Implementation of HDR panorama stitching algorithm

Piotr Ostiak*

Institute of Computer Graphics and Multimedia Systems
Technical University of Szczecin
Szczecin / Poland

Abstract

This paper covers an implementation of fully automated HDR panorama stitching algorithm. This problem involves image recognition as we need to know which part of panorama to stitch. We used the SIFT algorithm for recognition of the corresponding points; this method is invariant to changes in image scaling, rotation and in illumination. The SIFT algorithm was modified for work with HDR images. Perspective transformations were used to set up new positions of images in the panorama and normalization of luminance was made in order to remove seams from pictures. Results of implementation of the HDR panorama stitching algorithm are presented and discussed.

Keywords: HDR panoramas, HDR images, SIFT, local features, perspective transformations

1 Introduction

Panorama is a visual representation of the environment viewed from one 3D position [21]. In a traditional photography only a small part of the surroundings can be captured. Typical compact camera has a field of view of 50x35 degrees , whereas the human visual system has a field of view of around 200x135 degrees [1]. Therefore it is not possible to represent the entire environment with a single shot of compact camera. For this reason, a wide range of techniques is used in creating panoramic images.

The first step toward creating a panorama is to take a series of still pictures from one point in 3D space which covers certain parts of the environment. A sever problem we can encounter during that process is the variation of lighting conditions from one viewing direction in the scene to another. Any real-world scene has a significant difference in luminance levels that can be perceived by the human eye. A typical camera uses 8 bits per color channel to store brightness information, which provides 256 luminance levels. Human eye is able to distinguish a contrast of 10,000:1. Hence, the traditional stills are usually too bright or too dark in certain areas, which results in some of detailed information to be lost [14, 13]. The number of distinctive steps between the brightest and the darkest point is called a dynamic range. There are a number of techniques for representing a broader dynamic range of a scene. Images which use such techniques are called High Dynamic Range Images (HDR).

In the panoramic photography those brightness variations are especially noticeable thus taking advantage of HDR images in creating panoramas is reasonable.

There are applications for automatic panorama creation [15], some of them need user input [3] to properly register images, but most stitching tools do not take advantage of 16 bit/color source images such as OpenEXR [6, 10]. Our method, thanks to the PFS Tools package [9], allows working with most of the popular HDR image formats, such as OpenEXR and 32 bit RGBE.

In Section 2 we present in general the automatic process of making HDR panoramas, following subsections describe in details each stage of the process, including image matching and perspective transformations. Section 3 provides information about taking HDR images for panoramas and the modifications we made to the matching algorithm. Section 4 describes the implementation environment and details. Section 5 consist of results and conclusion. Finally, the last section points areas where our method could be improved.

2 Constructing a panorama

The process of building a panoramic image consists of five principal stages including: taking a series of still photos, locating correspondence points in each pair of images, estimating a transformation matrix between related photographs in order to calculate a new location of images in the panorama and, finally, stitching photos together (Figure 1).

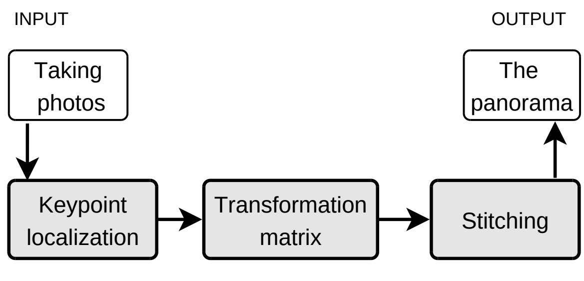

Figure 1: The process of creating panoramas.

*postiak@wi.ps.pl

A common way of putting together HDR panoramic images involves taking a sequence of LDR images with a different exposure, stitching each exposure set together to produce a sequence of LDR panoramas and, finally, making one HDR panorama by merging together all the LDR panoramas [7].

We propose a different approach based on creating a HDR panorama from a sequence of HDR images, rather than blending a series of LDR panoramas. The advantage of this method is the possibility of making panoramas from existing HDR images. Moreover, cameras capable of capturing HDR photos are becoming more and more popular, for it is HDR technology that looks upon future, and a method that would make better use of HDR images seems crucial.

2.1 Image matching

There are two main trends in automating image matching problem: direct methods and eature based methods.

Direct methods tend to iteratively estimate camera parameters by minimizing an error function based on intensity difference in the region of overlap. The advantage of direct methods lies in their use of all available data; as a result they provide a very precise registration. The disadvantage is that they are based on the assumption that the brightness of all the pixels of a region remains constant between two consecutive images ("brightness constancy" assumption) [2].

The second trend begins by establishing correspondences between points, lines or other geometrical entities [2]. Those correspondences are called local features.

Neither of the mentioned methods is invariant to scale, rotation or brightness change. Therefore they prevent a proper registration of photos that was taken in real-world circumstances.

For image matching we use a modification of the SIFT algorithm. The SIFT algorithm is a feature based method. This algorithm was chosen because it is invariant to image scaling, rotation and changes in illumination. The concept of the SIFT algorithm is explained later on in this section, whereas its modification for working with HDR images is described in the section 4.

2.1.1 SIFT algorithm

David G. Lowe in [11] suggests an algorithm for extracting local features which are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint [11]. The algorithm is called SIFT (Scale Invariant Feature Transform).

The major stages in extracting image features are:

1. Scale-space extrema detection.
2. Keypoint localization.
3. Orientation assignment.
4. Generation of keypoint descriptors.

2.1.2 Scale-space extrema detection

The first step in using the SIFT algorithm lies in extracting interest points. In order to find *interest points*, one has to search all over the scales and locations. This is implemented by building a multi-scale pyramid of Difference-of-Gaussian (DoG) images.

Given a Gaussian-blurred image:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y), \tag{1}$$

where: x,y - pixel position,
σ - denotes blur level,
$L(x,y,\sigma)$ - blurred image,
$G(x,y,\sigma)$ - Gaussian blurring function,

$$G(x,y,\sigma) = 1/(2\pi\sigma^2)\exp^{-(x^2+y^2)/\sigma^2}, \tag{2}$$

$I(x,y)$ - image being blurred.
Each pyramid's level consists of:

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma), \tag{3}$$

where, D is an image computed from the difference of two nearby levels separated by a constant multiplicative factor k $=\sqrt{2}$.

A Gaussian pyramid and a Difference-of-Gaussian pyramid are illustrated on Figures 2 and 3. The number of pyramid's levels is not fixed and may be different in various implementations.

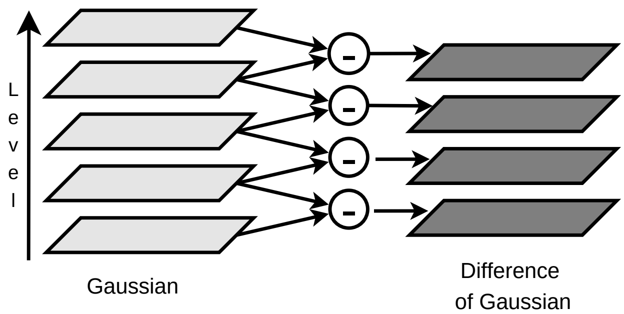

Figure 2: Adjacent images blurred with Gaussian function are subtracted to produce the DoG images [11].

In order to identify the keypoints we have to localize the local maxima and minima in the Gaussian pyramid across levels. Each pixel is compared to its 8 neighbors at the same level, plus 9 pixels on levels above and below (Figure 4). If such point is the local minimum or maximum, it should be marked as a "candidate point".

2.1.3 Keypoint localization

One should ignore those points among candidate points in which contrast is too low (the difference between its intensity and the intensities of its neighbors is too low), and

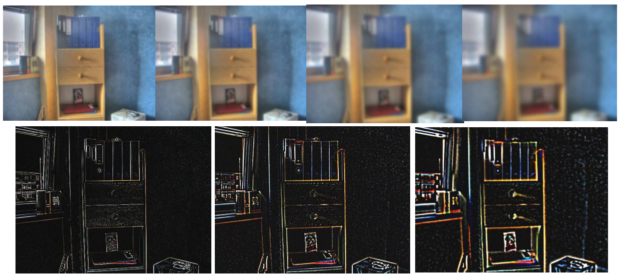

Figure 3: An example of blurred image pyramid and the DoG pyramid

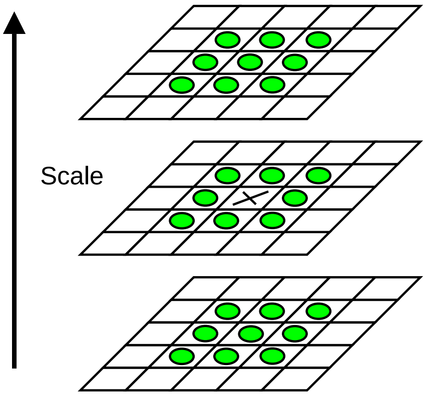

Figure 4: The pixel marked x is compared against its 26 neighbors in a 3x3x3 neighborhood that spans adjacent DoG images [11]

the edge responses are eliminated. When the initial stage is over, one should come up with a set of stable keypoints (Figure 5).

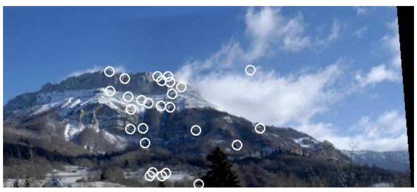

Figure 5: Two stitched images and their features

2.1.4 Orientation assignment and keypoint descriptors

The next step is to assign orientation to the keypoint. To determine the keypoint orientation, a gradient orientation histogram is computed in the neighborhood of the keypoint. The contribution of each neighboring pixel is weighted by the gradient magnitude and a Gaussian window with a sigma that is 1.5 times the scale of the keypoint [11].

The orientation corresponds to the histogram's bin with the maximum value. The histogram has 36 bins.

When keypoint orientations are assigned, the keypoint descriptor is computed from a set of orientation histograms in a 4x4 pixel neighborhood (Figure 6). The orientation histograms are relative to the keypoint orientation, the orientation data comes from the Gaussian image closest in level to the keypoint's level [11]. The contribution of each neighboring pixel is weighted by the gradient magnitude and a Gaussian window.

Each histogram has 8 bins, each descriptor consist of 4x4 histograms around the keypoint, leaving us with 4x4x8=128 element SIFT feature vector.

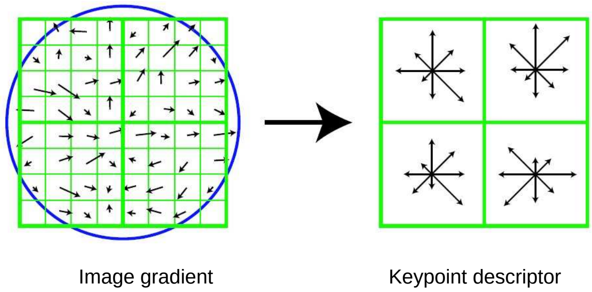

Figure 6: SIFT feature descriptor [11]

Lastly, the feature descriptor is normalized to improve invariance to change in illumination.

2.2 The Transformation Matrix

To compute a new position of pixels from an image *B* onto a final panorama *A*, we have to estimate transformation matrix *H* [19]:

$$x' \sim Hx, \tag{4}$$

where:
x - position of a pixel in the image *B*,
x' - position of a pixel in the final panorama,
$\sim$ - denotes similarity up to scale.

H is a 3x3 matrix which can be estimated by using the *Direct Linear Transform* algorithm [12]:

$$H = \begin{pmatrix} h1 & h2 & h3 \\ h4 & h5 & h6 \\ h7 & h8 & h9 \end{pmatrix}. \tag{5}$$

There is a given set of corresponding points *x* and *x*':

$$x_i = \begin{pmatrix} x_i \\ y_i \\ w_i \end{pmatrix}, x'_i = \begin{pmatrix} x'_i \\ y'_i \\ w'_i \end{pmatrix}, \tag{6}$$

where:
x_i,y_i - keypoint position,
w_i - is set to 1.

After subsequent transformations [12], the final equation is (7):

$$\begin{pmatrix} 0^T & -w'_i x_i^T & y'_i x_i^T \\ w'_i x_i^T & 0^T & -x'_i x_i^T \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = 0. \quad (7)$$

We may write (7) as:

$$A \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = 0. \quad (8)$$

Each pair of corresponding points adds two equations to the matrix A.

This is called homography or perspective transformation [19] it has 8 degrees of freedom. Perspective projection is used to merge images onto panorama as it is shown on Figure 7.

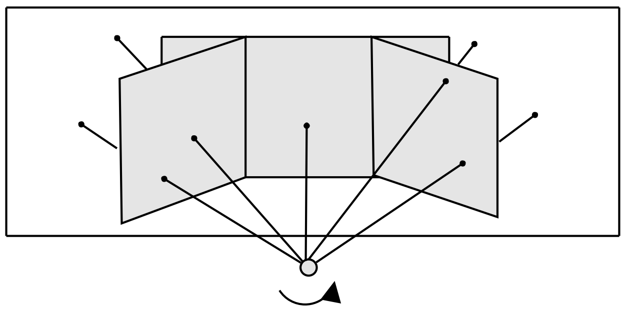

Figure 7: One image is selected as a reference, next homographies are computed for each neighboring image and finally all images are mapped to reference image plane [12]

3 HDR panorama stitching

3.1 Taking photos

Nowadays, the acquisition of high dynamic range images is an easy task and should become even simpler in the future. There are two principal methods for making HDR images: physically-based rendering and multiple-exposure image capture [20]. Due to the fact that the first method relates to the computer generated graphics, it is beyond the scope of this paper. For this reason, the paper will now focus on the second, more relevant, method of HDR image creation.

One approach to this problem is presented in [4], where they describe how multiple photographs at different exposures are to be merged together to create a single high-dynamic range image. Some commercial applications already use this technique [18]. The weakness of multi-exposure technique lies in the necessity for the camera to be mounted on a tripod. In addition, the depicted scene is forced to abandon all dynamic objects like moving cars or people.

Alternative method of taking HDR images is to use a camera that allows storage of a broader dynamic range. Fuji's SuperCCD S3 Pro is such a device. It has a chip with high and low sensitivity sensors per pixel location to increase dynamic range. The camera's image processor can then combine the values of these pixels to extend the dynamic range of the image [5].

As HDR capturing devices are very expensive and taking a series of multi-exposure pictures for panorama is problematic, because real world scenes are dynamic, Canon's RAW format comes in handy.

Canon RAW does not store the full spectrum of brightness but it broadens the dynamic range significantly because its sensors capture unprocessed data. .

One panorama produced for this article was made from Canon RAW images converted to Radiance HDR files. The other HDR images were made from a series of LDR photos.

3.1.1 The XYZ color space

High dynamic range images cannot be represented in the 24 bit RGB color space. Since the SIFT algorithm was designed to work with LDR images, it works in the RGB color space.

To fulfill HDR representation requirementsw e propose to move from the 24 bit RGB color space to the floating point XYZ color space. CIE XYZ represent full dynamic range and full color gamut. It is also used as internal PFStools format so converting to XYZ color space simplifies the implementation.

To convert a pixel form the RGB color space to the XYZ color space a color vector [R G B] must be multiplied by the transformation matrix M:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.4124 & 0.3575 & 0.1804 \\ 0.2126 & 0.7151 & 0.0721 \\ 0.0193 & 0.1191 & 0.9502 \end{pmatrix} * \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (9)$$

In The CIE XYZ color space the Y parameter is a measure of the color brightness. Since all of the primaries are non-negative we can use the total energy $E=X+Y+Z$ as the intensity level of the pixel and each computation in the modified SIFT algorithm is being made on the total energy.

3.1.2 More precise registration

HDR images contain broader scope of brightness information. Thus it is reasonable to use this for the advantage in modification of the SIFT algorithm. The SIFT algorithm works in the gradient domain. There is a plenty of real-world situations where the difference in gradients between the two objects is far too great to be represent in the 256 levels of luminosity. In HDR photography we can extend the gradient's range, yet, since the SIFT algorithm works on luminance that is normalized to the range of 0.0-1.0 and low-contrast features are rejected, resulting in a loss of details. Features are properly extracted from LDR images with contrast threshold set to around 0.007, it gives a reasonable number of stable features. While extracting features from HDR image the threshold value should be lowered to 0.0001 in order to keep additional HDR details.

3.2 Image blending

By multiplying each pixel by transformation matrix, we get pixel's new position.

Usually, two overlapping images were taken in slightly different lighting conditions. In results, frame edges between them are noticeable. One method dealing with this drawback in panorama creation process is the, so-called, weighted average method. Each image is multiplied by a weighting function which decreases monotonically across its border (Figure 8); the resulting images are then summed to form the mosaic [8].

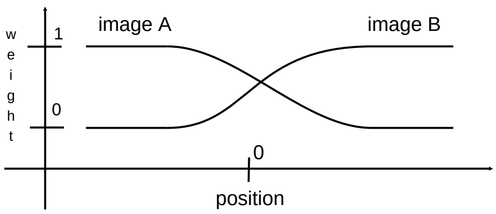

Figure 8: Example weighting functions are shown here in one dimension

While we work on the HDR images, blending two images together is being made by a normalization of the luminance and the chrominance values. Through comparison of the brightness level between the corresponding pixels, one should come up with the brightness ratio between the given photos.

The luminance values from the first picture are then proportionally adjusted to the values from the second one.

The process of blending the two HDR images should then cease to be troublesome, as we no longer have to worry about the brightness diversity nor the difference of color saturation.

4 Implementation

The method presented in this paper was implemented in c++ (gcc). All operations on HDR images were made in the *PFS Tools* package [9]. PFS Tools is a software package including a set of command line programs for reading, writing, manipulating and viewing HDR images and video frames. This package includes a c++ library that can be used for writing various applications that can be integrated with existing PFS Tool programs.

The PFS has a modular structure and it uses UNIX pipes. It makes the exchange of data between applications even easier. PFS Tools works both on a single image and a sequence of frames; therefore, it is an excellent environment for writing the panorama creation software.

An application which implement the presented method extends the PFS Tools and would certainly be included it the PFS Tools software package.

An example execution of the HDR panorama creation program would be:

```
pfsin|image1.hdr␣image2.hdr␣...␣imagen.hdr
|panohdr|pfsout␣panorama.hdr
```

where:

pfsin - is a program for reading HDR/LDR images, it gives PFS frame on the output

panohdr - is a program for creating panoramas

pfsout - is a program for writing a PFS frame as a HDR/LDR image

5 Results and conclusions

Three experiments were made to analyze the quality of panoramas stitched with our application. In each experiment a HDR and a LDR panoramas were produced to consider if the HDR panorama has any advantages over the LDR one.

All of the experiments were made in the Fedora Core Linux operating system, on Athlon 2000+ with 512 MB of RAM. HDR photos that were used in our tests were acquired in two manners. HDR images were created and its dynamic range was measured in Photomatix [17].

Images for the first panorama were converted from Canon RAW images and are of size of 775x512. The dynamic range of this photo is of 1510:1. Images for the LDR panorama were acquired by converting the RAW file to the LDR TIFF format. Figure 9 show pictures which were u sed to stitch the first set of panoramas.

Figure 9: HDR (upper) and LDR (lower) input pictures from which panoramas will be created.

HDR images for two other panoramas were capture with Canon PowerShot A510. The second panorama was made of HDR images acquired by merging a series of LDR images taken with aperture of F3.2 and exposures of: 1/4s, 1/10s, 1/40s, 1/160s and 1/640s. It gave a HDR photos with dynamic range of 10036:1. The LDR panorama was stitched from two LDR images taken with exposure time of 1/10s. The third panorama was created from HDR generated from LDR images taken with aperture of F6.3 and exposure times of: 1s, 0.4s, 1/10s, 1/40s, 1/160s and 1/640s. The dynamic range of this image is 534131:1.

In order to display a HDR image on a LDR display device it has to be tone mapped. Tone mapping is a process of converting the high dynamic range to the low dynamic range. HDR images which were used in this article were tone mapped with Photomatix Tone Mapping Tool.

The first set of panoramas (Figure 12) was made from three images taken from noticeably different angles, to examine if local feature will be properly extracted from differently rotated, translated and scaled images. Images were aligned properly and no registration errors are visible. It can be observed that the HDR panorama is more colorful.

Next there is a comparison of two panoramas depicting the same scene (Figure 13). The upper panorama covers high dynamic range, whereas the lower one has low dynamic range. It is clearly visible that on the LDR panorama some details are lost. Bright areas were overexposed and objects that are visible on the HDR picture are invisible on the LDR panorama. Objects that remain in shadows are also missing on the LDR panorama (Figure 10).

The last set of panorama portrays a scene with a very high dynamic range (Figure 14). On the first sight the differences between the HDR and the LDR panorama are not very distinctive. However on the enlarged figure (Figure 11) additional details can be noticed on the HDR photo.

High dynamic-range images store a full range of visible color. A HDR photograph will never be under- or overexposed, as it stores a full scope of brightness. This issue is imperative in the discussion of panoramic images, since the two images can vary significantly in luminance. By using HDR images for panoramas, one can handle those real-world lightning differences.

In the course of this paper, we used the Scale Invariant Feature Transform algorithm as a base method for feature matching. This algorithm was later modified for better compatibility with HDR images and was further enhanced by using the benefits of a higher dynamic-range in extracting some additional local features.

6 Future Work

There are some areas where our method could be improved. It does not deal with dynamic objects, which appear differently on two another photos. The simple image blending method we used should be replaced by a more robust multiresolution spline technique [8].

Another inconvenience is associated with photos' distortion that comes from the perspective transformations when a broad field-of-view is covered. Therefore the perspective transformations should be improved by applying a lens distortion correction method [16].

Figure 10: Shadowed objects are not visible on the LDR picture (lower).

Figure 11: Enlarged fragment of a HDR panorama (upper) and a LDR panorama (lower).

Figure 12: A HDR (upper) and a LDR (lower) panaroma made of three images.

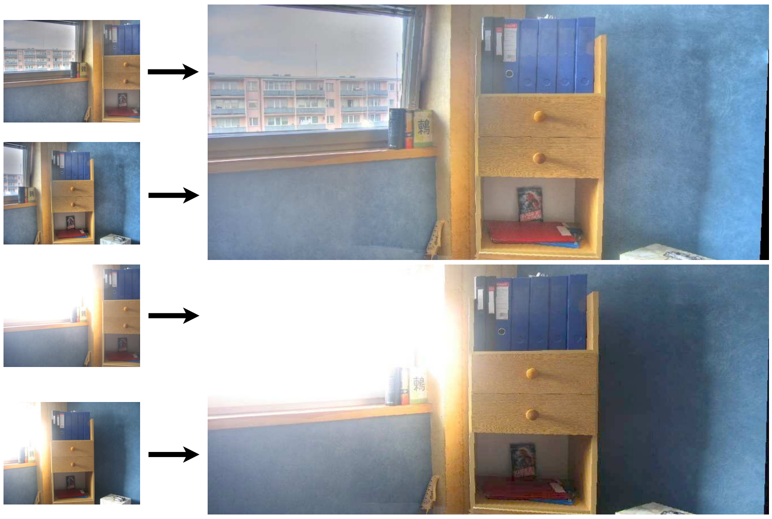

Figure 13: An indoor HDR panorama (upper) and LDR panorama (lower).

Figure 14: An indoor HDR panorama (upper) and LDR panorama (lower).

References

[1] M. Brown and D. Lowe. *Recognising Panoramas - slides*. University of British Columbia, 2004.

[2] M. Brown and D. G. Lowe. *Recognising Panoramas*. Ninth IEEE International Conference on Computer Vision (ICCV'03) - Volume 2, 2003.

[3] Pablo d'Angelo. Hugin. Technical report, 2005. Available at `http://hugin.sourceforge.net/`.

[4] Paul Debevec and Jitendra Malik. *Recovering High Dynamic Range Radiance Maps from Photographs*. University of California at Berkeley, 1997.

[5] Fuji. http://www.dpreview.com/reviews/fujifilms3pro/page18.asp. 1997.

[6] E. Goetze. Hdr for panoramas. Technical report, 2005. Available at `http://radio.weblogs.com/0110138/stories/2003/02/06/hdrForPanoramas.html`.

[7] HDRSOFT. Technical report, 2005. Available at `http://www.hdrsoft.com/examples.html`.

[8] J.Burt and E.H.Adelson. *A Multiresolution Spline With Application to Image Mosaics*. RCA David Sarnoff Research Center, 1983.

[9] G. Krawczyk and R. Mantiuk. Pfs tools. Technical report, 2005. Available at `http://www.mpi-sb.mpg.de/resources/pfstools/`.

[10] Idustrial Light and Magic. Openexr. Technical report, 2005. Available at `http://www.openexr.com/`.

[11] David G. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints*. Computer Science Department University of British Columbia Vancouve, 2004.

[12] Marcus Magnor. *Computer Graphics II - Camera Geometry*. 2005.

[13] S. K. Nayar and T. Mitsunaga. *High Dynamic Range Imaging: Spatially Varying Pixel Exposures*. Department of Computer Science Media Processing Laboratories, 2003.

[14] Bao Paul and Xu Dan. *Complex wavelet-based image mosaics using edge-preserving visual perception modeling*. Pergamon Press / Elsevier Science, 1999.

[15] A. Negri S. Negri and L. Robert. Realviz. Technical report, 2005. Available at `http://www.realviz.com/`.

[16] H. S. Sawhney and R. Kumar. *True multi-image alignment and its application to mosaicing and lens distortion correction*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999.

[17] HDR Soft. Photomatix. Technical report, 2005. Available at `http://www.hdrsoft.com/index.html/`.

[18] Idruna Software. Photogenics. Technical report, 2005. Available at `www.idruna.com/`.

[19] Richard Szeliski. *Image Alignment and Stitching:A Tutorial*. Microsoft Research, 2004.

[20] Greg Ward. *High Dynamic Range Imaging*. Exponent Failure Analysis Assoc, 2001.

[21] Y. Xiong and K. Turkowski. *Registration, Calibration and Blending in Creating High Quality Panoramas*. Jet Propulsion Laboratory, Black City, 1998.