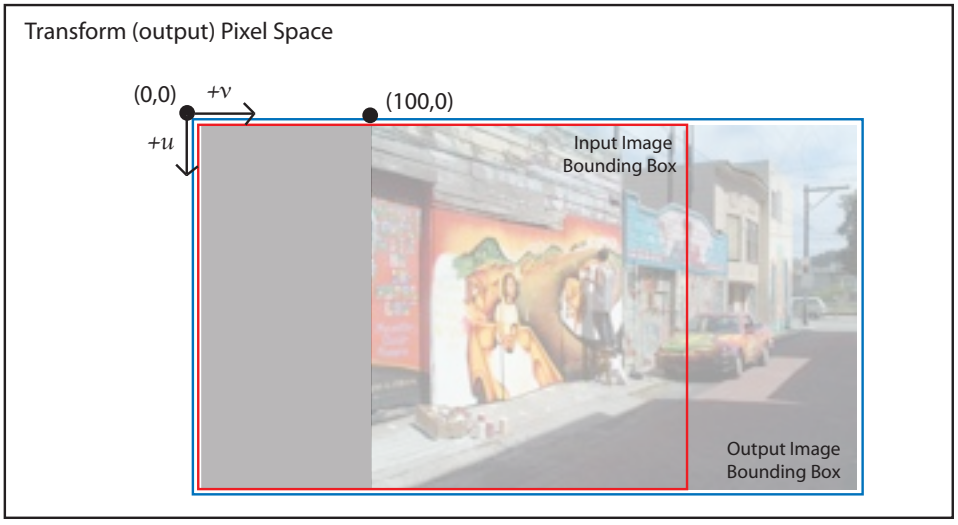


Figure 4.1: Using the `crop()` function, you can select any region of the transformed (in this case, translated) image that you need.

image, with the u coordinate increases as you move down the rows of the image. Figure 4.1 shows this coordinate system and the input and output bounding boxes in the case of the the cropped, translated image example we have been working with.

Using this intuition, we can now answer the second question posed above. When the pixels are translated off of the right edge of the image, they disappear unless they are explicitly selected using `crop()`. The only other reasonable behavior might have been to have the pixels wrap around and enter on the left side of the image. This is not supported using the Vision Workbench `translate()` function, however, as you will learn in the next section, such transformations are still possible using the general transform framework.

Finally, we arrive at the third question: what pixel value is used to fill area where the original image has no data? To answer this, think back to the discussion of edge extension modes for the filter routines in section 4.1.2. Edge extension behavior in the transform routines of the Vision Workbench are specified in an identical fashion.

```
result_image = translate(input_image,
                         u_translation, v_translation,
                         ConstantEdgeExtension());
```

In this example, the left 100x240 block of `result_image` will contain the "smeared out" pixels from the left side of the input image. Of course, this is probably not what you wanted, so the default behavior edge extension behavior for `translate()` is set to `ZeroEdgeExtension()`.

One final point before we move on to talking about image transformations more generally. Consider this block of code:

```
double u_transformations = 100.5;
double v_transformation = 30.7;
result_image = translate(input_image, u_translation, v_translation,
                         ConstantEdgeExtension(), BicubicInterpolation());
```

Type	Description
`NearestPixelInterpolation`	Use the nearest integer valued pixel location
`Bilinear Interpolation`	Linearly interpolation based on the four nearest pixel values
`Bicubic Interpolation`	Quadritic interpolation based on the nine nearest pixel values

Table 4.7: The Vision Workbench Interpolation Modes.

Here, the image is translated by a non-integer number of pixels. This is a totally reasonable thing to do, but it raises the question of how one accesses a non-integer pixel location in the source image. The answer: interpolation. As with edge extension, you can specify the interpolation mode by passing in a dummy argument to the `translate()` function. Table 4.7 shows the built-in interpolation types.

4.4.2 Creating a New Transform

Having now addressed some of the fundamental issues that arise when transforming images, we now turn our discussion to how one might formulate and implement a new image transformation algorithm.

In the most general sense, a transform is computed by performing the following two steps for every pixel in the output image.

- Given the coordinates X_{out} of a pixel in the output image, apply a transformation that yields the coordinates X_{in} of a source pixel in the input image.
- Use some edge extension and interpolation scheme to determine the pixel value of the input image at X_{in} (it may fall in between integer pixels coordinates or outside of the input image entirely) and set the value of the output image at X_{out} to this value.

When formulating a new image transformation algorithm, the first step where all of the interesting work happens. The code for interpolation and edge extension is important, but usually incidental to the transformation under development. Ideally, one would focus exclusively on writing code to perform the geometric calculations in step one. To help us with this task, we will introduce a new programming idiom that appears commonly in the Vision Workbench: the *functor*.

Technically, a functor is a C++ class that has implemented the `operator()` method. Once created, such a class can be called and passed around in place of a normal C++ function. It behaves identically except that, as a C++ object, the functor can maintain its own state (possibly initialized when the functor is constructed). In the Vision Workbench, we use this definition more loosely to mean any small function object that adheres to a small, pre-determined interface. But, rather than linger over semantic details, let's jump straight to an example so that you can see what we mean.

Let's look at the definition for the functor that describes image translation, shown in Listing 3. You'll notice that this class has defined three methods: a constructor and two methods called `forward()` and `reverse()`. The class also inherits from `TransformBase<>`, but that's not something to dwell on here. For now just be aware that `TransformBase<>` provides default implementations that throw `vw::UnimplErr()` exceptions in case the subclass does not implement both methods.

The constructor is used to initialize the state of this functor; in this case, an offset in x and y. The `reverse()` method is the most important. It performs step one in our list at the beginning of this section. Pretty simple, right? Although the transformation in this example is nothing special,

```
1  // __BEGIN_LICENSE__
2  // Copyright (C) 2006, 2007 United States Government as represented by
3  // the Administrator of the National Aeronautics and Space Administration.
4  // All Rights Reserved.
5  // __END_LICENSE__
6
7
8  class TranslateTransform : public TransformBase<TranslateTransform> {
9    double m_xtrans, m_ytrans;
10 public:
11   TranslateTransform(double x_translation, double y_translation) :
12     m_xtrans( x_translation ) , m_ytrans( y_translation ) {}
13
14   // Given a pixel coordinate in the ouput image, return
15   // a pixel coordinate in the input image.
16   inline Vector2 reverse(const Vector2 &p) const {
17     return Vector2( p(0) - m_xtrans, p(1) - m_ytrans );
18   }
19
20   // Given a pixel coordinate in the input image, return
21   // a pixel coordinate in the output image.
22   inline Vector2 forward(const Vector2 &p) const {
23     return Vector2( p(0) + m_xtrans, p(1) + m_ytrans );
24   }
25 };
```

Listing 3: [`transform-functor.h`] An example transform functor that performs image translation.

the `reverse()` method could potentially be long and complicated. So long as it returns a pixel position in the input image in the end, we're happy.

The `forward()` method performs the inverse operation of the `reverse()` method. This method is not always necessary. We'll discuss it more in Section 4.4.3.

The beauty of the `TranslateTransform` class, or any other class that defines a set of `forward()` and `reverse()` methods is that it can be passed as an argument to the `transform()` function.

```
result_image = transform(input_image,
                         TranslateTransform(u_translation, v_translation));
```

This block of code performs the very same transformations as our call to `translate()` in the previous section. (In fact, `translate()` is just a thin wrapper around `transform()` provided for convenience.) As with the previous example, the edge extend and interpolation modes can be supplied as dummy arguments to `transform()`.

As you can probably now see, the possibilities are endless! For example, we could also have used `HomographyTransform` (another built-in transform functor) to describe the same translation. The linear homogeneous transform that encodes a 100 pixel shift to the right is:

$$H = \begin{pmatrix} 1 & 0 & 100 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.1}$$

Dropping the `HomographyTransform` into `transform()` yields the same `result_image` once again.

```
vw::Matrix<double> H = ... // defined as above
result_image = transform(input_image, HomographyTransform(H));
```

Any transform functor that adheres to this simple interface, including one of your own devising, can be passed into `transform`. To summarize, you can create and use your own transformation functor `foo` by following these steps.

- Inherit from `public TransformBase<foo>`
- Define a constructor that stores any state information that you need
- Define a `reverse()` method
- Define a `forward()` method (optional, see Section 4.4.3)

Of course, some of the most common transform functors are provided for you as part of the Vision Workbench. Refer to Table 4.8 for a list of built-in classes available in `<vw/Image/Transform.h>`

Type	Function	Description
`ResampleTransform`	`resample()`	Scale an image, resizing the output image as needed
`TranslateTransform`	`translate()`	Translate an image
`HomographyTransform`		Apply a linear homogeneous transformation (3x3 Matrix)
`PointLookupTransform`		Apply a transformation based on a lookup table image
`PointOffsetTransform`		Apply a transformation based on an offset table image

Table 4.8: Built-in transform functors and (if available) their function interface.

4.4.3 Advanced Techniques

To wrap up our discussion of the transform methods, here are some advanced techniques that you may find useful when working with image transformations.

It is not uncommon to ask, for a given transform functor, what bounding box in the transformed image space contains the complete set of transformed pixels. The `compute_transformed_bbox()` routine answers this question by performing the `forward()` transformation for each pixel location in the input image and growing a bounding box to contain all of the forward transformed pixels. This bounding box can be passed directly as the second argument to `crop()`.

```
// Output image is cropped to contain all transformed pixels
BBox2f result_bbox = compute_transformed_bbox(input_image,
                                              MyTransformFunctor());
result_image = crop(transform(input_image, MyTransformFunctor()),
                    result_bbox);
```

For performance limited applications, you may find `compute_transformed_bbox_fast()` more appropriate. It computes the bounding box by applying `forward()` to the perimeter pixels of the input image only. This should produce identical results to the "slow" version so long as the

perimeter pixels of the input image form the perimeter of the output image under the transformation in question.

Finally, we would like to point out the existence of the `RadialTransformAdaptor` class. This class is useful when it is easier or more natural to describe a transformation in terms of polar coordinates $[r, \theta]$ instead of the usual cartesian coordinates $[u, v]$.

To use `RadialTransformAdaptor`, you write your transform functor as usual, but you interpret the components in the input and output `Vector2` to be $[r, \theta]$, in that order. Assuming you have created a class `MyRadialTransform` in this manner, you can apply it as follows:

```
result_image = transform(input_image,
                         RadialTransformAdaptor(MyRadialTransform(),
                                                input_image));
```

The `RadialTransformAdaptor` creates a polar coordinate system wherein the center of the image is the origin, and a value of $r = 1.0$ is equal to the distance from the center to the left edge of the image.

4.5 Pixel Mask

At this point is best to introduce a new pixel type, `PixelMask`. It was convientantly hidden from the previous chapter as it is not exactly just another pixel type. It does infact though still have that same blue collar past of the others from Table 3.2 and that means all operations the previous Pixel types could do, `PixelMask` can do as well. `PixelMask` just happens to append an additional boolean channel value to what ever pixel type it encapsulates.

Type	Description	Channels
`PixelMask<PixelT<T> >`	Additional valid bit	One plus whatever `PixelT` is

Table 4.9: An addendum to standard Vision Workbench pixel types. `PixelT` should generally be one of the types from Table 3.2 and `T` should generally be from Table 3.1.

This additional boolean channel is a pixel's validity. During FileIO the boolean channel can correctly be interpreted as determining if a pixel is transparent or not. This may seem as just a redundant binary version of an alpha channel and it can correctly be interpreted as this. Yet it has one ability that an alpha channel does not. The state of a `PixelMask` being invalid is quite like the 'T' virus from *Resident Evil* fame. It spreads to all other `PixelMask`s it is operated with. Meaning that given an image that has bad pixels in it, they can be labelled as invalid and will remain so through all operations.

A `PixelMask` passes all of it access functions on to its inner pixel type. So that a `pixel[0]` will still return a red channel from a pixel of type `PixelMask<PixelRGB<float> >`. Yet remember that it adds and an additional channel. So a `PixelMask`'s validity can be access with `pixel[3]` or instead with a `pixel.valid()`.

Testing for a pixel's validity via a `pixel.valid()` only works if a pixel is a `PixelMask`. But what if a function is written generally for all pixels and it must test for a valid pixel like `blob_index`? In that case, Vision Workbench provides a series of commands for working with a pixel validity despite the type. They are listed in Table 4.10.

Lastly, Vision Workbench provides a number of ImageView operators working pixel validity. They are listed in Table 4.11.

Function	Description
`is_transparent(PixelT)`	Tests if a pixel is transparent. In other words, not valid.
`is_valid(PixelT)`	Tests as expected.
`validate(PixelT)`	Forces a pixel valid *iff* `PixelMask`. Otherwise no-op.
`invalidate(PixelT)`	Forces a pixel invalid *iff* `PixelMask`
`toggle(PixelT)`	Changes the state of validity *iff* `PixelMask`
`remove_mask(PixelT)`	Strips off the `PixelMask` wrapper and returns the inner pixel type.

Table 4.10: Pixel type safe functions for working with a `PixelMask`'s valid bit.

Function	Description
`create_mask(im,s)`	Given a non `PixelMask` view `im`, returns an `ImageView` of type `PixelMask`. Pixels of valud `s` are listed as invalid.
`create_mask(im,lo,hi)`	As previously, except `lo` and `hi` describe the valid pixel range. Pixels outside of that range are listed as invalid.
`apply_mask(im,s)`	This removes the `PixelMask` type, and replaces invalid pixels with a value `s`.
`copy_mask(im,mask)`	Copies an image `mask`'s transparency on to an image `im` of non-`PixelMask` type.
`mask_to_alpha(im)`	Convert a `PixelMask` view to a view with a alpha channel.
`edge_mask(im)`	Searches for the edges of an image and marks the first valid pixel, invalid.
`invert_mask(im)`	Given an image of `PixelMask`s , inverts every pixels validity.
`validate_mask(im)`	Forces all pixels to be valid.
`invalidate_mask(im)`	Forces all pixels to be invalid.
`union_mask(im,mask)`	Unions the `PixelMask` of `mask` with `im`, and keeping the image data in `im`.
`intersect_mask(im,mask)`	Intersects the `PixelMask` of `mask` with `im`, and keeping the image data in `im`.

Table 4.11: Image operations for working with `PixelMasks`

Chapter 5

Vision Workbench Type System

The Vision Workbench is an example of what is often called a "multi-paradigm" C++ library. That is, different components of the library adopt different C++ programming models, such as the generic programming model or the object-oriented programming model, often in combination. At the core, however, is a set of data types and related tools that fall largely within the template-based generic programming paradigm. The purpose of this chapter is to describe this core type system in some detail. If your intention is simply to *use* the Vision Workbench for image processing tasks you can probably afford to skim or even skip this material. The primary intended audience is programmers who wish to extend the Vision Workbench's core capabilities in one way or another.

Data types in the Vision Workbench can be broadly divided into three categories: the fundamental data types, including the simple numeric types; the compound types, such as RGB pixel types and vectors; and the container types, such as images. We shall discuss each of these in turn. (Other C++ types, such as functors, do make an appearance in the Vision Workbench, but these are not *data* types *per se*.)

5.1 The Scalar Types

The most fundamental data types of all are the built-in C++ integral and floating-point numeric types. In scientific programming contexts like image processing and machine vision it is generally important to be able to specify the exact nature of the numeric data types that you are working with. Unfortunately the C++ language makes few promises about the sizes of data types such as `int` or even `char`. To work around this limitation, the Vision Workbench provides a number of portable `typedef`s that you are encouraged to use instead. These are listed in Table 5.1. (In fact most of these data types are simple wrappers around similar types provided by the Boost `cstdint` library.)

The Vision Workbench uses standard C++ complex numbers, as defined in the standard header file `<complex>`. The `std::complex<>` class takes a single template parameter, the underlying numeric type to be used for the real and imaginary components, which should usually be one of the scalar types listed in Table 5.1. In particular, it is almost always best to use one of the floating-point types for complex numbers. For example, `std::complex<float64>` is a good choice for use in frequency-domain image processing (discussed in Section 14.7).

There is a special type trait template, `IsScalar<>`, that you can use in template code to determine whether or not a type is a simple numeric type like we have described in this section. It inherits from either `boost::true_type` or `boost::false_type` accordingly. Its primary use is to prevent template functions such as scalar multiplication from being too general:

Type	Description	Notes
`int8`	Signed 8-bit integer	
`uint8`	Unsigned 8-bit integer	Most common for low-dynamic-range imaging
`int16`	Signed 16-bit integer	
`uint16`	Unsigned 16-bit integer	
`int32`	Signed 32-bit integer	
`uint32`	Unsigned 32-bit integer	
`int64`	Signed 64-bit integer	
`uint64`	Unsigned 64-bit integer	
`float32`	32-bit floating point	Most common for high-dynamic-range imaging
`float64`	64-bit floating point	

Table 5.1: The core Vision Workbench scalar `typedefs`, defined in `<vw/FundamentalTypes.h>`.

Class	Description
`SumType<T1,T2>`	Result type of a sum operation
`DifferenceType<T1,T2>`	Result type of a difference operation
`ProductType<T1,T2>`	Result type of a product operation
`QuotientType<T1,T2>`	Result type of a quotient operation

Table 5.2: The Vision Workbench type deduction classes, defined in `<vw/TypeDeduction.h>`.

```
template <class ScalarT>
typename boost::enable_if< IsScalar<ScalarT>, MyClass >::type
operator*( MyClass const& m, ScalarT s ) {
  /* compute result */
}
```

In this example we use the Boost `enable_if` library to restrict the definition of the `*` operator to cases where the second argument really is a scalar. Without this restriction this would have been an overly-general function definition and would likely have caused problems if we had attempted to defined any other product for the `MyClass` class later on. If you decide to extend the Vision Workbench to support additional scalar types, such as bigints, you should specialize `IsScalar<>` accordingly to ensure proper behavior.

5.2 Type Deduction

The C++ language has many intricate rules for type promotion and deduction in complex mathematical expressions. Unfortunately it provides no built-in mechanism to extend this automatic type deduction system or query its behavior. Consider adding two images with compatible but different pixel types: what should the resulting pixel type be? The Vision Workbench provides a standard set of type deduction traits classes, defined in `<vw/TypeDeduction.h>` and listed in Table 5.2, that allow you to both query and specialize the type deduction behavior of Vision Workbench types. Like all Vision Workbench type computation classes, they "return" their result types in a member type named `type`.

As a trivial example, imagine writing a template function that simply computes the sum of its two arguments. What should its return type be? We can use `SumType<>` to compute it:

Type	Description	Channels
`PixelGray<T>`	Grayscale	Grayscale value (`v`)
`PixelGrayA<T>`	Grayscale w/ alpha	Grayscale value (`v`), alpha (`a`)
`PixelRGB<T>`	RGB	Red (`r`), green (`g`), blue (`b`)
`PixelRGBA<T>`	RGB w/ alpha	Red (`r`), green (`g`), blue (`b`), alpha (`a`)
`PixelHSV<T>`	HSV	Hue (`h`), saturation (`s`), value (`v`)

Table 5.3: The Vision Workbench color-space pixel types, defined in `<vw/PixelTypes.h>`.

Syntax	Description
`PixelChannelType<PixT>::type`	The pixel type's underlying channel type
`PixelNumChannels<PixT>::value`	The number of channels in the pixel type
`PixelChannelCast<PixT,ChT>::type`	The same pixel type with a new channel type
`PixelIsCompound<PixT>`	Is the pixel type a compound type?
`PixelMakeReal<Pixt>::type`	Converts to the corresponding real channel type
`PixelMakeComplex<Pixt>::type`	Converts to the corresponding complex channel type

Table 5.4: The pixel traits and pixel type computation classes.

```
template <class T1, class T2>
inline typename SumType<T1,T2>::type sum( T1 a, T2 b ) {
  return a + b;
}
```

Remember that the C++ language does not allow this to be fully automatic. If you define a new type with an unusual addition operator, you will need to manually specialize `SumType<>` at the same time. However, the most common default behaviors are provided. For example, any built-in type is assumed to be promoted to any user-defined type, and any user-defined type operating with itself is assumed to return itself. These type deduction classes do also replicate the standard C++ promotion behavior when used with the built-in numeric types.

5.3 The Pixel Types

It is possible to use any of the fundamental scalar types described in the previous section as an `ImageView`'s pixel type. However in most circumstances a *compound* pixel type, consisting of one or more channels with associated semantics, is more appropriate. The Vision Workbench provides several compound pixel types corresponding to the most common color spaces used in image processing. These are listed in Table 5.3. Each is a template class taking one template parameter, the scalar type used to store the channels. For example, the native pixel type of a standard JPEG image is represented by `PixelRGB<uint8>`.

Several type trait classes are provided for use in writing generic pixel manipulation code and are listed in Table 5.4. The first section of the table lists the classes that you must specialize when you write a new compound pixel type. The second section of Table 5.4 lists convenience types that are defined in terms of the other, specialized types. For example, here is how the types are specialized for the RGB pixel type:

```
template <class ChannelT>
```

```
struct PixelChannelType<PixelRGB<ChannelT> > {
  typedef ChannelT type;
};

template <class ChannelT>
struct PixelNumChannels<PixelRGB<ChannelT> > {
  static const unsigned value = 3;
};

template <class OldChT, class NewChT>
struct PixelChannelCast<PixelRGB<OldChT>, NewChT> {
  typedef PixelRGB<NewChT> type;
};
```

When you define a new pixel type, you will usually want to define a provide a similar set of template specializations. To simplify the process, a macro is provided that you can use to automatically specialize the templates in the usual manner:

```
VW_DECLARE_PIXEL_TYPE( PixelRGB, 3 );
```

This macro expands to the same set of template specializations shown above, describing a pixel type named `PixelRGB` with three channels.

Note that it is *not* necessary to declare a type using this macro, or even to provide specializations for the traits templates described above, just to use that type as the pixel type for an image. These specializations are only necessary if you want to declare a type with multi-channel semantics. For any other type, the default Vision Workbench behavior is to treat the type as a single-channel pixel type whose channel type is equal to the type itself.

Several convenience functions are also provided to simplify working with pixels in generic template functions. The first is the `pixel_channel_cast<>()` function, which casts a pixel to a pixel of the corresponding pixel type but with the specified channel type. The syntax mirrors the built-in C++ casting functions, except the template parameter is the new channel type instead of the new type as a whole. In the following example we explicitly down-cast the channel type of a pixel from `float64` to `float32` in order to pass it to a function that happens to take a `PixelRGB<float32>` argument:

```
PixelRGB<float64> pixel;
some_function( pixel_channel_cast<float32>( pixel ) );
```

Note that this is not needed in the more common case that the function that you wish to call is itself generic and can accept any channel type.

Sometimes it is desirable to apply a function to each channel of a pixel, or to corresponding channels from two pixels. Re-scaling a pixel or adding two pixels on a per-channel basis can be cast into this form, for example. You can use the generic function `apply_per_pixel_channel()` to do this. Here is a trivial example that demonstrates re-scaling:

```
float32 triple(float32 v) { return 3*v; }
// Later, in some other function...
PixelRGB<float> pixel(.1,.2,.3);
PixelRGB<float> result = apply_per_pixel_channel(&triple,pixel);
```

In this case it would have been simpler to multiply the pixel by 3 directly, but the point is that we could have performed any arbitrarily complex operation on each channel instead. The binary form simply takes an extra pixel argument:

```
float32 sum(float32 a, float32 b) { return a+b; }
// Later...
PixelRGB<float> pixel1(.1,.2,.3), pixel2(.2,.1,.4);
PixelRGB<float> result = apply_per_pixel_channel(&sum,pixel1,pixel2);
```

Chapter 6

Core Module

The Core Module contains fundamental tools for building good software infrastructure.

6.1 Vision Workbench Exceptions

The Vision Workbench is intended in part to be used in flight systems, experimental multiprocessor systems, or other environments where exceptions may not be fully supported. As a result, the use of exceptions within the Vision Workbench is tightly controlled. In particular, the exception usage rules were designed to minimize the imact on platforms that do not support exceptions at all. There is a standard Vision Workbench "exception" class hierarchy which is used to describe errors and can be used even on platforms that do not support the C++ exception system.

The `vw::Exception` class serves as a base class for all VWB error types. It is designed to make it easy to throw exceptions with meaningful error messages. For example, this code:

```
vw_throw( vw::Exception() << "Unable to open file \"" << filename << "\"!" );
```

would generate a message like this:

```
terminate called after throwing an instance of 'vw::Exception'
     what():  Unable to open file "somefile.foo"!
```

Note that in the example the exception was thrown by calling the `vw_throw()` function rather than by using the C++ throw statement. On platforms that do support C++ exceptions the default behavior for `vw_throw()` is to throw the exception in the usual way. However, the user can provide their own error-handling mechanism if they choose. For example, the default behavior when exceptions are disabled is to print the error text to `stderr` and call `abort()`.

There are a number of standard exception types that derive from `vw::Exception`. These are shown in Table 6.1. In the above example, the exception should probably have been of type `vw::IOErr`.

Also, two macros, `VW_ASSERT(condition,exception)` and `VW_DEBUG_ASSERT(condition,exception)`, are provided, with the usual assertion semantics. The only difference is that the debug assertions will be disabled for increased performance in release builds when `VW_DEBUG_LEVEL` is defined to zero (which happens by default when `NDEBUG` is defined).

Exceptions are enabled or disabled based on the value of the `VW_ENABLE_EXCEPTIONS` macro defined in `vw/config.h`. This value can be set by passing the command line options `--enable-exceptions`

Function	Description
ArgumentErr	Invalid function argument exception
LogicErr	Incorrect program logic exception
InputErr	Invalid program input exception
IOErr	IO (usually disk IO) failure exception
MathErr	Arithmetic failure exception
NullPtrErr	Unexpected NULL pointer exception
TypeErr	Invalid type exception
NotFoundErr	Not found exception
NoImplErr	Unimplemented functionality exception
Aborted	Operation aborted partway through

Table 6.1: Vision Workbench exception types that derive from vw::Exception. All behave like C++ output stream classes, so you can associate an error message with the exception using the stream operator.

(the default) or `--disable-exceptions` to the configure script prior to building the Vision Workbench. This option also sets an automake variable called `ENABLE_EXCEPTIONS` which may be used by the build system to conditionally compile entire source files.

In either case the default behavior of `vw_throw()` may be overridden by passing a pointer to a user-defined object derived from `ExceptionHandler` to `vw::set_exception_handler()`. The user specifies the error-handling behavior by overriding the abstract method `handle()`.

6.2 The System Cache

The Vision Workbench provides a thread-safe system for caching regeneratable data. When the cache is full, the least recently used object is *invalidated* to make room for new objects. Invalidated objects have had the resource associated with them (e.g. memory or other resources) deallocated or freed, however, the object can be *regenerated* (that is, the resource is regenerated automatically by the cache) when the object is next accessed.

The `vw::Cache` object defined in `src/vw/Core/Cache.h` can be used to store any resource. For example, one common usage would be to create a cache of image blocks in memory. In this case, the cache enforces a maximum memory footprint for image block storage, and it regenerates the blocks (e.g. reloads them from a file on disk) when necessary if a block is accessed.

One can also cache more abstract resource types, such as std::ofstream objects pointing at open files on disk. The following section describes this use case in detail.

6.2.1 Example: Caching std::ofstream

Consider a situation wherein your system needs to open and read from tens of thousands of tiny files on disk. There is a high degree of locality of access, meaning that once you start reading from a file, you are likely to read from it again in the near future, but that likelihood diminishes as time goes on. It is impractical to simply open all of the files at once (this would eat up memory in your program, plus there is a hard limit in most operating systems on the number of open files you can have at any one time). However, it could potentially be very slow to close and re-open the file each time you attempt to read from it because of the time it takes to parse the file header or seek to the

correct location.

This situation calls for a cache.

We begin by specifying a *generator* class whose sole purpose is to regenerate the resource if it is ever necessary to reopen it. In this case, our class contains a `generate()` method that return a shared pointer to a newly open ifstream.

```
class FileHandleGenerator {
  std::string m_filename;
public:
  typedef std::ifstream value_type;

  FileHandleGenerator( std::string filename ) : m_filename( filename ) {}

  // The size is useful when managing items that have
  // a known size (e.g. allocated blocks in memory).
  // When caching abstract items, like open files, the
  // size does not matter, only the total number of open
  // files, hence the size in this case is 1.
  size_t size() const { return 1; }

  // Generate is called whenever there is a cache miss.  In
  // this example, we reopen the file in append mode.
  boost::shared_ptr<value_type> generate() const {
    return boost::shared_ptr<value_type> ( new std::ifstream(m_filename, ios::app) );
  }
};
```

Next, we create a `vw::Cache` object for storing instances of our `FileHandleGenerator` class. The cache itself can be declared without knowing the type of the generator(s) that will be inserted into it. The `vw::Cache` constructor takes only one argument specifying the size of the cache. This value will be used in conjunction with the `size()` methods its cache generators to determine when the cache is full. If the sum of the `size()` values from *valid* generators exceeds the max cache size, the least recently used cache generator is invalidated to make room for a new call to `generate()`.

```
// This cache maintains up to 200 open files at a time.
static vw::Cache filehandle_cache( 200 );

// Insert a ifstream generator into the cache and store a handle
// so that we can access it later.
Cache::Handle<FileHandleGenerator> file_ptr;
file_ptr = filehandle_cache().insert(GdalDatasetGenerator(filename));

// ... time passes ...

// We access the file generator like a pointer.  The cache
// will re-open the file if necessary.
char[2048] line_from_file;
*file_ptr.get_line(line_from_file, 2048);
```

Note in this example that our call to `insert()` returns a `Cache::Handle<>` object that "points at" the cache's version of the ifstream object. The cache handle behaves like a pointer, and we can use it just as we would a normal C++ pointer to a `std::ifstrem`. However, if the cache closes the file at any point to make room for other files, this cache handle ensures that the file is regenerated (re-opened) the next time we try to access it.

6.2.2 Performance Considerations and Debugging

Use of a cache can greatly increase the efficiency of a program by storing information that would be expensive to regenerate for as long as possible. However, depending on the size of the cache and the pattern you are using to access it, one can inadvertently end up in a situation where performance may suffer considerably.

For example, consider a scenario where you have a `DiskImageView<>` that points at a very large (22,000x22,000 pixels) 8-bit RGB image that is stored in a file format with a block size of 2048x2048 pixels. The `DiskImageView<>` caches these block as the image is accessed, and each block takes 96-MB of memory.

If we access the `DiskImageView<>` one scan line at a time (during a rasterization operation, for example), the cache will need to store at least $22,000/2048 = 10$ blocks in memory for efficient access (i.e our cache size must be ¿= 960-MB). If the cache is too small, the left-most image blocks in the row will be invalidated by the cache to make room for the right-most blocks, and vice versa as we traverse *each scanline* of the `DiskImageView<>`. Given that it takes a few seconds to regenerate each block, this would be hugely inefficient.

If your code is running much, much slower than you expect, you may have a similar cache miss problem. You can debug this by observing the "cache" *log namespace* using the system log (see Section 6.3 for details). The cache subsystem logs debugging information whenever cache handles are invalidated or regenerated. Use these messages to get a gross idea of how the cache is performing.

6.3 The System Log

As the Vision Workbench has become more parallelized, and as new subsystems (e.g. caching, fileio, threadpool management, etc.) have been added, it has become increasingly challenging to monitor and debug the code base. The Vision Workbench log class was designed to address the evolving needs of VW developers and users alike. The following design guidelines summarize the features of the system log facility provided in `vw/Core/Log.h`.

- **Thread Safety**: Log messages are buffered on a per-thread basis so that messages form different threads are correctly interleaved one line at a time in the log output.

- **Log Granularity**: Users will want to monitor different subsystems at different times. Each log message includes an associated *log namespace* and *log level*. Used in conjunction with the the `LogRuleSet` class, the log namespace and level allow the user to monitor only the log messages that concern them

- **Multiple Log Streams**: Log messages can be directed either to the user's terminal, one or more files on disk, or both at the same time. Each log stream has its own `LogRuleSet`, so it is possible to tailor the log output that appears in each file.

- **Runtime Log Adjustment**: When something goes wrong in the middle of a long job, it is undesirable to stop the program to adjust the log settings. The VW log framework will reload the log settings from the `~/.vwrc` file (if it exists) every 5 seconds, so it is possible to adjust log settings during program execution. See Section 6.3.2 below.

6.3.1 Writing Log Messages

Logging in the Vision Workbench is simple. Just call the `vw_out(log_level, log_namespace)` command, which will return a `basic_ostream` object that you can stream into using the standard C++ `<<` operator. For example,

```
// Record the default number of Vision Workbench threads to the log
int num_threads = vw::Thread::default_num_threads();
vw_out(vw::InfoMessage, "thread") << "The default number of threads is " << num_threads <<
```

This would generate a log message that looks something like this.

```
2007-Dec-28 14:31:52 {0} [ thread ] : The default number of threads is 8.
```

The log message includes an *infostamp* consisting of the timestamp, the unique id of the thread that generated the log message, and the `log_namespace` specified as an argument to `vw_out`.

```
<date> <time> {<thread_id>} [ <log_namespace> ] : <log_message>
```

Note that, for aesthetic reasons, log messages that go to the console only print the `log_message` after the colon; the *infostamp* is omitted. A new *infostamp* is prepended to the log stream each time `vw_out()` is called.

Take note of the newline character at the end of stream to `vw_out()` in the example above. The logging framework will cache the log message until it sees the newline *at the end* of a call to `operator<<`, at which point the log stream is flushed, and newline is written, starting a new line in the log file. Therefore, it is highly recommended that you end every log message with a newline character (or `std::endl`).

6.3.2 The Log Configuration File

The *log configuration file* can be used to change the Vision Workbench logging behavior in real-time, even while your program is running.

Every five seconds, or when a log message is generated using `vw_out()` (whichever is *longer*, the system log checks to see if the logconf file has been modified. If so, it erases all log streams and log rules, and reloads these settings from the file. If you modify the file while your program is running, you will see your changes take affect anywhere from 0-5 seconds from the time that you save your changes.

Using the file, you have full control over the system log: you can create as many log streams to files as you like, and you can adjust the log rules for both the console log and the log file streams. Note that syntax errors and malformed statements are silently ignored by the log configuration file parser, so check your file carefully for errors prior to saving.

An example log configuration file appears in Listing 4.

6.3.3 System Log API

If you would rather not use the log configuration file, you can adjust the system log settings directly by accessing the singleton instance of the vw::Log class using `vw::system_log()`. Once you have done this, you can explicitly add (or clear) new log streams (or `LogInstance` objects, in the parlance of the API), and you can adjust log rule sets (using the `LogRuleSet`) class.

Consult the API documentation for `vw/Core/Log.h` for more information.

```
# This is an example VW log configuration file. Save
# this file to ~/.vwrc to adjust the VW log
# settings, even if the program is already running.
#
# The following integers are associated with the
# log levels throughout the Vision Workbench.  Use
# these in the log rules below.
#
#    ErrorMessage = 0
#    WarningMessage = 10
#    InfoMessage = 20
#    DebugMessage = 30
#    VerboseDebugMessage = 40
#
# You can create a new log file or adjust the settings
# for the console log:
#
#   logfile <filename>
#       - or -
#   logfile console
#
# Once you have created a logfile (or selected the
# console), you can add log rules using the following
# syntax. (Note that you can use wildcard characters
# '*' to catch all log_levels for a given log_namespace,
# or vice versa.)
#
# <log_level> <log_namespace>
#
# Example: For the console log, turn on InfoMessage
# logging for the thread sub-system and log every
# message from the cache sub-system.

[general]
default_num_threads = 8
system_cache_size = 2048

[logfile console]
20 = thread
* = cache
20 = stereo
VerboseDebugMessage = *

# Turn on DebugMessage logging for both subsystems
# to a file on disk.
[logfile /temp/vw_test.log]
40 = cache
40 = thread
```

Listing 4: [`LogConf.example`] An example log configuration file.

Chapter 7

Camera Module

Cameras are the interface between images and the real world, and as such, their importance in computer vision cannot be understated. In fact, some would say that computer vision algorithms are distinguished by the fact that they endeavor to associate the processed pixel data with objects in the real world for the purposed of measurement, tracking, or display. This is achieved by modeling the geometric and physical properties of the device that was used to capture the original image. This is the purpose of the camera module.

The camera module includes built-in models for generic pinhole and line-scan imagers and a set of generic functions for linearizing (removing lens distortion) and epipolar rectifying (e.g. for stereo) when these operations are relevant. These classes and functions can be imported into your code by including `<vw/Camera.h>`.

Because you will likely encounter new camera geometries not supported by the built-in classes, the camera module is designed to be extensible. You can provide your own camera model by inheriting from and adopting the interface of the `CameraModel` abstract base class, which is defined in the header file `<vw/Camera/CameraModel.h>`.

Finally, the camera module provides a basic set of tools for working with images from real-world camera systems: bayer pattern filtering and EXIF parsing. We will cover all of these features in more detail, but we begin this chapter by establishing some terminology while exploring the most common camera geometry in use today: the pinhole camera model.

7.1 The Pinhole Camera Model

The pinhole camera model describes the geometry found in nearly all commercial digital camera systems today. It is characterized by a lens assembly that focuses light onto a two dimensional array of pixels (usually a sensor with light sensitive circuits such as a CCD or CMOS device). We will use the pinhole model to establish some terminology that will be used throughout the rest of this chapter. Be warned that the model we are about to develop is simplistic; many of the non-ideal characteristics of a real-world optical system (e.g. lens distortion) are not modeled in this simple example. Refer to the CAHVOR model in Section 7.3 if you require a pinhole camera model that more accurately models lens distortion.

7.1.1 Perspective Projection

Figure 7.1 shows the geometry of a basic pinhole camera. The light gray area represents the 2D array of pixels, and it is referred to as the *image plane*. The origin of the 3D coordinate system is

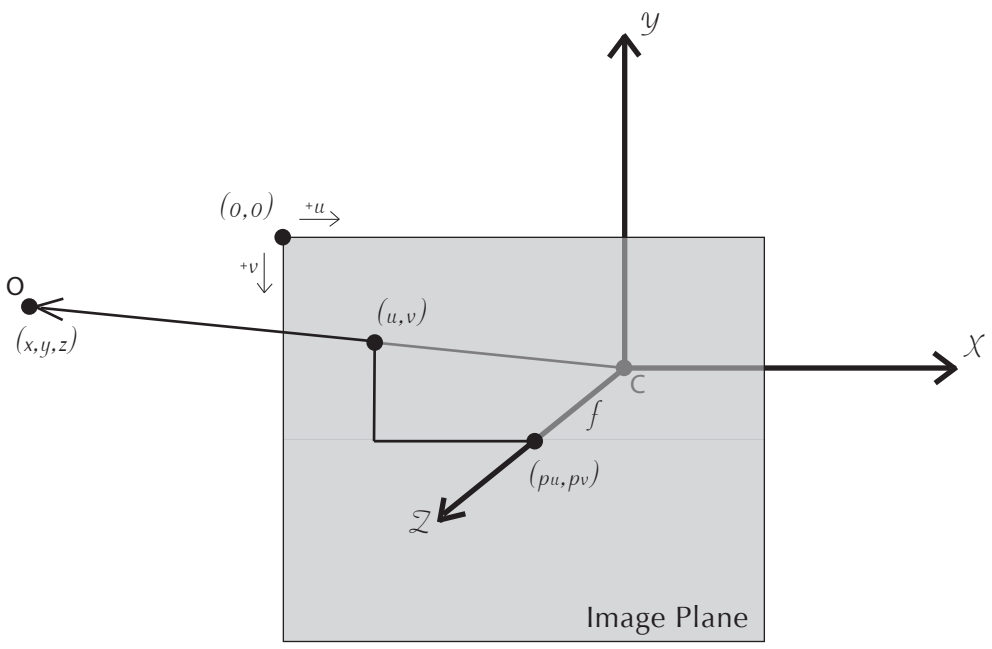


Figure 7.1: The basic pinhole camera model.

the point C, which is the center of projection or *camera center* of the imager. When a 3D point O is imaged by the camera, it appears at the pixel located where segment $\overline{OC}$ intersects the image plane at point (u, v). A line segment $\overline{OC}$ that is perpendicular to the image plane intersects this plane at the *principal point*, (p_u, p_v).

All of the points imaged by the camera appear on a line that passes through C. If the coordinates of $\mathbf{O}$ are (x, y, z), then the position of the point on the imager can be determined by projecting it onto the plane $z = +f$:

$$u = \frac{f}{\sigma}\left(\frac{x}{z}\right) - p_u \tag{7.1}$$

$$v = \frac{f}{\sigma}\left(\frac{-y}{z}\right) - p_v \tag{7.2}$$

Here, f is the focal length of the imager in meters, σ is the size of a pixel in *m/pixel*, and (p_u, p_v) are the offset in pixels of the principal point (this offset moves the origin of the image from the principal point to the upper left hand corner, which is the "origin" usually adopted when indexing images).

Equations 7.1 and 7.2 constitute the *forward projection* portion of the camera model; this is analogous to the process of "capturing" an image with a real camera. Our model should tell us exactly what what pixel location to look at if you wanted to see the point O in the image.

Notice how some information is lost during forward projection. Any point along $\overline{OC}$ will be imaged to the same point (u, v) on the image plane, so if we were to start with a point $P = (u, v)$ on the image plane, and we wanted to find the original 3D point O, the best we could do would be to say that it appears somewhere along the ray $\overrightarrow{CP}$. The origin and direction of the ray can be computed as follows:

$$\overrightarrow{CP}_{origin} = C \tag{7.3}$$

$$\overrightarrow{CP}_{direction} = \frac{(u + p_u, -(v + p_v), f)}{||(u + p_u, -(v + p_v), f)||_2} \tag{7.4}$$

This operation, which we call *back projection*, can still provide useful information that can be used in a full 3D reconstruction despite the ambiguity in the actual position of O. Imagine that you have two cameras that have imaged the same point O from two different viewpoints at pixel locations P_1 and P_2 in their respective image planes. Using simple geometry, you can reconstruct the position of O by computing the intersection of the two rays emanating from each camera center through P_1 and P_2. This is the technique commonly referred to as stereo reconstruction, and it is one of the many ways that you can make use of the information provided by back projection.

7.2 The Camera Model Base Class

As we have seen, a camera model provides a means for forward projection ("imaging" 3D points onto a 2D array of pixels) and back projection (finding the ray along which a 3D points must lie given a 2D pixel where it was imaged). All camera models in the Vision Workbench derived from the CameraModel abstract base class, which enforces this basic interface.

Forward projection of a 3D point is handled by the `point_to_pixel()` method.

```
CameraModel* camera_model = new MyDerivedCameraModelClass;
Vector3 world_coordinates;
Vector2 pixel_coordinates = camera_model.point_to_pixel(world_coordinates);
```

The back projection operation is split into two separate API calls. The `camera_center()` method returns the origin of the ray, and the `pixel_to_vector()` method returns its direction. Remember that any of the points that lie along this ray would have been imaged at (u, v), so this pixel-to-ray operation leaves some ambiguity about the true location of the point O.

```
CameraModel* camera_model;
Vector3 ray_origin = camera_model.camera_center(pixel_coordinates);
Vector3 ray_direction = camera_model.pixel_to_vector(pixel_coordinates);
```

Camera Coordinate Systems

Vision Workbench camera model classes take and return coordinates that are *not* homogeneous. That is, coordinates do not need to augmented with an additional homogeneous scaling element before being passed to camera module routines (e.g. a 2D vector $(325, 206)$ in cartesian coordinates is often represented as $(325, 206, 1)$ in homogeneous coordinates). Homogeneous coordinates have certain advantages in projective geometry (e.g. they allow a translation of the coordinates to be encoded as a matrix multiplication), however we have chosen not to adopt this convention.

Camera Model	*Header File*	*Imager Type*	*Details*
CAHV	`CAHVModel.h`	Pinhole	Basic pinhole camera model
CAHVOR	`CAHVORModel.h`	Pinhole	Models lens distortion
Linescane	`LinescanCameraModel.h`	Linescan	Generic Linescan Model
Linear Pushbroom	`LinearPushbroomModel.h`	Linescan	Assumes linear flight path
Orbiting Pushbroom	`OrbitingPushbroomModel.h`	Linescan	Models curvature of orbit

Table 7.1: Built-in camera models can be found in `vw/Camera/`

7.3 Built-in Camera Models

The Vision Workbench comes with several "built-in" camera models. These classes satisfy the needs of most common applications, and they can also serve as a design reference for your own camera model classes. Each class models a specific geometry and, to varying extents, the non-ideal characteristics of the camera system such as lens distortion.

The list of built-in models are summarized in Table 7.3. The following sections describes the two basic classes of built-in camera model: those that model pinhole cameras (where the imager is a 2D array of pixels), and those that model linescan cameras (where the imager is a 1D line of pixels).

7.3.1 Pinhole Cameras

The *CAHV camera model* has been widely used in NASA planetary mission for rover navigation and scientific camera systems [4]. It is a basic pinhole camera model that does not model lens distortion or other optical aberrations. The CAHV model is so named because the camera intrinsic and extrinsic parameters are jointly parameterized by four 3-dimensional vectors: C,A,H, and V. This compact representation leads to a very efficient forward projection operation, which is the strength of the CAHV model. Forward projection of a real world point O can be computed using

$$u = \frac{(O-C)\cdot H}{(O-C)\cdot A} \tag{7.5}$$

$$v = \frac{(O-C)\cdot V}{(O-C)\cdot A} \tag{7.6}$$

The user has two choices when initializing a CAHV camera model. First, they can construct the object by directly supplying four 3-vectors to the constructor.

```
Vector3 C,A,H,V;
CameraModel* cam = new CAHVModel(C,A,H,V);
```

Alternatively, users seeking to use the CAHV class as a general purpose pinhole camera model may find it easier to use the more verbose constructor wherein the camera extrinsics and intrinsics are explicitly supplied.

```
double focal_length;
Vector2 pixel_size;
double principal_point_h, principal_point_v;
```

```
Vector3 camera_center, pointing_vector;
Vector3 horizontal_vector, vertical_vector; // Defines image plane orientation

CameraModel* cam = new CAHVModel(focal_length, pixel_size, principal_point_h,
                                 principal_point_v, camera_center,
                                 pointing_vector, horizontal_vector,
                                 vertical_vector);
```

The *CAHVOR model* is an expanded camera model with two additional 3-vectors (O and R) that describe lens distortion introduced by the camera lens.

Refer the Doxygen-generated API documentation for more information about constructing CAHV and CAHVOR camera models.

7.3.2 Linescan Cameras

Linescan imagers capture images using a sensor containing a 1-dimensional array of pixels. The image is formed by capturing successive scan-lines as the camera platform is rotated or moved. For example, a flat-bed scanner or photocopier is a familiar example of such a system. The sensor is swept in the so-called *along-track* direction along the document, composing the final image by concatenating several thousand adjacent 1-dimensional images taken at evenly spaced positions.

Linescan sensors are fairly uncommon in commercial camera systems, but they appear frequently on satellites that capture photographs of terrain from orbit. The orbital motion of the satellite is used in much the same way as the motion of the sensor in the photocopier; the *across-track* dimension of the image corresponds to the projection of 3D points through the lens onto the sensor, and the along-track dimension of the image corresponds to successive scanlines taken at different times as the satellite moves.

The geometry of the linescan imager is subtly different from the geometry of the pinhole camera. See Figure 7.2. Although there is still a center of projection created by the camera's optics (and points are still imaged using a perspective projection in the across-track direction), this point moves as the camera moves, and as a result, the along-track position of a pixel in a linescan image is purely a function of the position and orientation of the camera; both a function of time. The orientation of the image (in the sense that u indexes the columns of an image and v indexes its rows) must be chosen consistent with Figure 7.2 when working with `LinescanModel` and its relatives.

In the special case where the motion of the linescan sensor is linear and its orientation is fixed, the projection of the points onto the image in the along-track direction is orthographic. These assumptions are the basis for the *Linear Pushbroom Model*, which can be found in the header file `<vw/Camera/LinearPushbroomModel.h>`. If you are interested in understanding this model in detail, we recommend you read the excellent paper by Gupta and Hartly [2].

If you must relax the assumption about a linear flight path somewhat to allow sensor pose to vary and the camera motion to lie along a curve (as is common with orbiting camera systems), the *Orbiting Pushbroom Model* is an appropriate choice. In the Orbiting Pushbroom model, the user supplies a series of evenly spaced samples of position and orientation and specifies the time interval (in seconds) between samples. A sparse set of samples is sufficient for this model: interpolation occurs for points in between the supplied positions and orientations.

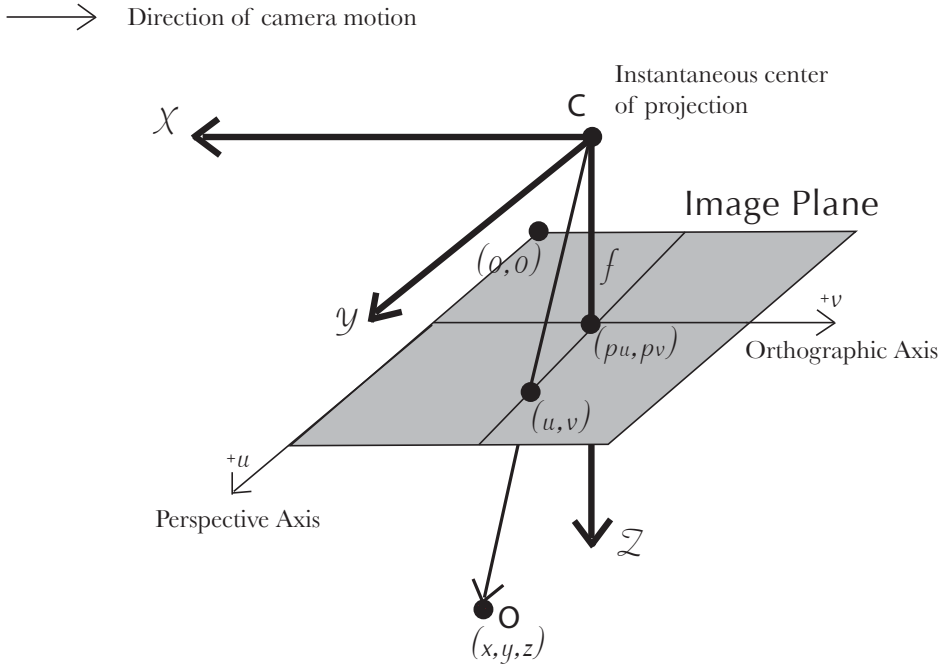


Figure 7.2: Geometry of the Linescan Camera Model.

7.4 Tools for Working With Camera Images

This section describes several tools that simplify the process of working with image captured by real cameras.

7.4.1 Inverse Bayer Pattern Filtering

Most imaging sensors are inherently grayscale capture devices. In order to capture color, some imagers have a hardware color filter placed in front of the pixels on the CCD. This is called a *Bayer filter*. The Vision Workbench provides the `inverse_bayer_filter()` function (found in the header file `<vw/Camera/BayerFilter.h>`) which interprets the raw, grayscale pixel values from the sensor and produces a color image by interpreting the Bayer filter effect.

7.4.2 Exif Exposure Data

Digital cameras store data about the settings used to take a picture in the image file according to the EXIF standard [1]. EXIF data is stored using the Image File Directory system described in the TIFF 6.0 standard. EXIF tags such as `FNumber` and `ExposureTime` can be useful for radiometrically calibrating your camera images. Unfortunately the standard is redundant and often poorly supported by camera manufacturers (for example, many hide the ISO setting in the maker note instead of storing it in the ISOSpeedRatings tag), so we cannot guarantee support for every camera.

The Camera module includes the `ExifView` class (defined in `<vw/Camera/Exif.h>`) for extracting this data from images. To create this class, you supply the filename of an image on disk.

Currently, JPEG and TIFF images are supported. ExifData and ExifView were based on `jhead`, an EXIF JPEG header and thumbnail manipulator program in the public domain [3].

```
// Reliably get F number.
ExifView view;
if (view.load_exif(``img.jpg'')) {
  double f = view.get_f_number();
  // ...
}
```

Bibliography

[1] "Exchangeable image file format for digital still cameras: Exif Version 2.2", (Japan Electronics and Information Technology Industries Association, 2002), http://www.exif.org/specifications.html.

[2] Gupta, Rajiv and Hartley, Richard. "Linear Pushbroom Cameras". IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.19 No. 9. September 1997

[3] Wandel, Matthias, "Exif Jpeg header and thumbnail manipulator program," 2006, http://www.sentex.net/ mwandel/jhead/.

[4] Yakimovsky, Y. and Cunningham R., "A System for Extracting Three-Dimensional Measurements from a Stereo Pair of TV Cameras " Computer Graphics and Image Processing 7, pp. 195-210. (1978)

Chapter 8

`Mosaic` Module

The `Mosaic` module provides a number of tools for assembling and generating image mosaics, i.e. large images that are composed of a number of smaller images. The applications of this module include panoramic imaging and aerial/satellite image processing. There are three major facilities provided at this time: compositing many images together, using multi-band blending to seamlessly merge overlapping images, and generating on-disk image quad-trees to efficiently store very large images.

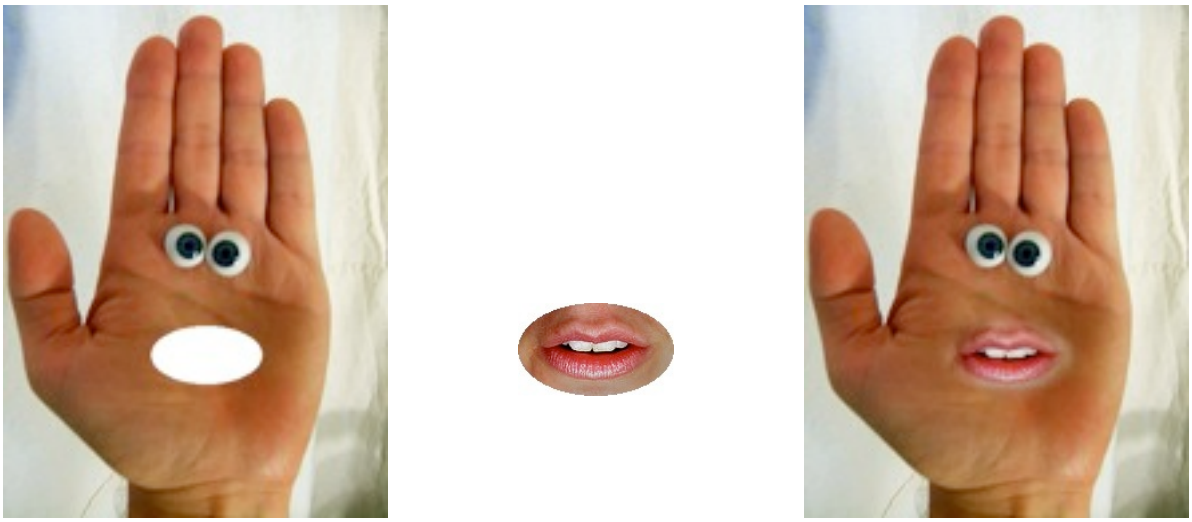
Note that the facilities described in this chapter are currently under active development, and there may be some API changes in future releases as new capabilities are added.

8.1 `ImageComposite` and Multi-Band Blending

The `ImageComposite` template class provides the ability to composite any number of source images together at arbitrary pixel offsets. It was originally designed for assembling tiled panoramas or aerial/satellite images that have been transformed into a common coordinate system, though it can be used for many other things as well.

The interface is fairly simple. Just like ordinary Vision Workbench images, an `ImageComposite` is templatized on its pixel type. In most cases you will want to use a pixel type that has an alpha channel, and if you want to perform image blending then the pixel type must be floating-point, so the most common pixel type is `PixelRGBA<float32>`. You can then configure whether you would like to use multi-band blending to merge your images or if you would simply like them overlayed by using the `set_draft_mode()` method. It takes a single argument which should be `true` if you simply want to overlay the images and `fales` if you want to use the blender. Blending is a significantly more expensive operation. If your images are simply tiles and do not overlap then draft mode is probably what you want. In blending mode you also have the option of asking the blender to attempt to fill in any missing (i.e. transparent) data in the composite using information from the neighboring pixels. You can enable or disable this behavior by calling the the `set_fill_holes()` method.

Once you have created the composite object, you add source images to it using the `insert()` method, which takes three arguments: the image you are adding, and the x and y pixel offset of that image within the composite. The `ImageComposite` does not store a copy of your image. Instead, it only stores a reference to it in the form of an `ImageViewRef` object. This means that you can easily do things like create a composite of images that could not all fit in memory simultaneously, e.g. by passing in `DiskImageView` objects. Note that only integer pixel offsets are supported: if you want to shift an image by a fractional amount you will first need to transform it accordingly. In most



(a) First source (b) Second source (c) Blended result

Figure 8.1: Example input and output images from the `ImageComposite` multi-band blender.[1]

cases you will need to pre-transform your source images anyway, so this applies no extra cost.

Once you have added all your images, be sure to call the `ImageComposite`'s `prepare()` method. This method takes no arguments, but it does two things. First, it computes the overall bounding box of the images that you have supplied, and shifts the coordinate system so that the minimum pixel location is $(0, 0)$ as usual. (For advanced users, if you prefer to use a different overall bounding box you may compute it yourself and pass it as an optional `BBox2i` argument to the `prepare()` method.) Second, if multi-band blending is enabled, it generates a series of mask images that are used by the blender. Currently these are saved as files in the current working directory. This is admittedly inconvenient behavior and will be changed in a future release.

Now that you've assembled and prepared your composite you can use it just like an ordinary image, except that per-pixel access is not supported. If the image is reasonably small then you can rasterize the entire image by assigning it to an `ImageView`. Alternatively, if the composite is huge the usual next step is to pass it as the source image to the quad-tree generator, discussed in the next section. You can also use `ImageComposite`'s special `generate_patch()` method to manually extract smaller regions one at a time. It takes a single `BBox2i` bounding-box, expressed in the re-centered coordinate frame, as its only argument.

Here's a simple example that illustrates how you might blend together a number of images on-disk. It assumes you already know the image filenames and their offsets within the composite, and that the total composite is small enough to sensible rasterize all at once.

```
ImageComposite<PixelRGBA<float> > composite;
for( int i=0; i<num_images; ++i ) {
  composite.insert( DiskImageView<PixelRGBA<float> >( image_filename[i] ),
                    image_offset[i].x(), image_offset[i].y() );
}
composite.prepare();
write_image( "composite.png", composite );
```

For a somewhat more fleshed-out example of how to blend images, see the example program `blend.cc` included with the `Mosaic` module sources.

8.2 ImageQuadTreeGenerator

The ability to assemble composites that are far larger than could be stored in memory all at once presents serious challenges. When viewing an image of that size, the ability to zoom in and out is of critical importance. Only a small fraction of the image data will ever be on-screen at a time at full resolution. However the entire data set may be visible at once at lower resolutions, and computing such reduced images on the fly can be prohibitively expensive. The usual solution to this problem is to pre-compute sub-sampled versions of the image at multiple levels of detail. The sub-sampling factors are often chosen to be successive powers of two, and the data at each resolution is typically chopped up into tiles for faster direct access.

The `ImageQuadTreeGenerator` class generates just such a representation of an image. You specify the tile size, the generator reduces the image by powers of two until it fits in a single tile. To generate each successive level of detail every tile is replaced by four tiles at twice the resolution. The resulting quad-tree of images is stored on disk in a hierarchical manner, making it easy to instantly access any tile at any resolution.

Like many things in the Vision Workbench, an `ImageQuadTreeGenerator` is templatized on its pixel type. The constructor takes two arguments, the pathname of the tree to be created on disk and the source image. You can then use several member functions to configure the quad-tree prior to generating it. The `set_bbox()` method, which takes a single `BBox2i` parameter, specifies that a particular region of the source image to be processed instead of the entire thing. The `set_output_image_file_type()` method sets the file type of the generated image files; the default is "`png`". The `set_patch_size()` function takes an integer argument specifying the patch size in pixels. Relatedly, the `set_patch_overlap()` function specifies how many pixels of the border of each patch overlap the neighboring patch. The default is 256-pixel patches with no overlap. Finally, the `set_crop_images()` method takes a boolean argument that controls whether or not the resulting images are cropped to the non-transparent region. Image cropping is enabled by default.

Once you have configured the `ImageQuadTreeGenerator` to your liking you can invoke its `generate()` method to generate the quad-tree on disk. It is stored in a directory with the name you provided as the first argument to the constructor with the extension "`.qtree`" appended. For example, if you specified the name of the quad-tree as `/home/mdh/example` then the result is stored in the directory `/home/mdh/example.qtree`. This directory typically contains three things. First is the lowest-resolution image of the tree, essentially a thumbnail, which is stored in an image with the same base name as the tree with the appropriate image file format extension appended. To continue the above example, if the file format type is "`png`" then the top-level image file's pathname will be `/home/mdh/example.qtree/example.png`. The next file in the top-level directory is a simple text file describing the bounding box of the top-level patch, with the same name but with the extension `.bbx` instead. The format of this file will be discussed below. Finally there is a subdirectory, which has the same name but no extension, that contains the next level of the tree.

Inside that subdirectory there are generally four image files, with names `0.png`, `1.png`, `2.png`, and `3.png`, containing to the four image patches at the second level of detail. The patches are numbered left-to-right and top-to-bottom, so `0` is the upper-left patch, `1` is the upper-right patch, and so on. There are also four corresponding `.bbx` files and four directories containing higher-resolution data for each patch. Each subdirectory likewise has four numbered images, bounding boxes, and further subdirectories. For example, the file `/foo/bar/myimage.qtree/myimage/0/1/3.png` would

[1] Original hand and face source images by `sheldonschwartz` and `vidrio`, respectively, and released under the Creative Commons license.

(a) Draft mode (simple overlay)

(b) Multi-band blending

Figure 8.2: A twelve-image mosaic composited using an `ImageMosaic` object, first (a) in draft mode and then (b) using multi-band blending.

Figure 8.3: Patches at three successive resolutions generated by an `ImageQuadTreeGenerator` constructed with the name "`Walker`". The files are named `Walker.qtree/Walker.png`, `Walker.qtree/Walker/1.png`, and `Walker.qtree/Walker/1/2.png`, respectively.

be an image at the fourth level of detail. The subdirectories at the highest level of detail have no further subdirectories. Note that if cropping is enabled then it is possible that some directories will not have all four images; this occurs if any of the images is entirely empty.

Each `.bbx` file contains eleven numbers, represented as text strings, each on a line by itself. The first is the scale factor of that tile: it has the value 1 for the highest-resolution patches and values of the form 2^n for lower-resolution patches. You can equivalently think of this number as describing the size of each pixel at this resolution, as measured in full-resolution pixels. The next four numbers describe the bounding box of the patch within the full-resolution coordinate system. First are the x and y coordinates of the upper-left pixel, and then come the width and height of the image in pixels. To reiterate, these are measured in *full-resolution* pixels, and so these numbers will generally be multiples of the scale factor.

After this come a similar set of four numbers describing the bounding box of the unique image data in this patch. This is generally the same as the previous bounding box if there is no patch overlap. However, if the patch overlap has been set to a nonzero value then this second bounding box will describe the portion of this patch that does not overlap the corresponding regions of the neighboring patches. In other words, taken together these second bounding boxes perfectly tile the entire image without overlap. Finally, the last two values describe the width and height of the entire source image in pixels.

This file format is obviously quite arbitrary, and was designed to make it easy to write scripts to manipulate the quadtree data later. If you prefer, you can subclass the quadtree generator and overload the `write_meta_file()` function to generate metadata of a different sort.

Chapter 9

High Dynamic Range Module

Photographers have long understood that the range of brightness levels in a real-world scene is considerably larger than the range that can be captured by a camera's film or imaging sensor. The luminance of a outdoor scene can easily span five orders of magnitude, however typical digital cameras encode the brightness at a single pixel location using a modest 8-bits (2 orders of magnitude). Pixels that fall outside the *dynamic range* of the camera sensor will either be underexposed or overexposed; their values will be at the minimum or maximum pixel value of the camera, respectively.

Some digital cameras can save RAW images with higher dynamic range. These cameras can capture 12 bits per pixel, or 4096 brightness levels. This expanded dynamic range is often sufficient to capture scenes with a moderate amount of contrast. However, to capture scenes with very high dynamic range, you must generate a HDR image from a set of *bracketed* exposures: a group of low dynamic range images of the exact same scene taken with different exposure settings that vary from under-exposed to over-exposed. This technique is subject of section 9.1.

The resulting HDR image is generally stored with a channel type of `float` or `int16` to accommodate the expanded dynamic range. To store HDR images on disk we recommend using the OpenEXR or TIFF file formats, both of which support 32-bit floating point channel types.

Just as capturing HDR images can be challenging, visualizing them is also difficult. Print media and most display devices can only manage about 2 orders of magnitude of dynamic range, whereas an HDR image may have 5 or more orders of magnitude. Simply scaling the image to the display range will cause it to look overly dark or washed-out. Section 9.2 discusses a technique called *tone mapping* that reduces the dynamic range of an image while preserving as much as possible the visual contrasts of the original scene.

9.1 Merging Bracketed Exposures

As discussed in the introduction to this chapter, the limited dynamic range of the sensors on modern cameras necessitates a different strategy for building true HDR images: exposure bracketing. This frees us from hardware limitations and allows us to set the dynamic range of the output image arbitrarily by adjusting the number of exposures in the bracket. The more exposures in the bracket, the higher the dynamic range merged HDR image.

For convenience, the exposure ratio between consecutive images is usually a fixed value to guarantee a wide exposure range while maintaining even brightness overlap between adjacent images in the bracket. A factor of two is generally recommended. The shutter speed is generally the preferred method of varying the exposure in a bracket; changing the aperture or ISO setting can

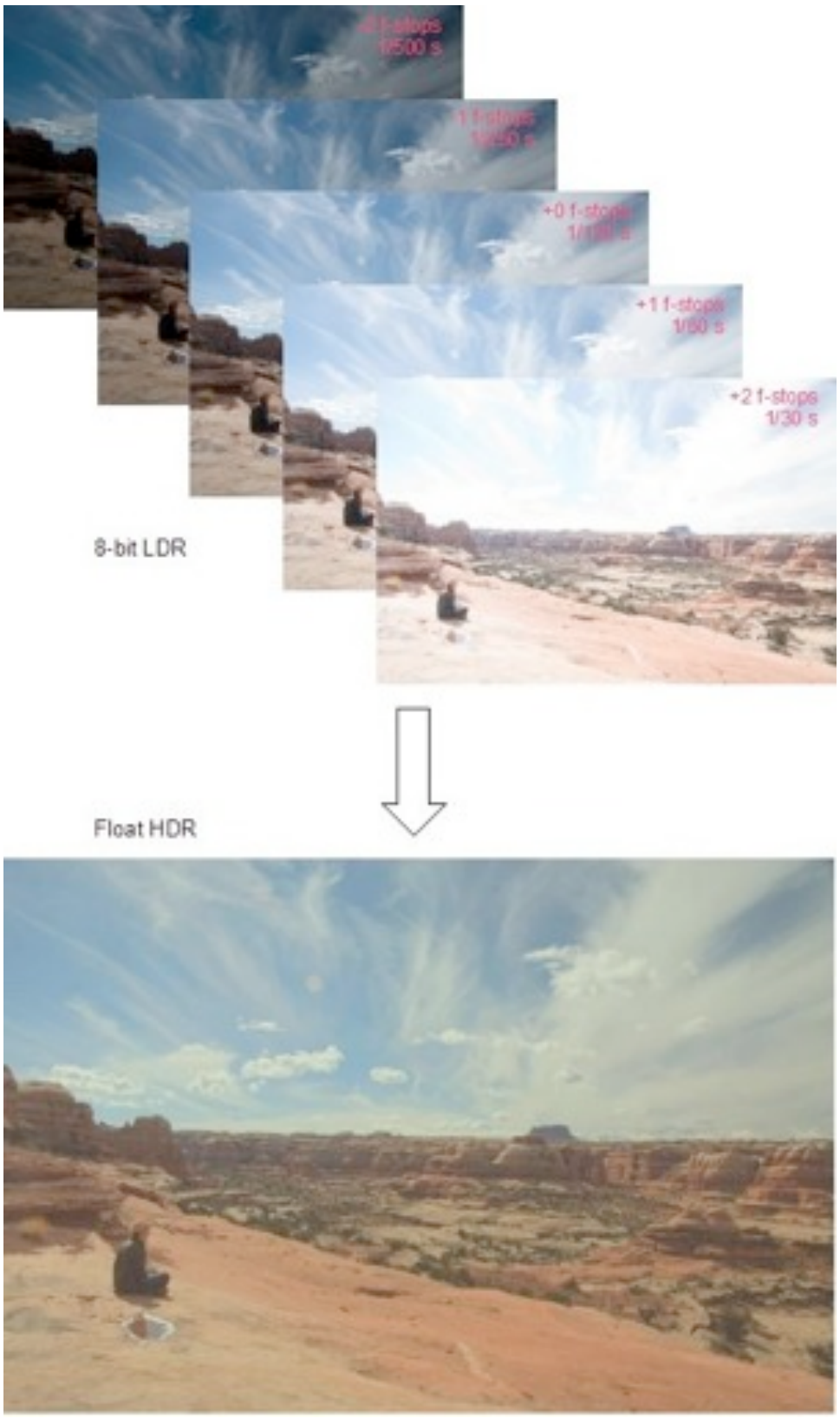


Figure 9.1: A stack of 8-bit-per-channel LDR images separated by one f-stop is merged a floating-point HDR image. The HDR image is tone-mapped for display using the Drago operator.

have the same effect on exposure but they may change the focus or increase noise.

9.1.1 Converting LDR Images to an HDR Image

The HDR module has a set of free functions that make stitching a stack of LDR images into an HDR image as simple as one function call. `process_ldr_images()` (in `<vw/HDR/LDRtoHDR.h>`) takes a `std::vector` of ImageViews (with grayscale or RGB pixels and a floating point channel type), sorted from darkest to brightest. This function assumes a constant exposure ratio between images; it can be specified or the default value of $\sqrt{2}$ can be used (corresponding to 1 f-stop, or a power of two in shutter speed). An overloaded version accepts a `std::vector<double>` of absolute brightness values for each image (as defined by the APEX system [6]) .

```
// Generate HDR image from HDR stack.
vector<ImageView<PixelRGB<double> > > images(num_images);

// ... Read input images ...

// Assum default exposure ratio
ImageView<PixelRGB<double> > hdr_image = process_ldr_images(images);
```

Most modern digital cameras store exposure information as EXIF metadata in the headers of images. The Vision Workbench supports reading this data from TIFF and JPEG files via the Camera Module, and the routine `process_ldr_images_exif()` capitalizes on this, generating an HDR image from an array of filenames of LDR images by computing brightness values from the files' EXIF data along the way.

9.1.2 The Camera Response Curves

The relationship between light entering a camera sensor and the pixel value that is recorded in the image is a non-linear. It depends on many factors, including the degree of gamma correction used in the image, the color balance across the range of brightness, and the camera's post processing settings (especially contrast). The function that encompasses all of these and maps the amount of light on the sensor (the luminance) to the pixel value is referred to as the camera response curve. See Figure 9.2 for an example.

To create a HDR image that truly represent the physical brightness levels in a scene, it is necessary to estimate the inverse of the camera response curves (we assume a separate curve for each channel) and apply it to the image data. That is, we would like to find a function that, given a pixel value in the image and its exposure information, returns the luminance of the scene at that point. `<vw/HDR/CameraCurve.h>` implements `estimate_inverse_camera_curve()`, a free function that estimates the inverse response curve as a polynomial. This routine takes a matrix of aligned pixel channel values from an HDR stack and their brightness ratios. This function is used internally by `LDRtoHDR.h`; most users will probably not need to call it directly.

The CameraCurve sources also supply `invert_curve()`, a free function that approximates the inverse of a polynomial curve, also as a polynomial. This is useful to recover the actual camera response curve (mapping a scaled luminance value to a pixel value) from the inverse response curve determined by `estimate_inverse_camera_curve()`. Re-applying the camera response curve to an HDR image after tone-mapping can improve the colors and overall visual appeal. See the code in Listing 5 for an example.

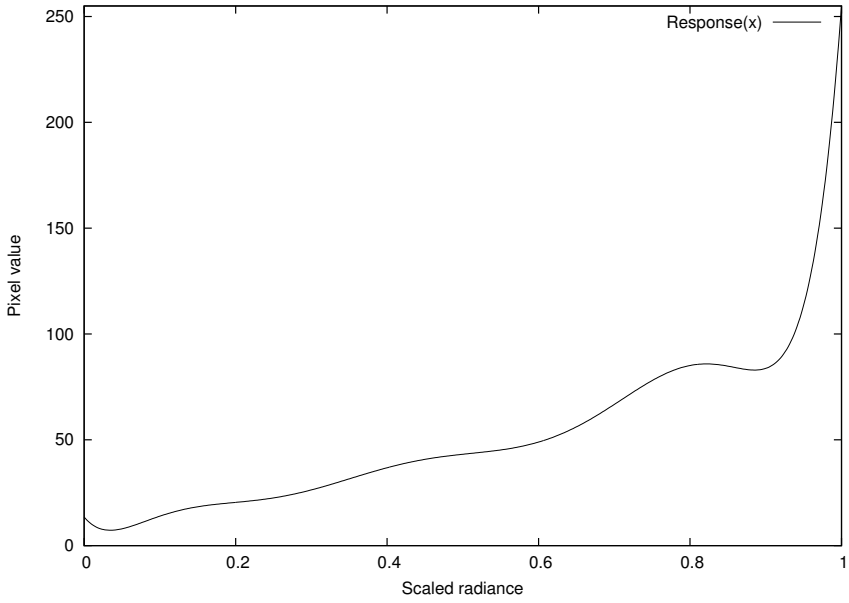


Figure 9.2: Camera response curve estimated from an HDR stack.

9.2 Tone Mapping

Since print media and most display technologies are inherently LDR, the dynamic range of a HDR image must be compressed before it can be displayed. Simply scaling the pixel luminances linearly yields poor results because the human visual system's response to luminance is approximately logarithmic rather than linear. A linear scaling tends to lose small details and local contrast, and the image as a whole will appear under or over-exposed.

A wide variety of tone-mapping operators have been proposed to compress the dynamic range of a HDR image while preserving details and local contrast as much as possible. Using an ideal tone-mapping operator, the observer of a tone-mapped LDR image would have a perceptual response matching that of the original HDR scene. Due to its greater realism, tone-mapping can vastly improves the appearance of the displayed image.

There are several broad classes of tone-mapping operators, including global operators, local operators, and operators that use the gradient or frequency domains. The HDR module currently includes one global operator and one local operator; they are described in the following sections.

9.2.1 Global Operators

Global tone-mapping operators apply the same compressive function to all pixels in the image. Such operators are implemented in `<vw/HDR/GlobalToneMap.h>`. Currently one such operator is implemented, the Drago Adaptive Logarithmic Mapping operator. For algorithm details see *High Dynamic Range Imaging* [9] or the original paper [3]. The Drago operator is currently the operator of choice in the HDR module. It is quite fast, produces good results for a wide range of images, and can usually be used with its parameters at default values. Optionally, the parameters bias

Figure 9.3: Top: an image taken in 12-bit RAW format. Middle: after tone-mapping with the Drago operator. Bottom: after tone-mapping with the Ashikhmin operator.

(controlling contrast, usually between 0.7 and 0.9), exposure factor (a simple multiplier to control the overall brightness of the image), and max display luminance (usually about 100) can be specified.

```
// Apply Drago tone-mapping operator
ImageView<PixelRGB<float> > tone_mapped = drago_tone_map(hdr_image);
```

9.2.2 Local Operators

Local tone-mapping operators compress a pixel's dynamic range in a way dependent on the neighborhood of pixels around it. These operators mimic the local adaptation of the human eye and are capable of more striking or artistic results than global operators, but they are also susceptible to artifacts such as excessive haloing and reverse gradients. `<vw/HDR/LocalToneMap.h>` currently implements the Ashikhmin local tone-mapping operator [1]. It is much slower than the Drago operator and more prone to artifacts, but may be useful for some images. Its only parameter is a threshold value (0.5 by default) which roughly controls the size of the neighborhood used for each pixel. A threshold value too large will result in haloing.

```
// Apply Ashikhmin tone-mapping operator
ImageView<PixelRGB<float> > tone_mapped = ashikhmin_tone_map(hdr_image);
```

9.3 Command Line Tools

The HDR module builds two small utilities for working with HDR images from the command terminal. If you simply type these command names with no arguments, you will see a list of acceptable arguments.

- `hdr_merge` : Merge LDR images into one HDR image
- `hdr_tonemap` : Tonemap an HDR image using the Drago operator

9.4 Other Resources

There are a number of freely available utilities which are useful for working with HDR images. The OpenEXR distribution [5] includes several utilities, including exrdisplay for displaying OpenEXR images. `exrtools` [2] provides utilities for converting between OpenEXR and other formats, performing basic operations on OpenEXR images, and a couple of tone-mapping utilities. `pfstools` [7] is a well-integrated set of utilities for reading, writing, manipulating and viewing HDR images. It has an associated pfstmo project that implements seven of the more prominent tone mapping operators.

```
1  // __BEGIN_LICENSE__
2  // Copyright (C) 2006, 2007 United States Government as represented by
3  // the Administrator of the National Aeronautics and Space Administration.
4  // All Rights Reserved.
5  // __END_LICENSE__
6
7
8  #include <vw/vw.h>
9  #include <stdio.h>
10 #include <vector>
11 using namespace vw;
12 using namespace vw::HDR;
13
14 int main(int argc, char** argv) {
15   std::vector<Vector<double> > curves;
16   std::vector<string> files;
17   for(int i = 1; i < argc; i++)
18     files.push_back(argv[i]);
19   // Process HDR stack using Exif tags
20   ImageView<PixelRGB<double> > hdr_exif = process_ldr_images_exif(files,
21                                           curves);
22   write_image("hdr.exr", hdr_exif);
23
24   // Apply Drago tone-mapping operator.
25   ImageView<PixelRGB<double> > tone_mapped = drago_tone_map(hdr_exif);
26   write_image("tm.jpg", tone_mapped);
27
28   // Apply gamma correction and save.
29   ImageView<PixelRGB<double> > gamma = pow(tone_mapped, 1.0/2.2);
30   write_image("tm_gamma.jpg", gamma);
31
32   // Re-apply camera response curves and save.
33   // First must invert curves calculated earlier.
34   std::vector<Vector<double> > inverse_curves(curves.size());
35   for (int i = 0; i < curves.size(); i++) {
36     invert_curve(curves[i], inverse_curves[i],
37                  VW_HDR_RESPONSE_POLYNOMIAL_ORDER);
38   }
39   psi(tone_mapped, inverse_curves);
40   write_image("tm_curved.jpg", tone_mapped);
41
42   // Apply gamma correction after response curves.
43   // Usually gives best results.
44   ImageView<PixelRGB<double> > tm_c_g = pow(tone_mapped, 1.0/2.2);
45   write_image("tm_c_g.jpg", tm_c_g);
46
47   return 0;
48 }
```

Listing 5: [ExifHDRExample.cc] This is a simple test program that stitches an Exif-tagged HDR stack into an HDR image, performs tone-mapping, and saves several versions with different post-processing applied for comparison. Usually the best image is produced by re-applying the camera response curves and then gamma correcting.

Bibliography

[1] Ashikhmin, Michael, "A Tone Mapping Algorithm for High Contrast Images," *Eurographics Workshop on Rendering*, 2002: 1–11.

[2] Biggs, Billy, "exrtools: a collection of utilities for manipulating OpenEXR images,", 2004, http://scanline.ca/exrtools/.

[3] Drago et al., "Adaptive Logarithmic Mapping For Displaying High Contrast Scenes," *Eurographics*, **22**(3), 2003.

[4] Fattal, Raanan, et. al, "Gradient Domain High Dynamic Range Compression," *ACM Transactions on Graphics*, 2002.

[5] Industrial Light and Magic, "OpenEXR," (Lucasfilm Ltd., 2006), http://www.openexr.com.

[6] Kerr, Douglas, "APEX–The Additive System of Photographic Exposure," 2006, http://doug.kerr.home.att.net/pumpkin/APEX.pdf.

[7] Mantiuk, Rafal, and Grzegorz Krawczyk, "pfstools for HDR processing," 2006, http://www.mpi-inf.mpg.de/resources/pfstools/.

[8] Reinhard, Erik, et. al. "Photographic Tone Reproduction for Digital Images," *ACM Transactions on Graphics*, 2002.

[9] Reinhard, Erik, Greg Ward, Sumanta Pattanaik, and Paul Debevec, *High Dynamic Range Imaging*, (Boston: Elsevier, 2006).

[10] Ward, Greg, et al., "A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes," *IEEE Transactions on Visualization and Computer Graphics*, 1997.

Chapter 10

Cartography Module

[*Note: The documentation for this chapter is not yet complete...*]

The earliest robots in space were not planetary rovers – they were unmanned probes that studied the planetary bodies in our solar system from afar. Today there are roughly twenty extraterrestrial spacecraft in active communication with earth (and only two planetary rovers), so the bulk of the extraterrestrial data that we receive consists of imagery that originated on round(ish) surfaces. The natural thing to do with this data is to merge it together into a map, but when doing so we are faced with the same problem that has plagued cartographers for hundreds of years: how does one flatten the globe? This is the job of the Vision Workbench Cartography module: to make maps.

Before diving into an introduction on planetary cartography, we will point out another problem that is relatively new to Cartography. The amount of map data that we have collected about Earth and the other bodies in our solar system is *immense*. It is often impossible to store an entire mapping data set in memory all at once, so intelligent paging, caching, and storage strategies must be used in order to make working with this data tractable. For this reason, the Cartography module is particularly powerful when used in conjunction with Mosaic module (see Chapter 8), which is designed to efficiently process and combine extremely large data sets.

We will begin this chapter with a quick summary of the third party libraries that are needed to compile the Vision Workbench. We will then describe the `GeoRefence` class, which creates a relationship between pixel coordinates in an image and coordinates on a globe. Next, we will discuss the `GeoTransform` class, which provides a simple means of re-projecting map data. We finish this section with methods for reading and writing image files with embedded geospatial metadata.

10.1 Software Dependencies

The Cartography module is currently built on top of two third party libraries:

- GDAL [`http://www.remotesensing.org/gdal/`]
- Proj.4 [`http://proj.maptools.org/`]

In order to enable the Cartography module, you must have these libraries installed on your system before you configure and build the Vision Workbench. You may need to use the `PKG_PATHS`

directive in the `config.options` file if you install them in a non-standard location as discussed in Chapter 2.

Once the library for the Cartography module has been built, the header files for GDAL and Proj.4 are no longer needed, so you can rely solely on linking in libraries when building your own application.

10.2 The `GeoReference` Class

When you point at a location on a map, you probably want to know where that location can be found in the real world. This relationship depends first and foremost on the familiar notion of a map's scale. However, this relationship is also affected by a subtle, but extremely important dependence on how the map is *projected*. That is, the image depicts a scene that sits on the surface of a spheroid. However, the image is flat, so at best it represents a very slightly distorted view of the surface.

One can imagine all sorts of different ways that the surface can be warped or projected onto a flat plane (or, at the very least, projecting onto a manifold that can be unfolded into a plane without distorting distances and areas – a sphere cannot be unfolded in this way). Generations of cartographers have struggled with this topological challenge, and as a result they have developed many different ways to "un-fold" the globe so that it can be represented as a flat image. Rather than attempt a description of these many techniques here, we suggest you look at this excellent web site describing all aspects of map projections.

```
http://www.progonos.com/furuti/MapProj/CartIndex/cartIndex.html
```

The Proj.4 manual is also recommended as a reference for the specific map projections supported by the Vision Workbench.

Now would be a good time to take a break from reading this section of the documentation to look over these references. When you return, we will dive into some code examples.

10.2.1 The Datum

A Vision Workbench `GeoReference` object is composed of three items:

- **The Projection**: As discussed above, this is the technique used to represent the round globe in a flat image.
- **The Affine Transform**: This is the geometric transformation between pixel coordinates in the image to coordinates in the map projection space.
- **The Datum**: Describes the approximate shape of the planetary body, as either a sphere or an ellipsoid.

10.2.2 The Affine Transform

Let's start by being explicit about the coordinate systems we will be working with. For images, we adopt the usual Vision Workbench coordinate system wherein the upper left corner of the image is the origin, the u coordinate increases as you move right along the columns of the image, and the v coordinate increases as you move down the rows.

Method	Description
`set_sinusoidal()`	Sinusoidal Projection
`set_mercator()`	Mercator Projection
`set_orthographic()`	Orthographic Projection
`set_stereographic()`	Stereographic Projection
`set_UTM()`	Universal Transverse Mercator (UTM) Projection (Earth only)

Table 10.1: Currently supported `GeoReference` map projections.

For a planetary body, the coordinate of a point on the surface is typically measured in latitude, longitude, and radius (ϕ, θ, r). Lines of latitude are perpendicular to the axis of rotation and are measured from the center line, the *equator* (+/-90 degrees). Lines of Longitude are vertical, passing through both the North and South poles of the planet. It is measured from a vertical arc on the surface called the *meridian*. We will generally adopt an East positive frame of reference (latitude increases to the east of the meridian 0-360 degrees). Finally, the radius is measured from the point to the planet's center of mass. Note that this coordinate system is similar but not identical to spherical coordinates in a mathematical sense, where "latitude" would be measured from the North pole rather than the equator.

Under this set of assumptions, if we have a point $P_{img} = (u, v)$ in the image, and we want to relate it to some planetary coordinates

10.2.3 Putting Things Together

10.3 Geospatial Image Processing

10.3.1 The `GeoTransform` Functor

10.4 Georeferenced File I/O

10.4.1 `DiskImageResourceGDAL`

Chapter 11

Interest Point Module

[*Note: Parts of this documentation may no longer applicable...*]

Interest points are unique point identifiers with in an image. There are usually many of them in an image and what defines them changes based on the algorithm used. For the most part though, interest points are defined at places where a corner exists, where at least 2 edges come together.

Interest points are helpful in locating the same feature in multiple images. There are 3 major parts to using interest points. First, all images are processed with some interest point / corner detection and all points are recorded. Secondly, all of the points found from the first step now have a unique identifier built for them that properly describes the feature and it texture surroundings. Lastily, captured interest points are compared across images. Features on separate images that have identifiers that seem approximately equal are then connected into a match file.

Uses for interest points are general corner detection, object recognition, and image alignment. The last one, image alignment, is a common use used within Vision Workbench. Using fitting functors that will be described later, transform matrices can be solved for to describe the relationship between images and later merge them (see `<vw/Image/TransformView.h>`). Also as a side note, interest points can be used for measurements for a bundle adjustment that would solve for the original placement of the cameras (see `<vw/Camera/BundleAdjust.h>`).

The interest point module includes a complete set of classes and functions for each step of interest point detection. They can be imported into your code by including `<vw/InterestPoint.h>`. The built-in classes `ScaledInterestPointDetector` and `SimpleInterestPointDetector` (defined in `<vw/InterestPoint/Detector.h>`) provide out-of-the-box support for detecting interest points, with or without scale space methods.

The interest point module is designed to be as flexible as possible in that it decouples each step in the process of interest point detection. Different interest measures and thresholding methods, built-in or user-defined, can be used with the `InterestPointDetector` classes. `ScaledInterestPointDetector` and `SimpleInterestPointDetector` can both easily be subclassed to further customize their operation, for example by implementing an alternative method for finding peaks in the interest image.

The interest point module also provides tools for generating descriptors to compactly describe the properties of an interest point.

11.1 Scale Space Methods

When detecting interest points, we want them to be invariant to changes in view perspective. The scale space is a standard tool for making a detection algorithm scale invariant [2].

The interest point module provides support for scale space detection methods based on the `ImageOctave` class. An octave is a subset of the scale space. It is a set of images (scales) formed by convolving the source image with Gaussian kernels having progressively larger standard deviations; the sigma used for the last scale in the octave is twice that used for the first scale. Given a source image and a number of scales per octave, `ImageOctave` will construct the first octave of the scale space of the source image. Successive octaves can be constructed with the `build_next()` method.

```
ImageView<double> source;
int scales_per_octave;
ImageOctave octave(source, scales_per_octave);
// Process first octave...
// Then build the second octave
octave.build_next();
```

Building the next octave is a destructive operation, as the previously computed octave data is not saved. If you need to retain all of the scaled images generated, e.g. for use in generating descriptors, `ImageOctaveHistory` can be used to store this data.

11.2 Measuring Interest

The interest point module includes both classes and free functions for computing interest images from source images using the standard Harris [1] and LoG (Laplacian of Gaussian) interest measures. They can be imported by including the file `vw/InterestPoint/Interest.h`.

The HarrisInterest and LoGInterest classes are intended for use in conjunction with the InterestPointDetector classes (next section). Creating your own interest measure classes for use with the Detector classes is straightforward. Subclass the `InterestBase` abstract base type. In the constructor, set `InterestBase<T>::type` to `IP_MIN`, `IP_MAX` or `IP_MINMAX`, depending on what type of peaks in the generated interest image represent interest points. Then overload the abstract virtual method `compute_interest` with your implementation of the interest measure.

11.3 The Interest Point Detector Classes

The InterestPointDetector classes in `<vw/InterestPoint/Detector.h>` form the heart of the interest point module. They integrate the various components of the module into an easy all-in-one interface for detecting interest points.

The InterestPointDetector class itself is an abstract base class. Two implementations of its interface are supplied, `ScaledInterestPointDetector` and `SimpleInterestPointDetector`. The Scaled version uses scale space methods, while the Simple version does not; otherwise they are identical. When constructing either type of detector, you specify an interest measure class and a thresholding class. Built-in thresholding classes are defined in `<vw/InterestPoint/Threshold.h>`.

```
LoGInterest<float> log;
InterestThreshold<float> thresholder(0.0001);
```

```
ScaledInterestPointDetector<float> detector(&log, &thresholder);
std::vector<InterestPoint> points = interest_points(src, &detector);
```

11.4 Flow of Data

Although designed primarily for flexibility, the interest point module takes care not to sacrifice efficiency by unnecessarily recomputing internal images such as gradients. If you take advantage of the module's flexibility by customizing its framework (for example, by implementing a new interest measure class), you will probably make use of `ImageInterestData`, a struct which holds a source image and several interesting related images, such as gradients and interest.

11.5 Generating Descriptors

A descriptor of an interest point represents the local image region around the point. It should be distinctive as well as invariant to factors such as illumination and viewpoint. The interest module contains basic functions and classes for generating descriptors in `<vw/InterestPoint/Descriptor.h>`.

Generating a descriptor for an interest point requires knowledge of the point's source image. Different descriptor classes may require different source data. The trivial `PatchDescriptor` uses only the source `ImageView` as its source data.

```
SimpleInterestPointDetector<float> detector(&harris, &thresholder);
std::vector<InterestPoint> points = interest_points(source_image,
                                                    &detector);
PatchDescriptor<float> pd;
generate_descriptors(points, source_image, pd);
```

Properly generating descriptors for interest points found with `ScaledInterestPointDetector` is more involved, as various blurred versions of the source image may be required to provide local image regions for interest points at different scales.

```
ScaledInterestPointDetector<float> detector(&log, &thresholder);
ImageOctaveHistory<ImageInterestData<float> > history;
detector.record_history(&history);
SIFT_Descriptor<float> sd;
generate_descriptors(points, history, sd);
```

11.6 Matching

Aliens have abducted this section. Are you man enough to save it? Huh, Punk?

11.7 RANSAC

RANdom SAmple Consensus *(or RANSAC)* is a method for sifting through messy data to remove outliers. RANSAC starts with the goal of fitting some objective to a mass of data. In the case of interest points it is usually fitting some transform matrix to represent to move of points from

one image's coordinate frame to another. The algorithm works by randomly selecting a minimal number of matches and fitting an initial transform to this small selected set. It then proceeds to to grow the initial set of matches from with matches from the original pool whose error stays within an inlier threshold. This process of randomly selecting a minimal set, fitting, and growing is repeated many times. The round that produced the most inliers is kept for a final stage where a better fitting algorithm can be applied to the entire final pool of matches. The transform solved on the last step is considered best solution that correctly maps the inliers. This shotgun method, though not efficient, gives the ability for coping with a large percentage of outliers. *Yet, also be aware that it is entirely possible that in a worst case scenario, RANSAC might fit itself to an interesting bunch of outliers.*

Vision Workbench's implementation can be found in `<vw/Math/RANSAC.h>`. `RandomSampleConsensus` expects 3 inputs during it's construction. It requires a fitting functor that describes the type of transform matrix used for fitting. It also needs an error metric functor. For the case of interest points, `InterestPointErrorMetric()` should do the job. Finally, an integer describing the inlier threshold is required that defines the greatest error allowed during fitting. An example of construction is below.

```
vw::math::RandomSampleConsensus<math::SimilarityFittingFunctor,
                    math::InterestPointErrorMetric > ransac(
                    vw::math::SimilarityFittingFunctor(),
                    vw::math::InterestPointErrorMetric(), inlier_threshold );
```

`RandomSampleConsensus` is operated via an overloaded `operator()`. It expects a container of Vector3s. Interest Point module provides a helpful tool for converting to `Vector3` from a `std::vector` of InterestPoints called `iplist_to_vectorlist`. Finally the overloaded `operator()` returns it's final transform matrix result that was used to select it's inliers, this can be stored for later image transform operations if desired. `RandomSampleConsensus` does not return a new list of inliers, instead it returns the index locations of the inliers. It is left up to the user to repackage the interest points to have only inliers. An example of operations is below.

```
std::vector<Vector3> ransac_ip1 = iplist_to_vectorlist(matched_ip1);
std::vector<Vector3> ransac_ip2 = iplist_to_vectorlist(matched_ip2);

Matrix<double> H(ransac(ransac_ip1,ransac_ip2));
std::vector<int> indices = ransac.inlier_indices(H,ransac_ip1,ransac_ip2);
```

Lastly, below is a listing of fitting functors that are available in `<vw/Math/Geometry.h>`.

Functor	Description
`HomographyFittingFunctor()`	8 DOF. Also known as Projective matrix.
`AffineFittingFunctor()`	6 DOF. Handles rotation, translation, scaling, and skewing.
`SimilarityFittingFunctor()`	4 DOF. Handles rotation, translation, and scaling.
`TranslationRotationFittingFunctor()`	3 DOF. Also known as Euclidean matrix.

Table 11.1: Fitting functors defined in `<vw/Math/Geometry.h>`.

11.8 Pre-built Tools

To further help the introduction to the use of interest points, VisionWorkbench supplies two utility programs for working with interest points. `Ipfind` and `Ipmatch` can be found built in VisionWorkbench's build path and the source code is available in `<vw/tools/>` .

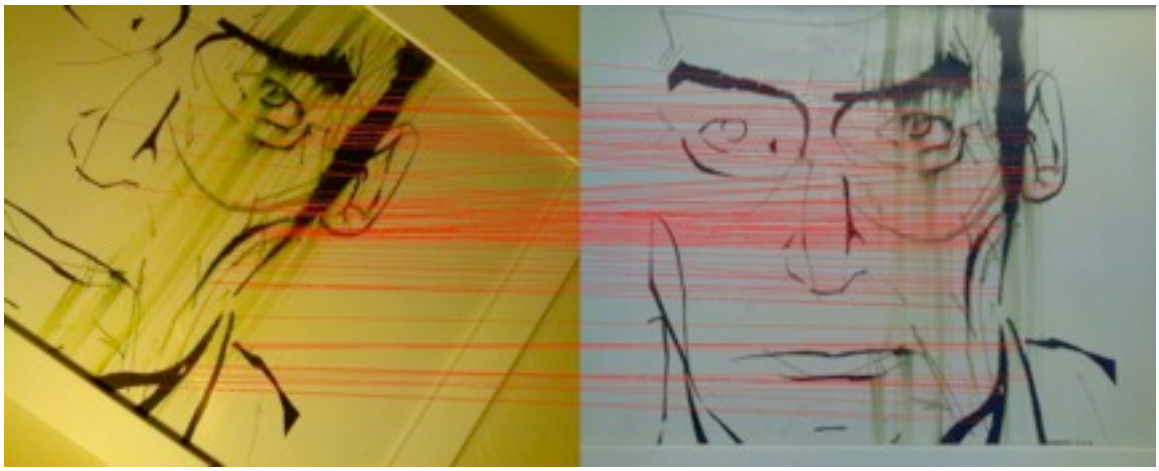


Figure 11.1: Example debug image from `Ipmatch`.

The above is an example of a result that can be created with `Ipfind` and `Ipmatch`. Here are the commands used to create it.

```
ipfind left.png right.png
ipmatch left.png right.png -r homography -d
```

Bibliography

[1] Harris, Chris, and Mike Stephens, "A Combined Corner and Edge Detector," Proc. 4th Alvey Vision Conf., Manchester, pp. 147-151, 1988.

[2] Lindeberg, Tony, "Feature Detection with Automatic Scale Selection," Int. J. of Computer Vision, Vol. 30, number 2, 1998.

[3] Lowe, David G., "Distinctive Image Features from Scale-Invariant Keypoints," Int. J. of Computer Vision, 2004.

Chapter 12

Stereo Module

Stereo module's focus is finding a dense correlation between two images taken of the same object from different locations. Dense correlation means that every pixel in the first image has been correctly matched to a pixel in the second image.

This dense matching is useful for recreating an ability every human has and that is stereo vision. Using Camera Models discussed in Camera Module and using a dense correlation solved for in this module, a 3D measurement can be created for every pixel. This has use in rover navigation, as it allows for the land ahead to measured for steepness and roughness of terrain. It can also be used for the construction topographical maps created from satellite imagery that allows for measurement of every mountain, canyon and crater.

The following chapter is order mostly in the fashion that 3D models are created. We'll start with a discussion of the storage format for a dense correlation called a disparity map. Then we'll cover solving for a rough integer estimate of a disparity map in Stereo Correlation. Next is finding the floating point solution of that disparity map call Subpixel Refinement. Finally we'll discuss the code that converts all of this data into 3D measurement with StereoView. We'll also point you to a demo tool that uses the discussed code.

12.1 Disparity Maps

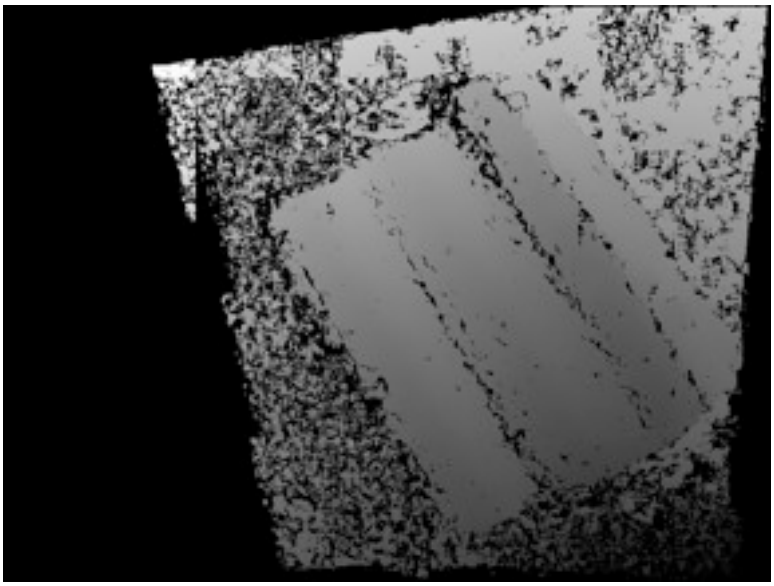
As mentioned before a disparity map is the data structure that we use to store the solution of the dense correlation. Simply it's just an `ImageView<PixelMask<Vector2f> >`. Disparity Maps are always associated with a left and a right image. Given a $I_L(x, y)$ in the left image, it's correspond index in the right image is found with $Il_L x, y) + D(I_L(x, y)) = I_R(x, y)$, where D is the disparity map. Pixel Mask is used to keep track of failures that is usually caused by an unstable solution or saturation of the input image (like a shadow or highlights). Figure 12.1 is an example of what a typical disparity map looks like created from images that were aligned before feeding to the `correlate` tool.

The header file `<vw/Stereo/DisparityMap.h>` provides a collection of useful tools for working with and modifying disparities. Those functions are listed in Table 12.1. Tasks that are not provided in the table are often possible with the generic routines for working with `PixelMasks`. For those abilities take look back at Table 4.11.

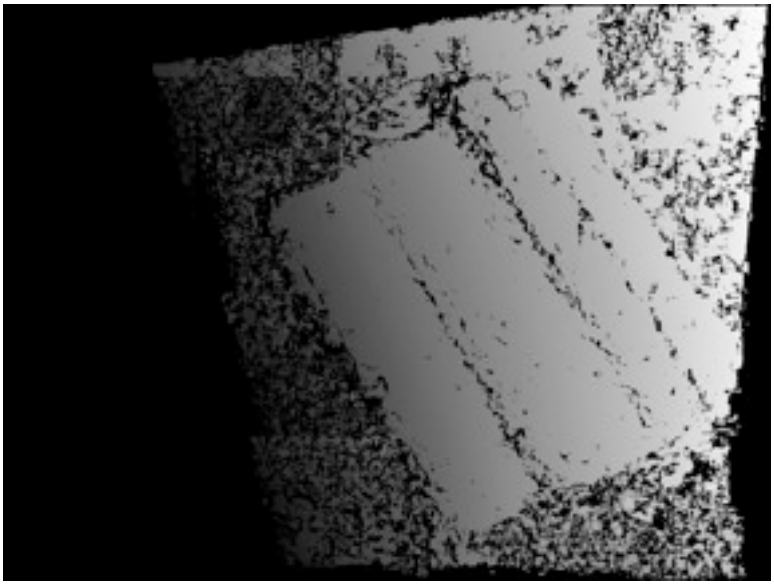
Of the above functions, `transform_disparities` is of particular interest. Before attempting to correlate to images it is best to do a rough align for the images (like making sure both images are upright). A transform can be solved for with interest points and then applied to the right image. After solving for the disparity map between the left and the transform right image,

(a) `Left Image`

(b) `Right Image`



(c) `Disparity Ch:0` *Horizontal offset*



(d) `Disparity Ch:1` *Vertical offset*

Figure 12.1: Example of a Disparity Map created from pictures of a Sun Chip.

Function	Description
`get_disparity_range(img)`	Returns a bounding box with a min & max describing the value ranges of the disparity map
`missing_pixel_image(img)`	Creates a new RGB image describing showing good pixel locations
`disparity_mask(disparity, Lmask, Rmask)`	Creates a new disparity map this valid only where the disparity links are valid in both the Left and Right images. `Lmask` & `Rmask` are uint8 images where 255 is valid.
`disparity_range_make(disparity, min, max)`	Invalidates pixel locations in a disparity map where the values exceed the values of `min` & `max`.
`remove_outliers(disparity, half_h_kern, half_v_kern, p_threshold, rej_threshold)`	An erosion like method to take out high frequency changes in the disparity map and label them as invalid.
`clean_up(disparity, half_h_kern, half_v_kern)`	Applies `remove_outliers` twice with second application targeting single pixel outliers
`std_dev_image(disparity, kern_w, kern_h)`	Remove pixels from the disparity map that correspond to low contrast pixels
`transform_disparities(disparity, trans)`	Applies a transform `tans` to a disparity map

Table 12.1: Functions provided for working with Disparity Maps.

`transform_disparities` can be applied to the disparity map with the inverse transform. Making the disparity map correlate between the original left and right images. This step of aligning the images before processing is sometimes called rectification and it can play a big part in the performance of stereo processing.

12.2 Stereo Correlation

In this section we'll cover the techniques and the code to solve for a disparity map with integer precision. First we'll start out with the general idea of correlation. Then we'll break out in sections with ways on how to speed of the process.

Correlation again is solving for a given point in the left image's corresponding point in the right image. This is performed with template matching.

Template matching is basically taking a crop of the image (called a kernel) around a point in the left image and then trying to find a similar point in the right image. A search range is defined for the right image and a kernel is slid across the entire range. To determine if a kernel position in the right image corresponds with the kernel in the left image, a cost function is used. Normally we work with the sum of absolute differences between the kernels, meaning that a low value is high correlation and a high value means those kernels just don't match. In our code we call this simply absolute difference. There are other cost functions as well, squared difference and normalized cross correlation. Yet all of them still boil down to per pixel difference being calculated between the 2 kernels.

Stereo Module does the task of template matching most plainly in the `ReferenceCorrelator` found in `<vw/Stereo/ReferenceCorrelator.h>`. Yet this is not recommended for general use, instead an better interface is provided called `CorrelatorView` found in `<vw/Stereo/CorrelatorView.h>`.

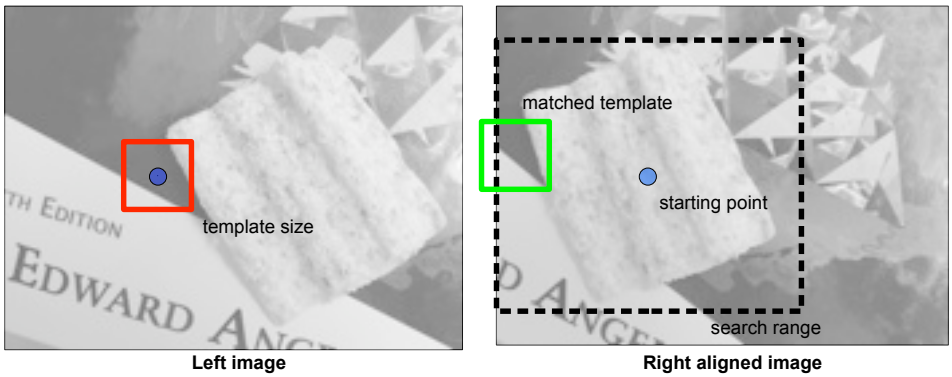


Figure 12.2: Template matching example

Here's an example for implementing it.

```
CorrelatorView<PixelGray<float>,vw::uint8,SlogStereoPreprocessingFilter>
  corr_view( left_disk_image, right_disk_image, left_mask, right_mask,
             SlogStereoPreprocessingFilter(1.5) );
corr_view.set_search_range(search_range);
corr_view.set_kernel_size(Vector2i(25,25));
corr_view.set_correlator_options(1, ABS_DIFF_CORRELATOR);

// Begin processing
ImageView<PixelMask<Vector2f> > disparity_map = corr_view;
```

That may seem like a lot of properties to set but don't let it frighten you away. The original inputs are the 2 input images, 2 mask images, and a preprocessing filter. Mask images are simply vw::uint8 images that are white where an image is valid. This is useful for masking off objects that are known to cause problems before hand like shadows, dust, or a certain someone who was accidentally touching the lens during the shot. The preprocessing filter options are defined in `<vw/Stereo/Correlate.h>`. The LoG filters are useful for making the correlator light invariant as it only see edges after it is applied. LoG filters are also a recommended filter for when trying out. Blur filter is recommended for only when working the normalized cross correlation cost function in template matching.

Function	Description
`SlogStereoPreprocessingFilter(arg)`	Sign of Laplacian of the Gaussian filter. Allows for efficient XOR comparisons. `arg` is the gaussian sigma size.
`LogStereoPreprocessingFilter(arg)`	Laplacian of Gaussian filter. `arg` is the gaussian sigma size.
`BlurStereoPreprocessingFilter(arg)`	Gaussian blur. `arg` is the gaussian sigma size.
`NullStereoPreprocessingFilter()`	No preprocessing.

Table 12.2: Built-in preprocessing filter options.

One the other lines we set the search range with a bounding box. Where origin of the bounding box is the starting search location in the right image. Kernel size is the same as the template

window's size. Finally the correlator options sets kernel size for an additional option blur that is applied internally and it sets the cost function used in template matching. Table 12.3 has the available options for cost functions.

Enumerator	Description
`ABS_DIFF_CORRELATOR`	Use the sum of absolute differences between paired kernels
`SQR_DIFF_CORRELATOR`	Use the sum of squared differences between paired kernels
`NORM_XCORR_CORRELATOR`	Normalized cross correlation cost function.

Table 12.3: Correlator cost function options.

This process of template matching is repeated for every pixel in the left image against a search range in the right image. It should be obvious why this is a CPU intensive process. Yet there are methods for doing template matching efficiently and they are using a box filter and by implementing a pyramid method to processing. The following subsections cover those improvements, but it's not required reading as `CorrelatorView` implements them on default.

12.2.1 Optimized Correlator

The OptimizedCorrelator employs a boxfilter-like accumulator to significantly speed up the template matching process by cutting out redundant calculations (See figure 12.3).

Its output should be identical to the ReferenceCorrelator, which does not attempt to optimize the template matching process in any way. (This makes it so slow it should only be used for testing purposes)

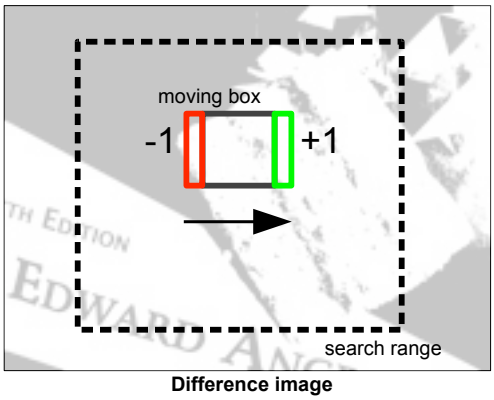


Figure 12.3: Box filter Optimization

12.2.2 Pyramid-based Search Refinement

Pyramid-based searches are pretty straight forward. Our first step is producing a gaussian pyramid of the input images. This means repeatedly blurring the image a small amount and then subsampling by 2. The highest level of the pyramid is the smallest image as it has been subsampled the most. The top level is where we start our template matching.

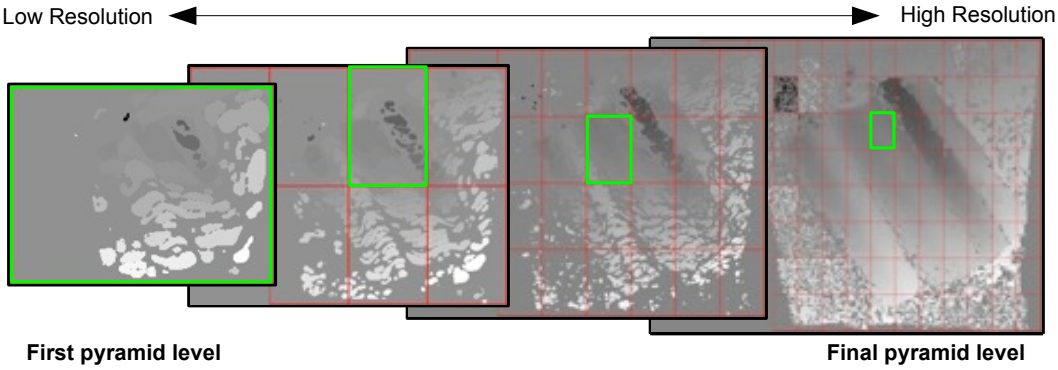


Figure 12.4: Pyramid-based search refinement

From the top level our search range for the kernel has been reduced greatly so it's processing time is very short. The disparity map created with the reduced imagery is then used as a seed for the next lower level of the pyramid. This is wonderful as the previous results from a higher level will have the template window searching roughly in the correct area and that means the search region is kept relatively small. Using the pyramid search, each level is used as a seed for the next lowest level until the bottom is reached. The result for the lowest level is the disparity map for the original resolution imagery.

12.3 Subpixel Refinement

After disparity image is created with `CorrelatorView`, it is very coarse and exhibits what looks like stair steps in the 3D model. This is because CorrelatorView has solved for the disparity using only whole numbers. Subpixel Refinement further polishes the disparity map into floating point precision. This additional step can be accessed can be found in `<vw/Stereo/SubpixelView.h>`.

```
disparity_map = SubpixelView<LogStereoPreprocessingFilter>(disparity_map_integer,
        left_image, right_image,
        25, 25,          // Kernel size
        true, true,      // Do H V subpixel
        0,               // Subpixel mode
        LogStereoPreprocessingFilter(1.5),
        false);          // Verbose option
```

Above is an example of code using `SubpixelView`. The first few arguments should be pretty straight forward. It's the disparity map created from `CorrelatorView` and then the left and right original images. The next two numbers are the kernel size to be used, these do not need to be the same as what was used in integer correlation. After that are two boolean conditionals that turn on horizontal and vertical subpixel. Normally both of those values should be true. Yet for some applications where speed is important, subpixel is only turned on for one direction.

The last three are self explanatory except Subpixel mode. Stereo module currently has 4 different sub pixel refinement algorithms and with each increment they have better quality but at the cost of speed. Mode 0 is parabola fitting and the fastest. It discussed in the next subsection. Modes 1-3 are forms of affine kernel algorithms and are discussed in the subsection 12.3.2. Of these modes, it

is suggested starting out with Mode 0 and then when time is not important jump directly to using Mode 3.

Mode	Name	Description
0	`No Subpixel`	Disable subpixel correlation
1	`Parabola Subpixel`	Simplest and fastest subpixel mode.
2	`Bayes EM Affine Subpixel`	Affine with estimated parameters for gaussian weighted window
3	`Experimental Subpixel`	Affine with estimated parameters for gamma and gaussian weighted windows that correspond to inlier and outlier models

Table 12.4: Available Subpixel modes.

12.3.1 Parabola Fitting

The parabola subpixel mode estimates the subpixel disparity offset by fitting a paraboloid to the template matching scores of the 8 pixels around the integer disparity estimate. The location of the minimum of this surface is used as the subpixel offset.

Although this is the fastest subpixel mode, it exhibits an artifact known as "pixel-locking": the subpixel offsets tend to fall near the integer disparity estimates. This causes stair-step like patterns on surfaces that should otherwise be smooth. To avoid this problem, use the affine-adaptive subpixel mode (Section 12.3.2).

12.3.2 Affine-adaptive Subpixel Refinement

Text

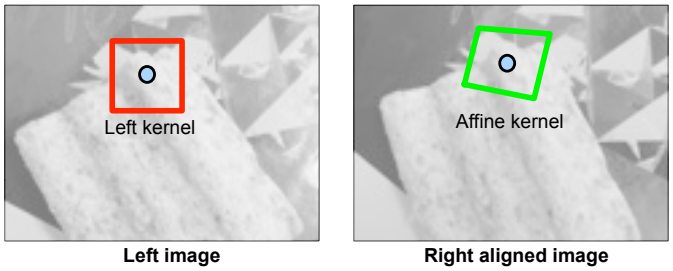


Figure 12.5: Affine refinement

Text

12.4 Point Clouds

12.5 Command Line Tool

Vision Workbench only comes with one tool that uses the Stereo Module. That is `correlate` which can be read up in the Tools Chapter in Section 13.2. That tool is only a demo of things possible

and was used to create the imagery in this chapter. For a more extreme example of what can be done with the stereo module, we encourage you to check out the Ames Stereo Pipeline.

Bibliography

Chapter 13

Tools

VisionWorkbench has various tools that are handy for working with graphics. The tools can be found in the `VisionWorkbench/src/vw/tools` folder.

13.1 `colormap`

The `colormap` tool visualizes the data encoded in a DEM by reading in a DEM and outputting a corresponding color-coded height image. The colors run from blue for the lowest height value to red for the highest height value. To adjust the image colors, use flags `--min` and `--max` to set the highest and lowest values on the blue-to-red scale. For example, suppose `firstDEM.tif` has heights in [0,100] and `secondDEM.tif` has heights in [0,120]. To visualize `firstDEM.tif` using the same color scale as `secondDEM.tif`, run `colormap --max 120 firstDEM.tif`.

Other command-line options for colormap are listed in Table 13.1.

Table 13.1: Command-line options for colormap

Option	Description
`--help`	Display a help message
`--input-file arg`	Explicitly specify the input file
`-s [ --shaded-relief-file ] arg`	Specify a shaded relief image (grayscale) to apply to the colorized image
`-o [ --output-file ] arg`	Specify the output file
`--nodata-value arg`	Remap the DEM default value to the min altitude value
`--min arg`	Minimum height of the color map
`--max arg`	Maximum height of the color map
`--verbose`	Verbose output

13.2 `correlate`

The `correlate` tools is mostly a demo for the Stereo Module (Chapter 12) yet that does not necessarily make it boring. It allows for playing around with settings provided in the `CorrelateView`. The user must specify 2 images taken of the same object (preferably aligned prior with a match file). This will then produce 2 output images that represent the horizontal and vertical disparities between the images. Again the Stereo Module demos results that can be created with `correlate`

in Figure 12.1.

All commands to `correlate` must be explict (as there are no positional inputs). An example of using the tool is given below.

```
correlate --slog 1.5 --left DSC00623.JPG --right DSC00625.JPG
  --xkernel 25 --ykernel 25 -xrange 100 -yrange 100 --pyramid
```

Table 13.2: Command-line options for correlate

Option	Description
`--help`	Display this help message
`--left arg`	Explicitly specify the 'left' input file
`--right arg`	Explicitly specify the 'right' input file
`--slog arg (=1)`	Apply SLoG filter with the given sigma, or 0 to disable
`--log arg (=0)`	Apply LoG filter with the given sigma, or 0 to disable
`--xoffset arg (=0)`	Overall horizontal offset between images
`--yoffset arg (=0)`	Overall vertical offset between images
`--xrange arg (=5)`	Allowed range of horizontal disparity
`--yrange arg (=5)`	Allowed range of vertical disparity
`--xkernel arg (=5)`	Horizontal correlation kernel size
`--ykernel arg (=5)`	Vertical correlation kernel size
`--lrthresh arg (=1)`	Left/right correspondence threshold
`--csthresh arg (=1)`	Correlation score rejection threshold (1.0 is Off – 2.0 is Aggressive outlier rejection
`--cost-blur arg (=1)`	Kernel size for blurring the cost image
`--correlator-type arg (=0)`	0 - Abs diff; 1 = Sq Diff; 2 - NormXCorr
`--hsubpix`	Enable horizontal sub-pixel correlation
`--vsubpix`	Enable vertical sub-pixel correlation
`--affine-subpix`	Enable affine adaptive sub-pixel correlation (slower, but more accurate)
`--reference`	Use the slower, simpler reference correlator
`--pyramid`	Use the pyramid based correlator
`--bitimage`	Force the use of the optimized bit-image correlator
`--nonbitimage`	Force the use of the slower, non bit-image optimized correlator

13.3 `hillshade`

An alternative visualization of the DEM data may be desired. For this, the `hillshade` tool reads in a DEM and outputs an image of that DEM as though it were a three-dimensional surface, with every pixel shaded as though it were illuminated by a light from a specified location.

Table 13.3: Command-line options for hillshade

Option	Description
`--help`	Display a help message

--input-file arg	Explicitly specify the input file
-o [--output-file] arg	Specify the output file
-a [--azimuth] arg (=0)	Sets the direction tha the light source is coming from (in degrees). Zero degrees is to the right, with positive degree counter-clockwise.
-e [--elevation] arg (=45)	Set the elevation of the light source (in degrees)
-s [--scale] arg (=0)	Set the scale of a pixel (in the same units as the DTM height values
--nodata-value arg	Remap the DEM default value to the min altitude value
--blur arg	Pre-blur the DEM with the specified sigma

13.4 geoblend

`geoblend` merges multiple DEMs into one large DEM. By default, it blends the DEMs so that the output is smooth at the boundaries. Disable this feature using the `--draft` flag. Other options are listed in Table 13.4.

Table 13.4: Command-line options for geoblend

Option	Description
--help	Display a help message
-o [--mosaic-name] arg (=mosaic)	Specify base output directory
-t [--output-file-type] arg (=tif)	Output file type
--tile-output	Output the leaf tiles of a quadtree, instead of a single blended image.
--tiled-tiff arg (=0)	Output a tiled TIFF image, with given tile size (0 disables, TIFF only)
--patch-size arg (=256)	Patch size for tiled output, in pixels
--patch-overlap arg (=0)	Patch overlap for tiled output, in pixels
--cache arg (=1024)	Cache size, in megabytes
--draft	Draft mode (no blending)
--ignore-alpha	Ignore the alpha channel of the input images, and don't write an alpha channel in output.
--nodata-value arg	Pixel value to use for nodata in input and output (when there's no alpha channel)
--channel-type arg	Images' channel type. One of [uint8, uint16, int16, float].
--verbose	Verbose output

13.5 georef

The `georef` tool lets you specify the geographical coordinates and projection method for an image taken of the surface of a planet.

Table 13.5: Command-line options for georef

Option	Description
General Options	
`-o [ --output-file ] arg (=output.tif)`	Specify the base output filename
`-q [ --quiet ]`	Quiet output
`-v [ --verbose ]`	Verbose output
`--cache arg (=1024)`	Cache size, in megabytes
`--help`	Display a help message
Projection Options	
`--copy arg`	Copy the projection from the given file
`--tfw arg`	Create a .tfw sidecar file with the given filename rather than a full copy of the image file
`--north arg`	The northernmost latitude in degrees
`--south arg`	The southernmost latitude in degrees
`--east arg`	The easternmost longitude in degrees
`--west arg`	The westernmost longitude in degrees
`--sinusoidal`	Assume a sinusoidal projection
`--mercator`	Assume a Mercator projection
`--transverse-mercator`	Assume a transverse Mercator projection
`--orthographic`	Assume an orthographic projection
`--stereographic`	Assume a stereographic projection
`--lambert-azimuthal`	Assume a Lambert azimuthal projection
`--utm arg`	Assume UTM projection with the given zone
`--proj-lat arg`	The center of projection latitude (if applicable)
`--proj-lon arg`	The center of projection longitude (if applicable)
`--proj-scale arg`	The projection scale (if applicable)
`--nudge-x arg`	Nudge the image, in projected coordinates
`--nudge-y arg`	Nudge the image, in projected coordinates
`--pixel-as-point`	Encode that the pixel location (0,0) is the center of the upper left hand pixel (the default, if you specify nothing, is to set the upper left hand corner of the upper left pixel as (0,0) (i.e. PixelAsArea).

13.6 `image2qtree`

`image2qtree` turns a georeferenced image (or images) into a quadtree with geographical metadata. For example, it can output a kml file for viewing in Google Earth.

Table 13.6: Command-line options for image2qtree

Option	Description
General Options	
`-o [ --output-name ] arg`	Specify the base output directory

Continued on next page

Table 13.6 – continued from previous page

-q [--quiet]	Quiet output
-v [--verbose]	Verbose output
--cache arg (=1024)	Cache size, in megabytes
--help	Display a help message
Input Options	
--force-wgs84	Use WGS84 as the input images' geographic coordinate systems, even if they're not (old behavior)
--pixel-scale arg (=1)	Scale factor to apply to pixels
--pixel-offset arg (=0)	Offset to apply to pixels
--normalize	Normalize input images so that their full dynamic range falls in between [0,255]
Output Options	
-m [--output-metadata] arg (=none)	Specify the output metadata type. One of [kml, tms, uniview, gmap, celestia, none]
--file-type arg (=png)	Output file type
--channel-type arg (=uint8)	Output (and input) channel type. One of [uint8, uint16, int16, float]
--module-name arg (=marsds)	The module where the output will be placed. Ex: marsds for Uniview, or Sol/Mars for Celestia
--terrain	Outputs image files suitable for a Uniview terrain view. Implies output format as PNG, channel type uint16. Uniview only
--jpeg-quality arg (=0.75)	JPEG quality factor (0.0 to 1.0)
--png-compression arg (=3)	PNG compression level (0 to 9)
--palette-file arg	Apply a palette from the given file
--palette-scale arg	Apply a scale factor before applying the palette
--palette-offset arg	Apply an offset before applying the palette
--tile-size arg (=256)	Tile size, in pixels
--max-lod-pixels arg (=1024)	Max LoD in pixels, or -1 for none (kml only)
--draw-order-offset arg (=0)	Offset for the ¡drawOrder¿ tag for this overlay (kml only)
--composite-multiband	Composite images using multi-band blending
--aspect-ratio arg (=1)	Pixel aspect ratio (for polar overlays; should be a power of two)
Projection Options	
--north arg	The northernmost latitude in degrees
--south arg	The southernmost latitude in degrees
--east arg	The easternmost longitude in degrees
--west arg	The westernmost longitude in degrees
--force-wgs84	Assume the input images' geographic coordinate systems are WGS84, even if they're not (old behavior)
--sinusoidal	Assume a sinusoidal projection
--mercator	Assume a Mercator projection

Continued on next page

Table 13.6 – continued from previous page

`--transverse-mercator`	Assume a transverse Mercator projection
`--orthographic`	Assume an orthographic projection
`--stereographic`	Assume a stereographic projection
`--lambert-azimuthal`	Assume a Lambert azimuthal projection
`--lambert-conformal-conic`	Assume a Lambert Conformal Conic projection
`--utm arg`	Assume UTM projection with the given zone
`--proj-lat arg`	The center of projection latitude (if applicable)
`--proj-lon arg`	The center of projection longitude (if applicable)
`--proj-scale arg`	The projection scale (if applicable)
`--std-parallel1 arg`	Standard parallels for Lambert Conformal Conic projection
`--std-parallel2 arg`	Standard parallels for Lambert Conformal Conic projection
`--nudge-x arg`	Nudge the image, in projected coordinates
`--nudge-y arg`	Nudge the image, in projected coordinates

13.7 `ipfind`

The `ipfind` tool processes images for interest points. Interest points are features with in an image that can be reliably located across multiple images. `ipfind` is a tool that is best paired with another tool, `ipmatch`, described in the next Section 13.8.

The `--interest-operator` option allows for the user to change the algorithm used to detect interest points. `[LoG]` is the simplest and also the slowest, but is often pretty reliable. All method work to some extent by identify points where the image gradients all point inwardly as on a hill. They also identify the polar opposite, all gradients pointing away.

The `--descriptor-genetator` option allows for the user to change the algorithm use to describe an interest point. Interest points are described so that they may have a unique identity so that they maybe located in different images. `[patch]` is the simplist descriptor, as it is actually the image cropped around an interest point.

The result of `ipfind` is a special file called a *Vision Workbench interest point file*, and it has the extension `.vwip`. The *Interest Point Module* (Chapter 11) contains routines that can open and process such files.

Usage: `ipfind [options] <image files>...`. Other command line arguments are listed in Table 13.7

Table 13.7: Command-line options for ipfind

Option	Description
`--help`	Display this table
`--num-threads arg (=0)`	Set the number of theads for interest point detection. Setting the num_threads to zero causes ipfind to use the visionworkbench default number of threads.
`-t [ --tile-size ] arg (=2048)`	Specify the tile size for processing interest points. (Useful when working with large images).

-l [--lowe]	Save the interest points in an ASCII data format that is compatible with the Lowe-SIFT toolchain.
-d [--debug-image]	Write out debug images. This will highlight the found interest points.
--interest-operator arg (=LoG)	Choose an interest point metric from [LoG, Harris].
--log-threshold arg (=0.03)	Sets the threshold for the Laplacian of Gaussian interest operator.
--harris-threshold arg (=1e-5)	Sets the threshold for the Harris interest operator.
--max-points arg (=0)	Sets the maximum number of interest points you want returned. The most ïnterestingp̈oibts are selected.
--single-scale	Turn off scale-invariant interest point detection. This option only searches for interest points in the first octave of the scale space. This means faster operation at a cost of quality.
--descriptor-generator arg (=patch)	Choose a descriptor generator from [patch,pca].

13.8 ipmatch

The `ipmatch` tool processes images for interest points and matches them across images pair wise. It exactly one step longer than the `ipfind` tool. This tool will load up the images and their corresponding `.vwip` files and will run a matching algorithm looking for interest points with similair descriptors. If a corresponding `.vwip` file can not be found for an input image, `ipmatch` will actually call `ipfind` using the default settings to create a interest point file.

To filter out mismatched interest points, a RANSAC method is used. The fitting functor used to determine if a match is valid or not can be selected via the `--ransac-constraint` option. The options available are transform matrices with varing degrees of freedom.

The result of `ipmatch` is a special file called a *Vision Workbench match file*, and it has the extension `.vwip`. The *Interest Point Module* (Chapter 11) contains routines that can open and process such files.

Usage: `ipmatch [options] <input file1> <input file2>`. Other command line arguments are listed in Table 13.8

Table 13.8: Command-line options for ipmatch

Option	Description
--help	Display this table
-t [--matcher-threshold] arg (=0.8)	Threshold for the interest point matcher.
--non-kdtree	Use an implementation of the interest matcher that is not reliant o a KDTree algorithm.
-r [--ransac-constraint] arg (=similarity)	RANSAC constraint type. Choose one of: [similarity, homography, or none].
-i [--inlier-threshold] arg (=10)	RANSAC inlier threshold. Increase the number to allow more matches through that might agree with RANSAC.

-d [--debug-image]	Write out debug images. This will highlight all the matches found between two image.

13.9 slopemap

The slopemap tool takes in an image with geodetic coordinates and calculates the gradient angle (steepness of slope) and aspect (direction of slope) in radians of each point on a DEM. The gradient angle is calculated as the angle between the steepest slope and the horizon; the aspect is the angle clockwise between the normal of the surface and a vector pointing to north, both projected onto the tangent plane of the unelevated surface at that point.

Outputs include two georeferenced tifs encoding float values for aspect and gradient angle (by default), as well as one RGB image displaying a colorized representation of both gradient angle and aspect in which aspect is represented as hue and gradient angle is represented as a combination of saturation and value (needs flag). The output files are named according to the command line argument output-prefix, with corresponding suffixes of _aspect.tif, _gradient.tif, and _pretty.tif

The implementation calculates gradient angle and aspect by one of several methods: [horn, fh, sa, planefit]. planefit refers to fitting a plane by least squares to all nine points of a 3x3 window around the point in question, weighted equally. The implementation involves using singular value decomposition to solve a homogeneous linear system of equations and is slower than the other methods.

horn, fh, and sa are modified versions of finite difference methods. They correspond to modified versions of Horn's method, Fleming and Hoffer's method (also known as Ritter's method as well as the rook's case), and Sharpnack and Akin's method (also known as the queen's case). They are all variations of a similar method: approximate a west-east gradient and a south-north gradient and subsequently calculate gradient angle and aspect from that. The modification to these algorithms is that they do not require sampling from a square grid; when there is such a grid, the results are the same as the unmodified versions. There is also a flag, spherical, that determines whether the datum is interpreted to be spherical or flat (defaults to spherical), for an additional modification, as the original algorithms assumed a flat sampling grid.

Usage: slopemap [options] <input file>. Other command line arguments are listed in Table 13.9

Table 13.9: Command-line options for slopemap

Options	Description
--help	Display this help message
--input file arg	Explcitly specify the input file
-o [--output-prefix]	Output prefix
--no-aspect	Do not output aspect ([output-prefix]_aspect.tif)
--no-gradient	Do not output gradient ([output-prefix]_gradient.tif)
--pretty	Output colored image ([output-prefix]_pretty.tif)
--algorithm arg (=moduneven)	Choose an algorithm to calculate slope/aspect from [horn, fh, sa, planefit]. Horn: Horn's algorithm; FH: Fleming & Hoffer's (rook's case); SA: Sharpnack & Akin's (queen's case)

`--spherical arg (=1)`	Spherical/elliptical datum (recommended); otherwise, a flat grid

Chapter 14

Advanced Topics

Alas, this chapter is only partially written. Contact the authors for assistance with any of these topics.

14.1 Lazy Evaluation: Working with Views

In the Vision Workbench, the `ImageView<>` class gives you access to raw, reference-counted pixel data stored in RAM. However, this is just one instance of a more general *View* concept that is central to the inner workings of the Vision Workbench library. In this section, we will discuss the view concept in detail and show how you can leverage the full flexibility of VW by creating and optimizing your own *Views*.

14.1.1 The View Concept

Consider what could happen when you chain three image processing operation together.

```
result = image1 + image2 + image3;
result = transpose( crop( image, x, y, 31, 31 ) );
```

In order to perform this computation, the C++ compiler introduces a temporary image like this:

```
ImageView<PixelGray<uint8> > tmp = image1 + image2;
result = tmp + image3;
```

And this:

```
ImageView<PixelGray<uint8> > tmp = crop( image, x, y, 31, 31 );
result = transpose(tmp);
```

For simple operations such as addition, the introduction of a temporary image like this can be terribly inneficient because it causes a large, image-sized block of memory to be allocated and deallocated, and it requires a second pass over the data.

It is more efficient to allocate a single temporary `PixelGray<uint8>`; using it as you iterate over the image computing the result one pixel at a time. This is how a hand-coded implementation would be written by a cognisant programmer.

Getting dimensions	`img.cols()`
	`img.rows()`
	`img.planes()`

Accessing pixels	`img(col,row)` `img(col,row,plane)`

STL iterator	`ImageViewBase<...>::iterator`
	`img.begin()`
	`img.end()`

Pixel iterator	`ImageViewBase<...>::pixel_iterator` `img.origin()`

Table 14.1: In the Vision Workbench, a *View* is any object that supplies the above methods and inherits from ImageViewBase.

To avoid such enefficiencies, the Vision Workbench resorts to *lazy evaluation*; postponing operations on *Views* until the last possible moment before the result is needed. For example, applying the addition operator to a pair of images does not immediately compute their sum. Instead, *the + operator returns an image sum object.* This object is a *View*: an object that behave just like an image, but represents a processed view of the underlying data (see Table 14.1). The actual computation of the sum is postponed until later when it is needed; for example when the user assigns it to an `ImageView<>`, thereby causing the sum to be computed and stored in memory. This process of finally evaluating one view into another is called *rasterization*.

Nested function calls produce nested View types, e.g. a sum object containing a sum object plus an image, and so on. In this fashion, a series of image processing operations builds up a *View Tree* that can represent an extensive chain of operations. The tree of operations are collapsed into the inner loop when they are rasterized, allowing the compiler to efficiently allocate small tempories. Rasterization is also an opportunity to make decisions about how to optimize the image processing operations. For example, it may actually be *more* efficient in some cases to introduce an entire temporary image in between two operations. A convolution operation followed by another convolution operation is much more efficient when implemented in this manner. The Vision Workbench rasterization engine does its best to make efficient decisions about when and how to introduce temporaries when rasterizing a view tree. We'll have more to say on this subject in the sections below.

```
 /// An image view for performing image correlation
template <class ImageT, class PreProcFuncT>
class CorrelatorView : public ImageViewBase<CorrelatorView<ImageT, PreProcFuncT> > {

  typedef typename ImageT::pixel_type pixel_type

  ImageViewRef<pixel_type> m_left_image, m_right_image;
  PreProcFuncT m_preproc_func;

  // Settings
  BBox2i m_search_range;
  Vector2i m_kernel_size;

  public:
```

```
  typedef typename ImageT::pixel_type pixel_type;
  typedef pixel_type result_type;
  typedef ProceduralPixelAccessor<CorrelatorView> pixel_accessor;

CorrelatorView(ImageViewBase<ImageT> const& left_image, ImageViewBase<ImageT> const& ri
               PreProcFuncT const& preproc_func) :
  m_left_image(left_image.impl()), m_right_image(right_image.impl()),
  m_preproc_func(preproc_func) {

  // Set some sensible default values
  m_search_range = BBox2i(-50,-50,100,100);
  m_kernel_size = Vector2i(24,24);
}

// Standard ImageView interface methods
inline int32 cols() const { return m_left_image.cols(); }
inline int32 rows() const { return m_left_image.rows(); }
inline int32 planes() const { return 1; }

inline pixel_accessor origin() const { return pixel_accessor( *this, 0, 0 ); }

inline pixel_type operator()(double i, double j, int32 p = 0) const {
  vw_throw(NoImplErr() << "CorrelatorView::operator()(double i, double j, int32 p) has
  return pixel_type(); // Never reached
}

/// \cond INTERNAL
typedef CropView<ImageView<pixel_type> > prerasterize_type;
inline prerasterize_type prerasterize(BBox2i bbox) const {

  // The area in the right image that we'll be searching is
  // determined by the bbox of the left image plus the search
  // range.
  BBox2i left_crop_bbox(bbox);
  BBox2i right_crop_bbox(bbox.min() + m_search_range.min(),
                         bbox.max() + m_search_range.max());

  // The correlator requires the images to be the same size. The
  // search bbox will always be larger than the given left image
  // bbox, so we just make the left bbox the same size as the
  // right bbox.
  left_crop_bbox.max() = left_crop_bbox.min() + Vector2i(right_crop_bbox.width(), right

  // Finally, we must adjust both bounding boxes to account for
  // the size of the kernel itself.
  right_crop_bbox.min() -= Vector2i(m_kernel_size[0], m_kernel_size[1]);
  right_crop_bbox.max() += Vector2i(m_kernel_size[0], m_kernel_size[1]);
```

```
left_crop_bbox.min() -= Vector2i(m_kernel_size[0], m_kernel_size[1]);
left_crop_bbox.max() += Vector2i(m_kernel_size[0], m_kernel_size[1]);

// Log some helpful debugging info
vw_out(DebugMessage, "stereo") << "\t   search_range: " << m_search_range << std::end
vw_out(DebugMessage, "stereo") << "\t left_crop_bbox: " << left_crop_bbox << std::end
vw_out(DebugMessage, "stereo") << "\tright_crop_bbox: " << right_crop_bbox << std::en

// We crop the images to the expanded bounding box and edge
// extend in case the new bbox extends past the image bounds.
ImageView<ImagePixelT> cropped_left_image = crop(edge_extend(m_left_image, ZeroEdgeEx
ImageView<ImagePixelT> cropped_right_image = crop(edge_extend(m_right_image, ZeroEdge
ImageView<MaskPixelT> cropped_left_mask = crop(edge_extend(m_left_mask, ZeroEdgeExten
ImageView<MaskPixelT> cropped_right_mask = crop(edge_extend(m_right_mask, ZeroEdgeExt

// We have all of the settings adjusted.  Now we just have to
// run the correlator.
vw::stereo::PyramidCorrelator correlator(BBox2(0,0,m_search_range.width(),m_search_ra
                                         Vector2i(m_kernel_size[0], m_kernel_size[1])
                                         m_cross_corr_threshold, m_corr_score_thresho
                                         m_cost_blur, m_correlator_type);

// For debugging: this saves the disparity map at various pyramid levels to disk.
if (m_debug_prefix.size() != 0) {
  std::ostringstream ostr;
  ostr << "-" << bbox.min().x() << "-" << bbox.max().x() << "_" << bbox.min().y() <<
  correlator.set_debug_mode(m_debug_prefix + ostr.str());
}

ImageView<pixel_type> disparity_map = correlator(cropped_left_image, cropped_right_im
                                                 cropped_left_mask, cropped_right_mas
                                                 m_preproc_func);

// Adjust the disparities to be relative to the uncropped
// image pixel locations
for (int v = 0; v < disparity_map.rows(); ++v)
  for (int u = 0; u < disparity_map.cols(); ++u)
    if (!disparity_map(u,v).missing())  {
      disparity_map(u,v).h() += m_search_range.min().x();
      disparity_map(u,v).v() += m_search_range.min().y();
    }

// This may seem confusing, but we must crop here so that the
// good pixel data is placed into the coordinates specified by
// the bbox.  This allows rasterize to touch those pixels
// using the coordinates inside the bbox.  The pixels outside
// those coordinates are invalid, and they never get accessed.
```

```
    return CropView<ImageView<pixel_type> > (disparity_map, BBox2i(m_kernel_size[0]-bbox.
                                                                   m_kernel_size[1]-bbox.
                                                                   bbox.width(), bbox.hei
  }

  template <class DestT> inline void rasterize(DestT const& dest, BBox2i bbox) const {
    vw::rasterize(prerasterize(bbox), dest, bbox);
  }
  /// \endcond
};
```

14.2 Working with Shallow Views

14.3 Efficient Algorithms and `pixel_accessor`

14.4 Rasterization, Efficiency, and Tiled Computation

14.5 Generic Image Buffers

14.6 The File I/O System

14.7 Frequency-Domain Image Processing

Chapter 15

A Vision Workbench Cookbook

This chapter provides simple bite-sized examples of how to use the Vision Workbench to perform a range of common tasks.

15.1 Removing Camera Lens Distortion

All digital camera systems introduce some amount of distortion in the image. In some high-precision cameras this distortion may be very small, while in other such as those with fisheye lenses it may be very large. In either case removing camera lens distortion is the first step in many image processing algorithms. The idea is to transform the image so that it looks as if it were taken with a perfect pinhole camera. The resulting image is usually called either "linearized", "undistorted", or "rectified".

First you must create a camera model object, such as a `PinholeModel` or `CAHVORModel`, containing a reasonably accurate model of your camera. (Computing one if you don't have one already is another topic altogether.) Then you can simply compute the linearized image like this.

```
result = linearize_camera_transform( image, camera_model );
```

This function is declared in the header `<vw/Camera/CameraTransform.h>` and located in the `vw::camera` namespace. By default it uses the usual default edge extension and interpolation modes, which may not be best depending on your needs. Often you will get better linearization results by overriding those defaults using the optional arguments.

```
result = linearize_camera_transform( image, camera_model,
                                     ConstantEdgeExtension(), BicubicInterpolation() );
```

In many cases you would also like to know the linearized camera model that the linearized image corresponds to. You can compute it directly using the `linearize_camera` function.

```
linearized_model = linearize_camera( camera_model, image.cols(), image.rows() );
```

The linearization function needs to know the image dimensions so that it can select the linearized model that best fits the distorted model over the given range. If you do use that function to compute the linearized model in advance, or if you want to specify an alternate linearized camera model of your own, then you should use the more general camera transformation function.

```
result = camera_transform( image, camera_model, linearized_model );
```

This function allows you to transform images between any two concentric camera models; undistorting an image is just a special case. You can again override the default edge extension and interpolation settings if you wish.

Finally, if you require an even greater degree of control you can achieve it by manually creating a `CameraTransform` object and appying it to the source image directly using the regular `transform` function. This is in fact exactly what happens inside the convenience functions described above. Here is a simple example, assuming the CAHVOR camera model.

```
CAHVORModel cahvor( "my_camera_model.cahvor" );
CAHVModel cahv = linearize_camera( cahvor, image.cols(), image.rows() );
CameraTransform<CAHVORModel,CAHVModel> ctx( cahvor, cahv );
result = transform( image, ctx );
```

You might choose to do it this way if you want to compose the camera transform with another transform in a single step for efficiency, for example.