RC22747 (W0303-086) March 18, 2003
Computer Science

IBM Research Report

Digital Image Acquisition and Continuous Zoom Display from Multiple-Resolution Views Using Heterogeneous Image Pyramids

Belle L. Tseng
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Digital Image Acquisition and Continuous Zoom Display from Multiple-Resolution Views Using Heterogeneous Image Pyramids

Belle L. Tseng
IBM T. J. Watson Research Center
19 Skyline Drive, Hawthorne, NY 10532
Belle@us.ibm.com

ABSTRACT

There are many ways of capturing images to represent a detailed scene. Our motivation is to use inexpensive digital cameras with little setup requirements and to allow photographers to differentially capture both low-resolution overviews and high-resolution details. We present the heterogeneous image pyramid as a non-uniform representation composed of multiple captured multi-resolution images. Each resolution image captures a specific portion of the scene at the photographer's discretion with the desired resolution. These images are highly correlated since they are captured from the same scene. Consequently, these images can be registered and represented more compactly in a 3-dimensional spatial image pyramid called the heterogeneous image pyramid.

Keywords: image pyramid, mosaic, spatial representation, multiresolution, super-resolution, image resolution.

1 INTRODUCTION

We capture images to represent real-world scenes and display them back to share with others. The displayed images can be shown to users with varying viewpoints and different resolutions. The images can be used as texture maps to build virtual environments, or generate a mosaic to provide a panorama of the view from the mountain peak. There are many image representations of the desired scene as they are specific to the final application and task. In David Sarnoff Research Center, a taxonomy of mosaic representations is developed, where they discuss the completeness of the representations and examine how the various mosaics can be used for different applications.[13] Kumar *et al*. presents a hierarchical framework for scene representation where four levels in the hierarchy support various types of tasks so that the overall structure grows in capability as more information about the scene is acquired.[14] In our introduction, we describe a different workflow for scene representations.

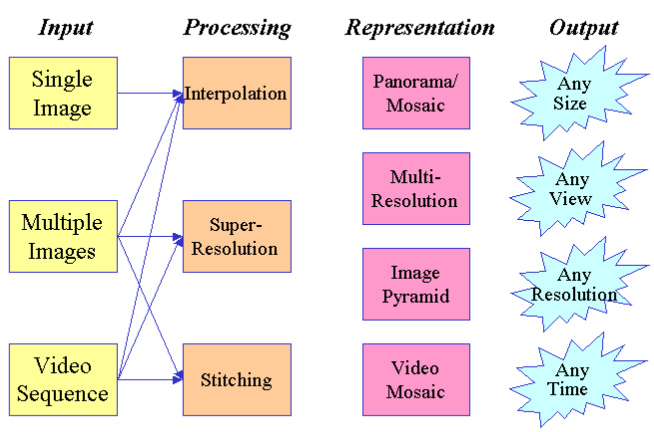


Figure 1: Workflow from ***Image Inputs*** to ***Image Processing*** to ***Scene Representations*** to ***Output Capabilities***.

In Figure , we illustrate the four major stages from the forms of captured images to the final displayed capabilities of the output. In the *Input* stage, the photographer captures one or more images of the real-world scene. In the *Processing* stage, the input images undergo some image processing steps. In *Representation*, the processed images are stored in some optimal image representation. Finally in the *Output* stage, the display can show the user different aspects of the captured scene.

Three categories of inputs can be captured by the photographer to illustrate the desired scene. The first category is a single image. These images tend to of high resolution so as to capture the details of the scene. Also, wide-angle images can be captured using Fish-eyes lens. Furthermore, hemispherical images using special camera attachments can capture the full 360 degree in a single image. From a single image, the interpolation process offers the user to view the scene in varying but limited resolutions and sizes.

In the second input category, multiple images are captured. These images can be captured from one camera or multiple cameras. Consequently, they may capture both static and dynamic scenes, leading to some desirable and undesirable attributes. One common capture process of multiple images is to capture a panorama with a regular camera. The photographer takes a series of pictures with some translation and small overlap. In this case the use of the stitching process can be use to combine and create one panoramic mosaic. On the other hand, the multiple images can be captured with different viewpoints, resolutions, sizes, and overlaps. In this case, the use of the super-resolution can be used to fuse the overlapping image regions.

The third input category is video sequences. Video is the natural extension to capturing multiple continuous images. In a video sequence, consecutive frames are highly correlated in both image intensity and region overlap. As a result, stitching can easily be performed resulting in smoother mosaics. With the high region overlap of consecutive frames, the super-resolution process can be incorporated to improve the resulting image quality.

One caveat in both multiple image and video input captures is the difficulties in representation of dynamically changing scenes. In our work, the input category is multiple images with the use of inexpensive digital cameras and little setup requirements. In addition, the photographer is allowed to differentially capture both low-resolution overviews and high-resolution details of a static scene. The rest of this section is followed by prior work then an introduction into our heterogeneous image pyramid (HIP).

1.1 Prior Work

In recent years, panoramic images have been widely used to represent virtual environments from real-world scenes. The traditional approach of representing scene information is to relate each captured image to an abstract 3D coordinate system. Another direct approach represents the relationships among the collection of captured images. Over 30 years of research in scene representation, computer vision has demonstrated the difficulty of computing the 3D object representations from image sources.

The first *Input* category is single image captures. One way to increase the resolution of an image is to interpolate the pixel intensities. Some of the well known interpolation algorithms include nearest-neighbor, bilinear, and variants of cubic spline interpolation.[9] Shultz and Stevenson provide a survey of several more sophisticated interpolation algorithms including regularization approaches, edge-preserving techniques, and Bayesian algorithms.[16] Interpolation will work well if the input images are fairly high resolution to start with; however, interpolation becomes much more difficult as the image size gets smaller and does not contain the details to begin with.[1][2]

The second *Input* category uses stitching and super-resolution algorithms on multiple images captured from the scene. Super-resolution is the process of combining multiple low resolution images to form a higher resolution image. Numerous super-resolution algorithms exist with varying performances. While super-resolution generally offer huge improvements over the original input images, for large magnification factors the high frequencies are generally not reconstructed very well.[1]

The first step to super-resolution is to register the images. Registration is a fundamental problem in image processing used to match two or more images. Registration computes the displacement of pixels from one image to the others. Over the years, there has been a broad range of registration techniques including edge-based approaches[11], frequency domain techniques, and multi-resolution feature-based methods[12]. A good survey of image registration techniques are given in [10], where they organize the techniques by establishing the relationship between the distortions in the input images and the type of registration techniques that are most suitable.

After image registration for the second step of super-resolution, the low resolution input images need to be fused to form the high resolution images. Baker and Kanade propose learning a prior on the image gradient in their super-resolution algorithm to yield 4-8 fold improvements in resolution using as few as 2-3 images.[2] Elad and Feuer reconstruct a super-resolution image sequence from an input low resolution sequence using a generalization of the stochastic estimation based methods for the restoration of single blurred and noisy images.[5][6] The blur, decimation, and noise degradations are modeled as sparse matrices linear equation connecting the measurements to the ideal required image.

The third *Input* category captures the real world in a video sequence, say using a hand-held camcorder. The construction of one super-resolution panoramic image of a real-world scene from video sequences is highly desirable. Panoramic mosaics are an efficient way to represent a collection of images as well as video segments. The mosaic provides a significant reduction in the total amount of data needed to represent the scene. Panoramic stitching methods can be used to select a set of optimal keyframes in the video. Chen *et al.* describe their real-time system to construct panoramic images from video sequences automatically using frame stitching by minimizing the gradient error.[4]

Shum and Szeliski present techniques for automatically deriving realistic 2D scenes and 3D texture mapped models from video sequences.[17][18] They begin by performing image mosaic on flat scenes and panoramic scenes, then progress to more complicated scenes resulting in full 3D models. By mapping the texture mosaic onto an arbitrary polyhedron surrounding the origin, we can explore the virtual environment using standard 3D graphics viewers and hardware without requiring special-purpose players.

For super-resolution video reconstruction, Patti et al. proposes a complete model of video acquisition and a corresponding super-resolution algorithm.[15] Their videos are acquired with an arbitrary input sampling lattice, a sensor element's physical dimensions, the aperture time, focus blurring, and additive noise.

Our system differs from previous works in that our inputs and generated outputs are not of uniform resolution. A multi-resolution image is an image with different resolutions in different places. Many painting and editing tools work on much multi-resolution images, allowing an artist to work on a simple image at various resolutions on different parts of the picture. Berman *et al.* uses the Haar wavelet decomposition of a multi-resolution image to support for fractional-level zooming and editing and a variety of compositing operations.[3] Finkelstein *et al.* presents a multi-resolution video representation for time-varying data that allows for varying and arbitrarily high spatial and temporal resolutions in different parts of a video sequence.[7][8] Their representation is based on a sparse binary tree of sparse quadtrees. The binary tree encodes the flow of time, and each quadtree encodes the spatial decomposition of a frame.

1.2 HIP Overview

Our work deals with the problem of acquisition, storage and display of image data. For example we may wish to allow a user to browse an image of a natural scene with provisions for zooming certain portions of the image that the author of the image has deemed of interest. To acquire images that allow considerable zooming requires expensive photographic equipment that is usually based on photographic film, followed by scanning the developed film. The increasing availability of digital processing has made it feasible to use a method that permits more inexpensive digital capturing devices, such as digital cameras. However, these typically have much lower resolution than photographic film. To make up for this lack of resolution, an author who wishes to convey an impression of a scene would in all likelihood take a picture of the overall scene at low angular resolution, and take additional higher resolution pictures of interesting areas. Note that interesting in this case does not necessarily imply areas with a large amount of information. For example, a crowd in a stadium when viewed at a large distance, while containing much detailed information, may still be deemed uninteresting. In business and electronic commerce applications interesting might mean detail that an advertiser might coax a viewer to uncover after selecting the image region with an interactive input device such as a mouse. A good example would be allowing a user to zoom in on an image of the fabric of a dress.

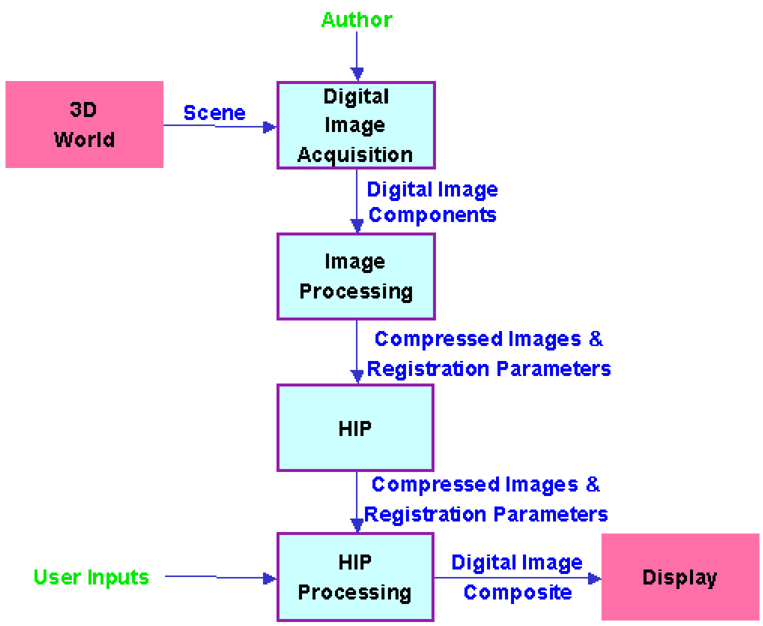


Figure 2: Overview of the System Block Diagram using HIP to represent the 3D world.

The framework, as illustrated in Figure 2, allows the author to select a collection of resolutionally non-uniform, captured digital image views of a static 3D world, thus allowing the author to define the more interesting sections of the scene from the less interesting sections. The system performs image processing to obtain the registration information between the input images. The registration data along with the multiresolution images are compactly stored in a compact representation that we refer to as the heterogeneous image pyramid (HIP). Following, when the user requests to view the scene, HIP processing provides for continuous zooming and display of the composite image representation having different levels of resolution at different locations. The main advantages of the HIP include:

(A) Define the more interesting portions of the scene from the less interesting ones by the author.
(B) Accept a non-uniform distribution of multiple resolution images.
(C) Unlimited capture of digital images where camera rotation and magnification are freely selected.
(D) Unlimited image format, size, pixel resolution, and spatial overlap.
(E) Enable arbitrarily continuous zoom and display of the scene.
(F) Reduce total data size required to represent the target scene.
(G) Use optimized hardware graphics accelerator for continuous zoom display.

2 SYSTEM OVERVIEW

The system block diagram, as shown in Figure 3, illustrates a system for capturing images of the 3D world, then processing the images for storage, followed by user-dependent processing of the stored data, for final rendering of an image on the display. In the 3D world, an author selectively captures 2D images of the scene then digitizes them to yield digital image components, as represented by the digital image acquisition block. With these digital image components, image coding and image registration are performed to give us a corresponding set of compressed images and registration parameters, as depicted by the image processing block. The derived compressed images and registration parameters are stored on disk for later reuse or for transmission, as shown by the storage block. Upon receiving the compressed images and registration parameters by the user, the compressed images are decoded and combined with the registration parameters to construct the heterogeneous image pyramid (HIP), as depicted in the HIP processing block. Furthermore, the HIP representation enables construction of the digital image composite, which is dependent on the user input. Upon updated user inputs, a new digital image composite is derived and the image is rendered on the screen, as illustrated by the display block. The next section discusses the first module.

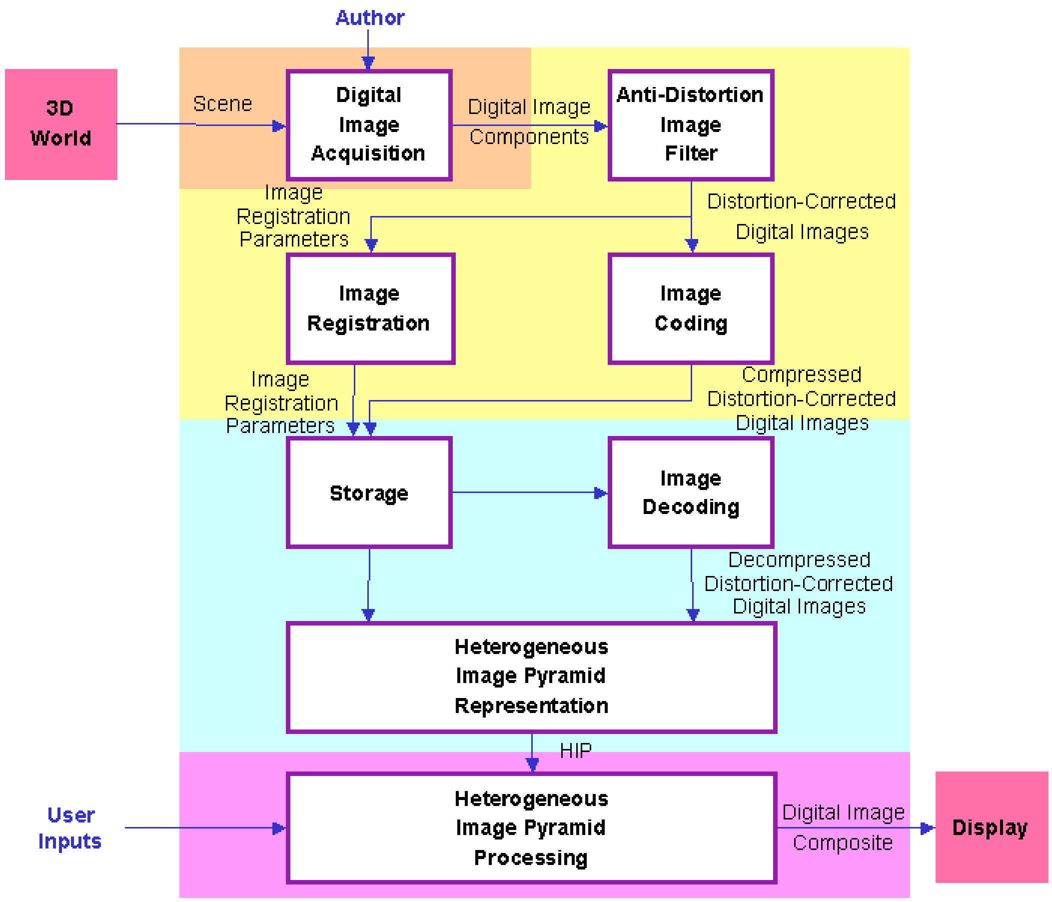


Figure 3: System Block Diagram divided into Four Modules: (1) Digital Image Acquisition, (2) Image Processing, (3) HIP Representation, and (4) HIP Processing.

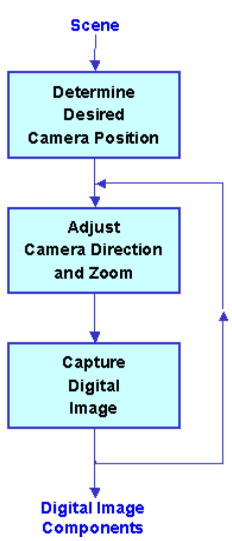


Figure 4: Digital Image Acquisition Block Diagram.

3 DIGITAL IMAGE ACQUISITION

The digital image acquisition block diagram in Figure 4 allows the author to capture a scene of the 3D world and output a set of digital image components. The author is a photographer and an image digitizer. First, the author selects the position for the camera device to capture the best viewpoint of the scene, as illustrated by the determine desired camera position block. After setting and fixing the camera position, the author adjusts the camera direction and camera zoom, as shown by the adjust camera direction and zoom block. The camera direction includes the camera pan, tilt, and revolution. Next, the author allows the camera device to capture an image of the selected scene and, if necessary, digitally scan the image to yield a digital image component, as depicted by the capture digital image block. Finally, the author determines if another digital image of the scene is desired. If another image is wanted, then the author returns to adjust the camera direction and camera zoom using the fixed camera position. Following, the author captures a new digital image and this process is repeated, resulting in a collection of digital image components. Thus the collection of digital image components is an author selected set of desired views of the same scene and may be of different image sizes, different pixel resolutions, and different spatial locations, but where they have some degree of spatial overlap.

The preferred mode of author operation is as follows. After determining the desired camera position, adjust the camera direction and zoom so that an image encompassing the whole target scene can be captured, usually resulting in a fairly low resolution digital image. Then depending on where the areas of interest or importance are in the scene, adjust the camera direction and vary the camera zoom to obtain more higher resolution digital images. Keeping in mind that the camera direction and zoom can be adjusted to capture the desired digital images, but the camera position must not be changed. Thus the resulting collection of digital image components captures the author selected set of resolutionally non-uniform representations of the desired scene, where higher resolution digital images can be taken of higher detailed or more important sections of the scene, and lower resolution digital images are taken of lower detailed or less important sections.

For example, assume we wish to capture a resolutionally non-uniform collection of digital image components of an ordinary textbook cover. Using one CCD camera device to capture multiple digital images of the book, the author first determines the optimal viewpoint of the camera device and fix the camera to that position. Adjusting the camera direction and zoom, one low resolution digital image is captured of the whole book cover. Next, the author focuses the camera toward the textbook title by slightly rotating the camera direction and increasing the camera zoom to capture a higher resolution digital image of the title text. Following, the author focuses the camera toward the textbook author names by readjusting the camera direction and modifying the camera zoom to capture another high resolution digital image of the author names. These three digital image components then represent a resolutionally non-uniform distribution of multiple resolution scene data. Due to the nature of the CCD camera device, the resulting collection of digital image components are of the same fixed image sizes and same spatial resolutions, but of different pixel magnifications and different spatial locations.

4 IMAGE PROCESSING

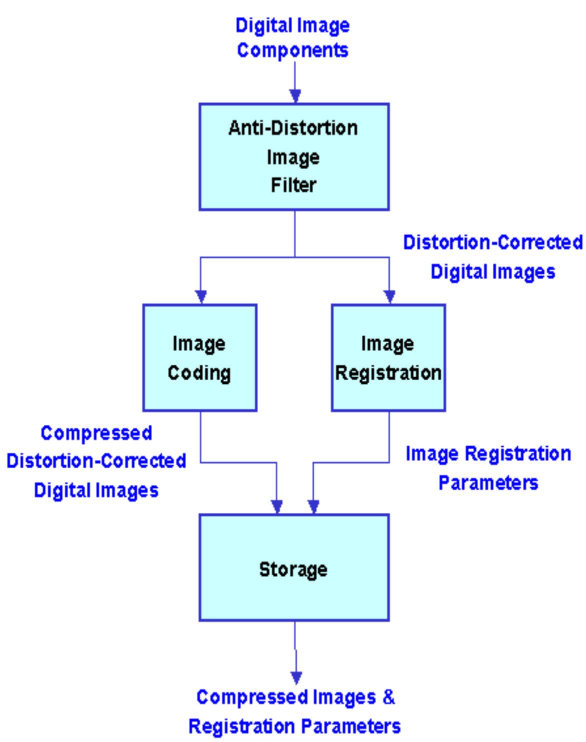


Figure 5: Image Processing Block Diagram.

The image processing block diagram in Figure 5 takes the digital image components and computes a corresponding set of compressed images and registration parameters. The first step is to correct for the undesirable non-linear optical distortions resulting from most image acquisition devices, as illustrated by the anti-distortion image filter block. Image registration does not work well when the images are affected by these lens distortions. For example, inexpensive video-conferencing cameras usually have short focal length lenses resulting in a lot of barrel (or pin-cushion) distortions. Object features far away from the optical axis known to be straight are imaged as curved lines. In other applications such as remote sensing, lens distortion is much less severe. In any event, an anti-distortion image filter obtained from a camera calibration is highly desirable. The anti-distortion image filter is used to calibrate and rectify all the digital image components, resulting in a corresponding collection of *distortion-corrected digital images*.

Secondly, the *distortion-corrected digital images* are compressed for storage and retrieval purposes yielding a corresponding set of *compressed distortion-corrected digital images*, as shown by the image coding block. The image coding stage represents any form of image compression for disk space constrained storage limitations. Two types are the wavelet image compression and DCT-based compression.

4.1 Registration

Simultaneously, the *distortion-corrected digital images* are analyzed to derived a corresponding set of *image registration parameters*, as depicted by the image registration block. The image registration output assumes a parameterized transform from one digital image to another. One such transformation is the affine image transformation r2 = A * r1 + b. Where r2 is the transformed two-dimensional image vector, r1 the given image vector and A and b are parameters. In a simple case, the two images may be linearly shifted versions of each other. Another example is the known homographic transformation r2 = (A * r1 + b) / (c * r1 + 1), containing 8 parameters. The latter transformation and other more general transformations handle perspective as well. The image registration problem finds the set of parameters that minimizes a distance metric (mismatch) defined for a reference image and its transformed counterpart.

Finally, the collection of *compressed distortion-corrected digital images* and their corresponding *image registration parameters* are saved in storage for later retrieval as *compressed images and registration parameters*, as illustrated by the storage block.

4.2 Register

The image registration block diagram determines the parameterized transforms from each *distortion-corrected digital image* to a reference coordinate system, yielding a corresponding set of *image registration parameters*. For our digital image acquisition system where the camera position is fixated, the parameterized transforms only include but are not limited to the *spatial offset*, *spatial rotation*, and *resolution magnification* for each digital image relative to some reference coordinate system.

The image coordinate system (ICS), denote *ICS(u,v)*, is a discrete 2D coordinate system with orthogonal *u*- and *v*-axes, where *u* and *v* are discrete and integral. For a digital image with index *j*, denoted I_j, let $ICS_j(u,v)$ be its image coordinate system and $I_j(u,v)$ be its image pixel intensities where *u* ranges from 0 to (M-1) and *v* ranges from 0 to (N-1) for an MxN image size. Furthermore for a total collection of J digital images, index *j* ranges from 1 to J.

The global coordinate system (GCS) is chosen to be the reference coordinate system corresponding to the Cartesian image coordinate system of the lowest resolution digital image; i.e., the image taken with the lowest zoom magnification. Consequently, the global coordinate system denoted by *GCS(x,y)* is a discrete 2D coordinate system with orthogonal *x*- and *y*-axes, where *x* and *y* are discrete but not necessarily integral, with horizontal scale unit x=1 corresponding to that of u=1 and vertical scale unit y=1 corresponding to that of v=1 of the lowest zoom magnification digital image *I(u,v)*.

After determining the global coordinate system, each *distortion-corrected digital image* I_j is iteratively analyzed against the reference coordinate system to determine its *spatial offset*, *spatial rotation*, and *resolution magnification*. The *spatial offset* (x_j, y_j) for digital image I_j corresponds to the translation component of the image coordinate system $ICS_j(u,v)$ relative to the origin of the global coordinate system *GCS(x,y)*. The *spatial rotation* r_j for digital image I_j corresponds to the rotational component of the image coordinate system $ICS_j(u,v)$ relative to the orthogonal axes of the global coordinate system *GCS(x,y)*. The *resolution magnification* m_j for digital image I_j corresponds to the scaling components of the image coordinate system $ICS_j(u,v)$ relative to the unit scale of the global coordinate system *GCS(x,y)*.

5 HETEROGENEOUS IMAGE PYRAMID

The HIP processing block diagram receives the *compressed images and registration parameters* from the storage block and generates the *digital image composites* dependent on *user inputs*. Initially, the *compressed distortion-corrected digital images* are uncompressed to yield the *decompressed distortion-corrected digital images* as illustrated by the image decoding block. Combining the *image registration parameters* and the *decompressed distortion-corrected digital images*, a data structural representation thereafter referred to as the *heterogeneous image pyramid* (*HIP*) is constructed, and an example of the HIP is depicted in Figure 6. Using the *HIP* and the response from the *user inputs*, a *digital image composite* is generated for the screen display to reflect every user response..

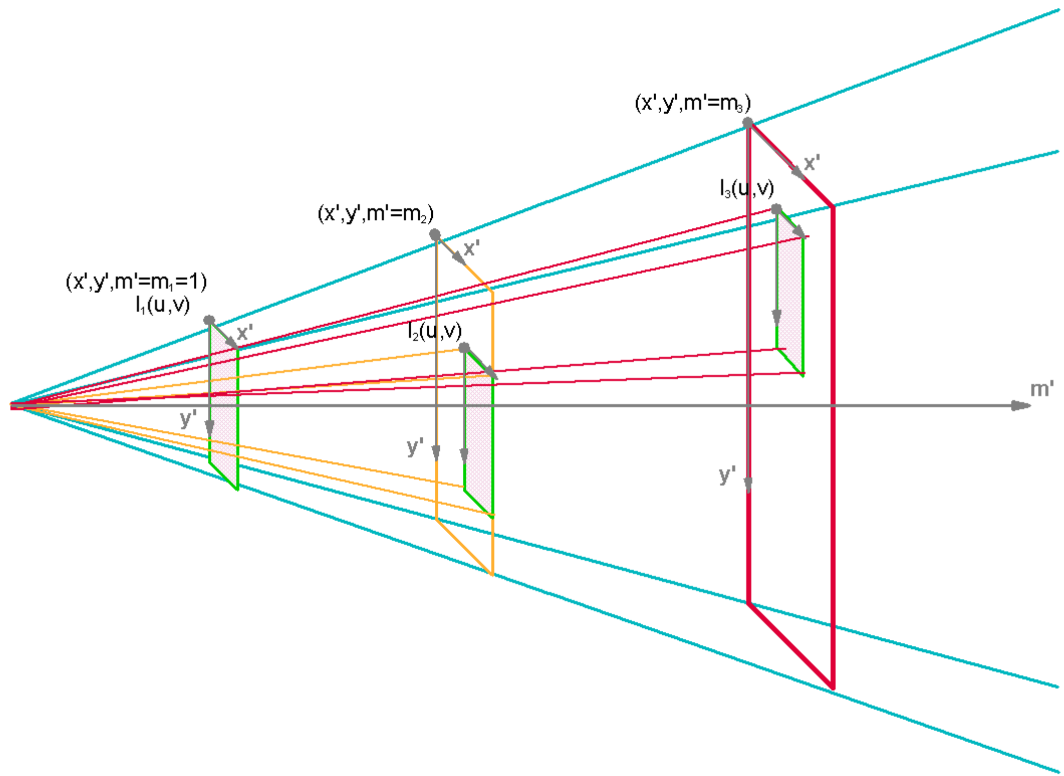


Figure 6: Heterogeneous Image Pyramid Representation.

5.1 Heterogeneous Image Pyramid Representation

The heterogeneous image pyramid representation composes and generates an image structural representation called HIP from the collection of *decompressed distortion-corrected digital images* and their corresponding *image registration parameters*. The HIP representation maintains and manages the different levels of resolution magnification images with different spatial offsets and spatial rotations. The foundation of the HIP is an *image pyramid* build using the *image registration parameters*. Simultaneously, the *decompressed distortion-corrected digital images* are ordered using the respective *image registration parameters*, resulting in a set of *ordered images*. Finally, the *ordered images* are placed in the *image pyramid* to construct the *HIP*.

The ordering and indexing of the *decompressed distortion-corrected digital images* are performed as follows. The *image registration parameters* for each digital image I_j provide the *spatial offset* (x_j , y_j), the *spatial rotation* r_j , and the *resolution magnification* m_j for the corresponding image coordinate system $ICS_j\ (u,v)$ with respect to the global coordinate system $GCS(x,y)$. Using the *image registration parameters*, the *decompressed distortion-corrected digital images* I_j are ordered according to their resolution magnification m_j , where the digital image with the lowest resolution magnification is assigned to be the image with the lowest index j=1. Thus digital images I_1 , I_2 , I_3 , ... , I_J are ordered with respective increasing resolution magnifications m_1 , m_2 , m_3 , ... , m_J . Following without loss of generalization, the resolution magnifications are normalized such that the lowest resolution magnification is equal to one, $m_1=1$. This can be achieved by dividing each resolution magnification m_j by the lowest resolution magnification m_1.

For the formulation of our multi-variable image pyramid as illustrated in Figure 6, three axes and their scale units need to be defined. Two orthogonal axes are defined to correspond to those of the global coordinate system *GCS(x,y)*, and are denoted by variables x' and y'. The third orthogonal axis is defined to be the resolution magnification, and is denoted by variable m'. The volumetric image pyramid denoted by $IP(x',y',m')$ is a 3D coordinate system where (x',y') corresponds to the global coordinate system and m' corresponds to the resolution magnification. The relationship between the global coordinate system *GCS(x,y)* and the image pyramid is the following constraints:

$$x' = m' \bullet x \qquad \text{and} \qquad y' = m' \bullet y .$$

Consequently for limited ranges of variables x and y, the ranges of variables x' and y' are directly proportional and linearly increasing with respect to the magnification variable m', yielding a pyramidal volume along the m' axis.

Having defined the *image pyramid*, the set of *ordered images* is incrementally placed in the volumetric pyramid representation. For each ordered image I_j, starting with that of the lowest resolution magnification (namely I_1) and increasing in index j, copy the digital image I_j into the image pyramid such that the image coordinate system $ICS_j(u,v)$ is transformed to the image pyramid coordinate system $IP_j(x',y',m')$ using its image registration information.

For instance digital image I_1 with image coordinate system $ICS_1(u,v)$ and spatial offset $(x_1, y_1) = (0, 0)$, spatial rotation $r_1 = 0$, and resolution magnification $m_1 = 1$, discrete image intensities $I_1(u, v)$ for $u \in [0,M)$ and $v \in [0,N)$ are copied onto the volumetric pyramid at the image pyramid coordinates $IP_1(x'=u,\ y'=v,\ m'=m_1=1)$. For digital image I_j with image coordinate system $ICS_j(u,v)$ and spatial offset (x_j, y_j), spatial rotation r_j, and resolution magnification m_j, discrete image intensities $I_j(u, v)$ for $u \in [0,M)$ and $v \in [0,N)$ are copied onto the volumetric pyramid at the image pyramid $IP_j(x', y', m'=m_j)$. Figure 6 illustrates this for digital image I_1, I_2, and I_3.

5.2 HIP Processing and Display

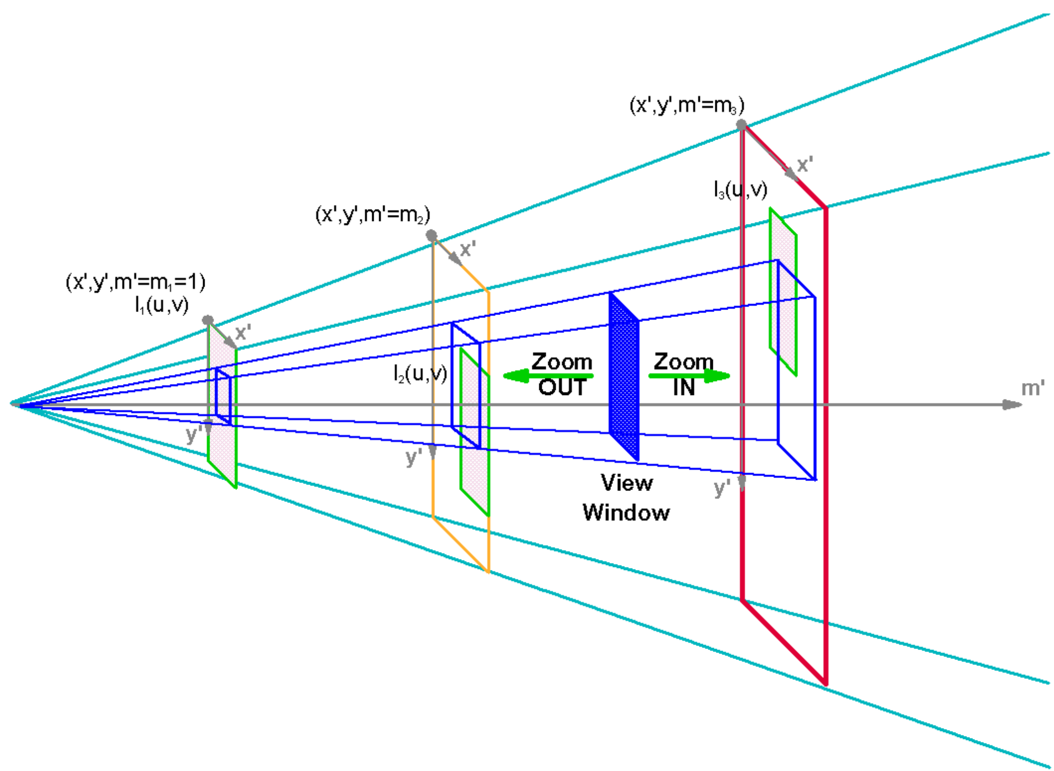


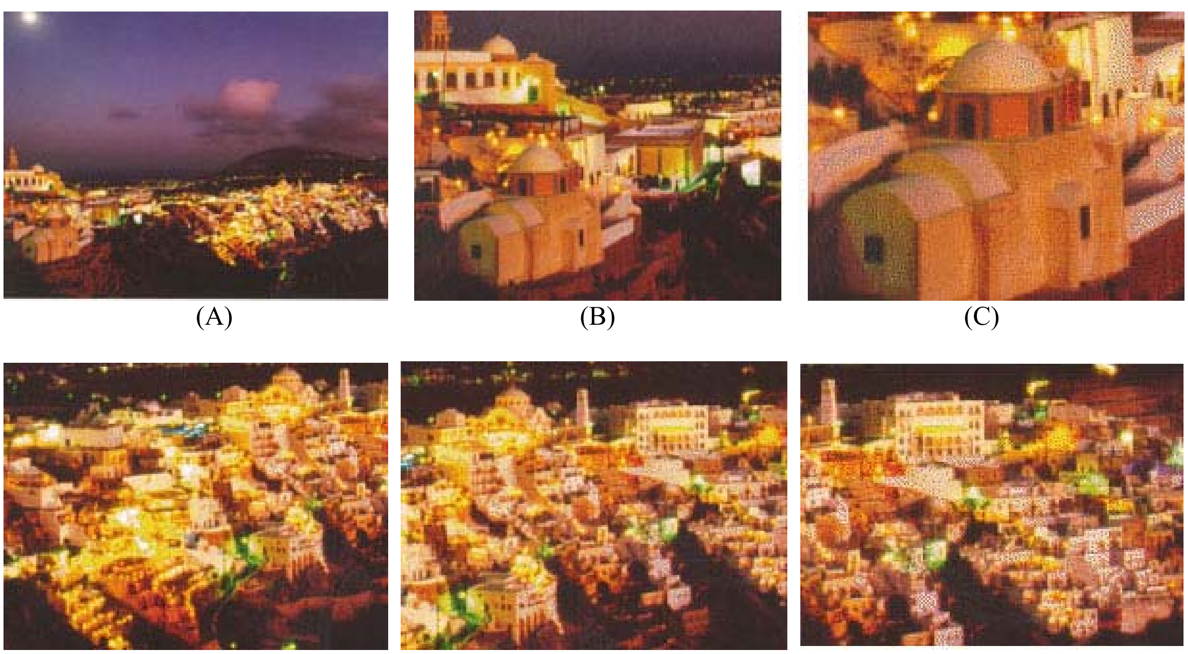
Figure 7: HIP Processing to Determine View Window and Project the View Window.

From the stored HIP representation, a user can select to display any view, size, and resolution of the scene. The four steps for the final image composite includes: (1) determine the view window within the HIP, (2) project the view window onto each image in the HIP, (3) select the appropriate regions on each image, and (4) perform image processing to combine selected regions. As a result, the final combined image composite is display to the user at the selected viewpoint and resolution magnification.

The user selects the desired display view, including magnification zoom and spatial position. The view window parameters for the display is converted with respect to the HIP coordinate system. The view window parameters determine which part of the scene and at what magnification the final image composite is to be displayed. Having determined the view window parameters, the pixels from the input images are used to contribute to the desired display view.

The view window parameters are indexed relative to the HIP coordinate system, and thus also determine the relevant spatial ranges of image region from the input images residing in the HIP. For each input image component, a window of its pixel data is determined to contribute to the final display view. Figure 7 illustrates the view window in the HIP and how the view window is projected onto the magnification levels of the input images. The collection of pixel data from the images residing in the HIP is referred to as the selected image regions. It is these selected image regions and their corresponding positions in the HIP that determines the final output.

The final image composite is determined through image process techniques like interpolation, super-resolution, and possibly stitching of the collections of selected image regions. In addition, the image interpolation stage may include one or a combination of the following processes: up-interpolation, down-interpolation, and bi-interpolation. Subsequently, because aliasing artifacts may be evident when spatial image magnification and reduction are performed, an anti-aliasing image filter may be inserted. Finally, the image composite output is rendered on the display device at the user specified display size, spatial offset, spatial rotation, and resolution magnification.



(A) (B) (C)

(D) (E) (F)

Figure 8: Six Varying-Resolution Digital Images Acquired from the Same Scene.

6 CONCLUSION

We present a framework for authors to capture images of a scene with varying importance and resolution to different regions through the use of a heterogeneous image pyramid representation(HIP). The HIP is a collection of input images captured of the same scene or object, where the images may be of different image sizes, different pixel resolutions, and different spatial locations, but where they have some degree of spatial overlap. The volumetric HIP is thus a non-uniform representation composed of multiple captured multi-resolution images. Through the HIP processing, users can select to display any view, size, and resolution of the scene. The user will be able to see high resolution renderings of important regions and lower resolution for other parts. The HIP representation enables arbitrarily continuous zoom and display of the multi-resolution scene. The HIP also reduces the total data size required to represent an otherwise super-resolution and uniform-resolution scene.

In figure 8, six digital images are acquired of a scene with varying magnification resolutions and at different angular viewpoints. Figure 8A is the lowest resolution capture of the scene covering a wide-angle view of the city Santorini, Greece. The other images are zoomed in to capture the interesting parts to the author. Using the HIP representation and processing, the full heterogeneous image composite is rendered in Figure 9. As a user views this city in a virtual environment from a far distance, the low level input image in Figure 8A is sufficient to render that display. As the user navigates closer to the city, specific parts of the city renders higher in detail as the higher resolution images are used in the final image composite. Furthermore, with the use of optimized hardware and graphics accelerators, the HIP representation can be rendered without special players.

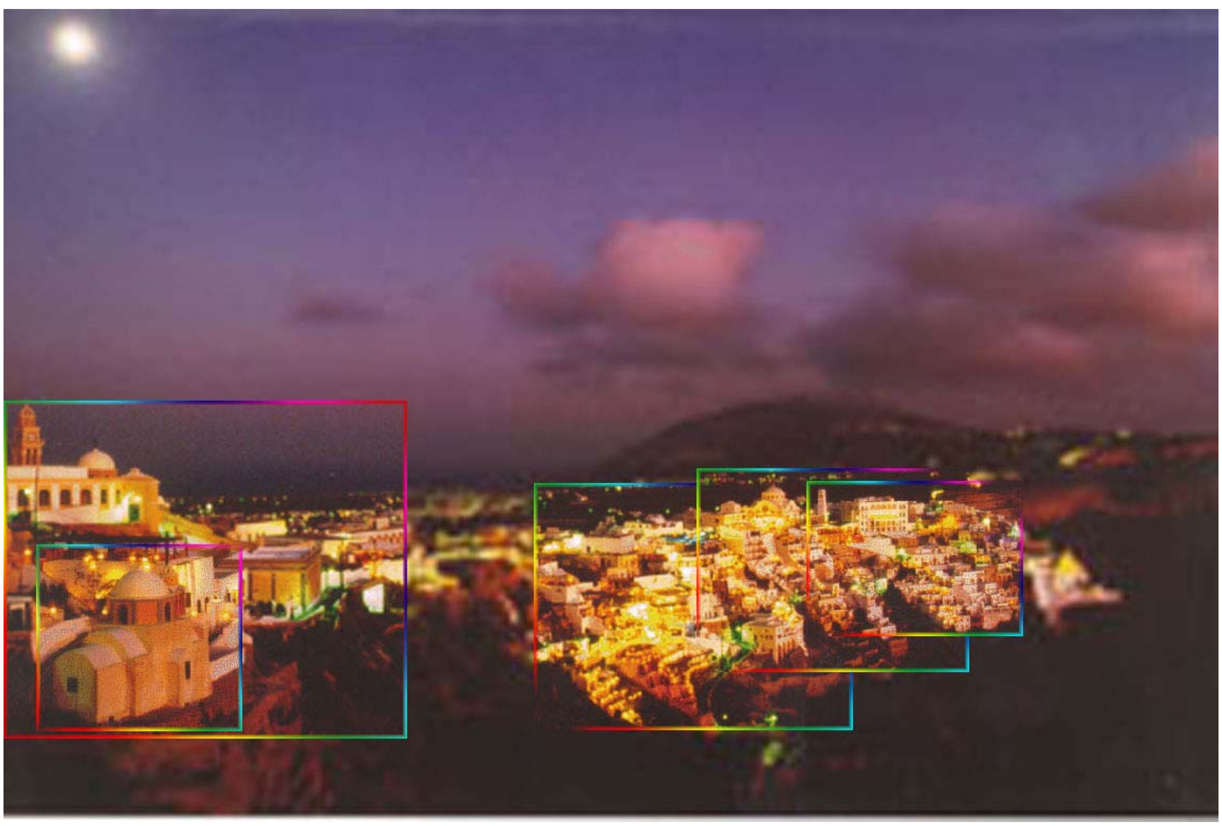


Figure 9: Digital Image Composite Generated from the HIP Processing of the Images Acquired in Figures 8.

7 REFERENCES

[1] S. Baker and T. Kanade, "*Limits on Super-Resolution and How to Break Them*", Proc.CVPR00, 2000.

[2] S. Baker and T. Kanade, "*Hallucinating faces*", Proc. of the Fourth ICAFGR, 2000.

[3] Deborah Berman, Jason Bartell, and David Salesin, "*Multiresolution Painting and Compositing*", Proceedings of the ACM Computer Graphics, SIGGRAPH '94, pp. 85-90, 1994.

[4] Ding-Yun Chen, Murphy Chien-Chang Ho, and Ming Ouhyoung, "*VideoVR: A Real-Time System for Automatically Constructing Panoramic Images from Video Clips*", Proc. of IFIP 1998 Workshop on Modeling and Motion Capture Techniques for Virtual Environments (CAPTECH98), pp. 140-143, Geneva, Switzerland, Nov. 1998.

[5] M. Elad, "*Super-Resolution Reconstruction of Image Sequences - Adaptive Filtering Approach*", PhD thesis, The Technion - Israel Institute of Technology, Haifa, Israel, 1996.

[6] M. Elad and A. Feuer, "*Super-Resolution Restoration of Continuous Image Sequence - Adaptive Filtering Approach*", IEEE Trans. Image Processing, December 1995.

[7] Adam Finkelstein and Marisa Range, "*Image Mosaics*", Lecture Notes in Computer Science, vol. 1375, 1998.

[8] A. Finkelstein, C. E. Jacobs, and D. H. Salesin, "*Multiresolution Video*", Proceedings of the ACM Computer Graphics, SIGGRAPH '96, New Orleans, pp. 281-290, August 1996.

[9] J. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. "Computer Graphics, Principles and Practice, Second Edition." Addison-Wesley, Reading, Massachusetts, 1990.

[10] Lisa Gottesfeld Brown, "*A Survey of Image Registration Techniques*", ACM Computing Surveys, 24(4), pp. 325-376, December 1992.

[11] J. W. Hsieh, H. Y. M. Liao, K. C. Fan, M. T. Ko and Y. P. Hung, "*Image Registration Using a New Edge-Based Approach*", Computer Vision and Image Understanding, Vol. 67, No. 2, pp. 112-130, August 1997.

[12] C. Hsu and R. Beuker, "*Multiresolution Feature-Based Image Registration*", Visual Communications and Image Processing 2000, Proceedings of SPIE vol. 4067, Australia, pp. 20-23, June 2000.

[13] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, "*Efficient representations of video sequences and their applications*", Signal Processing: Image Communication, vol. 8, pp. 327-351, 1996.

[14] R. Kumar, P. Anandan, M. Irani, J. Bergen, and K. Hanna. "*Representation of scenes from collections of images*", Proc. IEEE Workshop on Representations of Visual Scenes, pp. 10-17, 1995.

[15] A. Patti, M. Sezan, and M. Tekalp, "*Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time*", IEEE Trans. on Image Processing, pp. 1064-1078, August 1997.

[16] R. Schultz and R. Stevenson, "Extraction of high-resolution frames from video Sequences", Trans on Image Processing, vol. 5, pp. 996-1011, 1996.

[17] H. Y. Shum and R. Szeliski, "*Panoramic Image Mosaics*", Microsoft Research Tech. Rep. MSR-TR-97-23, 1997.

[18] R. Szeliski, "*Image Mosaicing for Tele-Reality Applications*", Technical Report 94/2, Digital Equipment Corporation, Cambridge Research Lab, June 1994.

[19] A. Zomet and S. Peleg, "*Applying super-resolution to panoramic mosaics*", WACV, 1998.