Wavelet-Based Resolution Enhancement of Omnidirectional Images*

PENG Qimin JIA Yunde
Department of Computer Science
Beijing Institute of Technology, Beijing 100081, P.R. CHINA
yjiar@bit.edu.cn

Abstract

In this paper we present an approach for resolution enhancement of omnidirectional images based on wavelet transform. First, the degradation model of omnidirectional image is given. Second, the resolution enhancement of an image is achieved by using local extrema extrapolation of wavelet coefficients. Then, fusion operation is applied to the coefficients of registered pixels in the enhanced images of an image sequence. Finally, a fine resolution enhancement image is reconstructed via inverse wavelet transform. The experimental results show the proposed approach is feasible and efficient for resolution enhancement of omnidirectional image.

Key Words: omnidirectional image, resolution enhancement, wavelet transform, image fusion

1. Introduction

Omnidirectional cameras have several distinguished merits compared with conventional cameras. Omnidirectional cameras can easily identify slim rotation from translation without ambiguity due to its 360° view field [Nayar,1997][Kang and Weiss,1999]. Especially, the rotation invariant of an omnidirectional image is suitable for surveillance, robot navigation, image-based rendering and video-conferencing [Onoe *et at.*, 1998][Boult, 1999][Gaspar *et at.*, 2000]. Unfortunately, non-uniformity and low resolution of omnidirectional images limit its application. One way to solve this problem is trying to compensate the non-uniform projection with a special CCD array particularly designed for a omnidirectional camera, but it is very expensive. Another way is to enhance resolution with a computational compensation. A lot of work has been done on resolution enhancement and much progress has been reported [Hunt, 1995][Elad and Feuer,1999][Freeman and Pasztor,1999] [Schultz and Stevenson,1996]. Most of existing approaches fall into the following three categories: frequency-based, spatial-temporal-based and hybrid methods [Tekalp, 1995] [Sementilli *et at.*,1993][Lorette *et at.*, 1997]. Resolution enhancement is an ill-posed inverse problem, and a spatial domain approach can conveniently integrate a priori knowledge for the regularization of the problem. As a result, spatial domain approaches are used more often than frequency domain approach. More attention has been paid to the interpolation of non-uniform spaced samples and nonlinear models capable of bandwidth extrapolation [Cyetkovic and Vetterli, 1995][Aizawa *et at.*, 1991].

In this paper, we present an approach to resolution enhancement of omnidirectional images. We integrate single frame enhancement and image fusion from an image sequence based on the addition theorem, spatial-temporal consistency and self-similarity of wavelet transform. The experimental results demonstrate that the proposed algorithm is feasible for resolution enhancement of omnidirectional images.

2. Degradation Model

An omnidirectional camera consists of a catadioptric mirror and a CCD camera. The mirror shape can be spherical, parabolic, and hyperbolic [Ollis *et at.*, 1999].

The imaging procedure can be treated as the combination of a projection from a scene to the surface of the mirror and a reflection from the mirror to the CCD array. Both are nonlinear projections (see Fig. 1), and reflection results in a non-uniform sampling.

The non-uniform resolution of omnidirectional images can be verified with a simple calculation: corresponding to the same ken, the inside track of the image which contains fewer pixels has lower resolution in comparison with the exterior track of the image.

Let $x_{sc}(m_1,m_2)$ be a scene containing $m_1 \times m_2$ visible points, and $x_o(n_1,n_2)$ the omnidirectional image with size of $n_1 \times n_2$ The procedure of imaging can be written as

$$x_o(n_1,n_2) = T_r T_p x_{sc}(m_1,m_2) + n(n_1,n_2) \quad (1)$$

*This work is partially supported by Chinese Natural Science Foundation K60075005

where T_r and T_p represent projection transform and reflection transform, both being nonlinear functions corresponding to the surface equation of the catadioptric mirror.

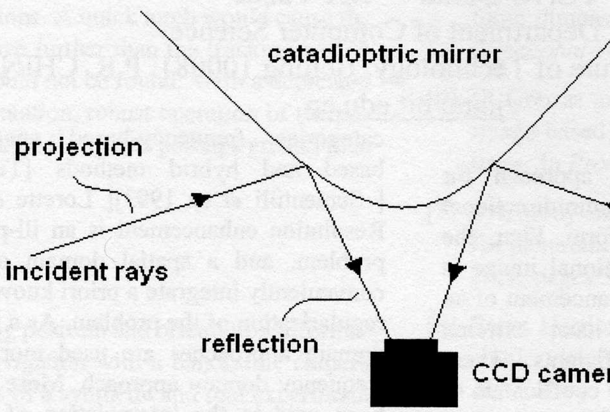

Figure 1. Projection and reflection

Along the radial direction of the omnidirectional image, a pixel can be given as

$$x_o(p,q) = x * h = \iint_{\Omega} x_{sc}(u,v)h(p,q)dxdy \tag{2}$$

where h(p,q) is a shift-variant spatial Point Spread Function (PSF) of the camera, is a region on the mirror surface corresponding to a point in the image.

The discrete form of Equation (2) can be rewritten as:

$$x_o(p,q) = \sum_{i\in\phi(p)} \sum_{j\in\varphi(q)} x_{sc}(i,j)h(p(i),q(j)) + n(p,q) \tag{3}$$

where $\phi(p)$ and $\varphi(q)$ define a region in the scene, in which all of visible points that are projected to the pixel $x_o(p,q)$ are related through the equation of mirror, parameters of the CCD camera and the pixel location. n(p,q) is additive noise. It has been proved that resolution increase nonlinearly along the radial direction from the center to exterior in an omnidirectional image. The omnidirectional image can be treated as a seriers of concentric circles. Those pixels located on the same concentric circle have the same spatial resolution, sequentially. Resolution is consistent in each row and variant in each column when it is expanded to cylinder image. Therefore, it is the nonlinear catadioptric mirror that introduces the resolution degradation of omnidirectional image and makes resolution enhancement more difficult. This problem is difficult to overcome by a single image without special CCD arrays. Thus we propose an intuitive algorithm of image fusion from an image sequence aimed at reconstructing a uniform resolution image with more details of a scene from the image sequence. Note that each incident ray from a scene visible point to the focus of catadioptric mirror, passing through the cylinder around the omnidirectional camera, uniquely corresponds to a pixel in the image, and resolution enhancement can be implemented on the expanded cylinder image of its original image.

When the relative position between the scene and the omnidirectional camera is changed, the visible point might be located at the different area on the surface of catadioptric mirror, which probably causes the resolution change as the change of its neighbor size. Compared with the resolution enhancement of conventional cameras, the resolution enhancement of omnidirectional image is able to achieve more compensation from an image sequence. Each frame in the sequence might contain additive useful information of the scene due to its space-variant resolution and relative movement between the camera and the scene. In order to get a uniform resolution and an enhanced image, the procedure can be completed in three steps. First, enhance each image at consistent resolution using local extrama extrapolation based on wavelet transform. Secondly, fuse images from a sequence to obtain a new image that contains more detail of the surrounding scene. Finally, resample the cylinder image at desired grid to get a series of uniform resolution images. However, if there is no relative movement, it will degenerate to regular resolution enhancement that can be done with conventional methods [Borman and Stevenson, 1998].

3. Single Frame Enhancement

It is well known that reconstruction of a high-resolution image from low-resolution images is an ill-posed inverse problem. To solve this problem, necessary constraints must be provided as prior knowledge to stabilize the solution. For high-resolution reconstruction from a omnidirectional image, this problem becomes more serious.

Although the problem can be regularized in principle, it is time consuming because of nonlinearity and resolution varietion. Besides spatial integration modeled by a shift variant PSF, original omnidirectional images also suffer from aliasing due to sub-Nyquist sampling and relative motion between the scene and the camera during the image capture. For simplicity, we assume that scene is stationary.

Consider decomposing an image f(x,y) with wavelet basis, the approximation at a resolution 2^j is characterized by the inner products

$$f^j[u,v] = < f(x,y), \phi_{j,u}(x)\phi_{j,v}(y) > \quad (4)$$

where $\phi_{j,u}(x)\phi_{j,v}(y)$ are orthogonal bases of space V_j of dilated separable scaling functions

$$\phi_{j,u}(x)\phi_{j,v}(y) = \sqrt{2^j}\phi(2^j x - u)\sqrt{2^j}\phi(2^j y - v) \quad (5)$$

The high frequency components of wavelet transform show self-similarity among different scales, as shown in Fig 2. Unfortunately, the wavelet coefficients cannot be used directly in practice because they are not translation invariant. Therefore, the wavelet extrema can be translated without being modified [Mallat and Zhong, 1992].

The local regularity at location (x_0,y_0) can be quantified with Lipschitz exponent $\alpha (0 \le \alpha \le 1)$:

$$|f(x,y) - f(x_0,y_0)| \le K(|x-x_0|^2 + |y-y_0|^2)^{\alpha/2} \quad (6)$$

where K is a non-zero constant, and a characterizes the type of singularity at that location. The larger the a is, the more regular that location is, but not differentiable except at zero, implying the discontinuity. The local extrema of wavelet coefficient modulus correspond to the singularity at that location and can be propagated across the scales. Our approach based on local extrema extrapolation makes use of this propagation property.

According to the theory of Multi-Resolution Analysis (MRA) proposed by S. Mallat, low-frequency component A_j at scale j is decomposed into two parts, high frequency component D_{j+1} and low frequency component A_{j+1}. The original image can be reconstructed with inverse wavelet transform on these components. It is known that some of high frequency components are lost during the image formation by the degradation model, i.e., D_1 is lost during the image formation, so the problem turns into one of estimating high frequency component $\hat{D}_1$ from existing image.

As mentioned in section 2, the compensation of $\hat{D}_1$ should vary spatially depending on the image coordinates. Because the highest scale is the most sensitive to noise, we find extrema at the second scale and extrapolate them to the highest scale in proportion of the ratio determined by the a certain criterion [Baker and Nayar, 1998]. An approximate choice is ratio = (n-m) / n, where n represents the total number of rows in the expanded cylindrical image, and m is the current row number.

There is an alternative scheme. In this method, each pixel in an omnidirectional image is mapped to an actual row in the expanded cylindrical image but not in column. Each row is treated as a wave and is enhanced in a similar way using 1D extrema extrapolation. Once all of the rows have been enhanced and resampled at a uniform resolution, bilinear interpolation is applied to each column. This method needs extra global constraints to ensure the fidelity of the result with respect to the original image.

Because most of the extrema will be removed by smoothing at larger scales, and the remainder's location might be changed. Some constraints should be observed in estimating $\hat{D}_1$:

(1) To ensure that the reconstructed image is consistent with the original, the local extrama should be determined by equation (6), where constant K can be computed from the local extrema at a larger scale using optimization method.

(2) To maintain original high frequency, edge information should be used to selectively suppress false high frequency components in the estimated $\hat{D}_1$.

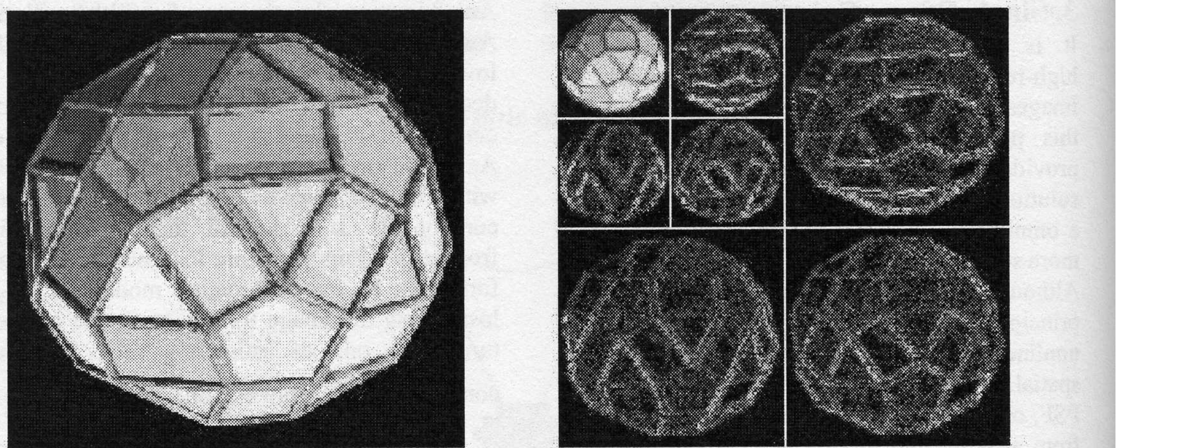

Figure 2. Self-similar of wavelet coefficients

Thus, the approach is to find local extrema in the high frequency components of the original image and then extrapolate them at fine scales to estimate $\hat{D}_1$.

The reconstruction can be described as

$$\hat{f}(x,y) = A_n \tilde{\phi}_{n,u}(x)\hat{\phi}_{n,v}(y) + \sum_{j=2}^{n} D_j \tilde{\phi}_{j,u}(x)\tilde{\phi}_{j,v}(y) + \hat{D}_1 \tilde{\phi}_{1,u}(x)\tilde{\phi}_{1,v}(y) \quad (7)$$

The result should satisfy the following optimization criteria:

$$\hat{f}(x,y) = \underset{f^i(x,y)}{\arg\min}\{\|f_o(x,y) - Pf^i(x,y)\|^2 + \lambda\|Ff^i(x,y)\|^2\} \quad (8)$$

where $f_o(x,y)$ is the original omnidirectional image, $Pf^i(x,y)$ represents mapping cylinder image to the omnidirectional image, and $Ff^i(x,y)$ a high-pass filter to the *ith* result of following the iterative algorithm:

$$f^i(x,y) = f^{i-1}(x,y) + \alpha(E(f(x,y) - f^{i-1}(x,y))) \quad (9)$$

The constraint function is composed of two parts. One is the error denoted by $\|f(x,y) - Pf^i(x,y)\|^2$ whose minimization ensures the consistency of the enhanced image to the original, and the other is a stabilizing function denoted by $\|Ff^i(x,y)\|^2$ whose minimization suppresses high frequency components caused by noise amplification. This function can be weighted by a regularization parameter λ which controls the fidelity to the original image and the smoothness of the reconstruction.

The resolution enhancement of a single frame can be summarized as following:

1. Expand the omnidirctional image to a cylindrical image.
2. Decompose the cylindrical image with the wavelet transform.
3. Extrapolate the local extrema from the lower scales to 1st scale under the edge constraints in proportion of its position.
4. Iteratively refine the result, according to Equations (8) and (9).
5. Reconstruct the image from the final coefficients with the inverse wavelet transform.

4.Enhancement from an Image Sequence

The resolution enhancement algorithm from one image can improve the quality of the image to some extent. However, many of the high frequency components cannot be reconstructed reliably by this method. This problem can be solved with an image sequence. Each frame in the sequence contains additive useful information from the relative movement between a scene and a camera, especially for a omnidirectional camera with a non-uniform resolution. Integration of all relevant high-resolution regions spreading among the sequence is an intuitive solution.

Consider the fact that the wavelet extrema are translated without being modified when the image is translated. Therefore, wavelet transform can be used in image fusion. If we select one frame as the reference frame, all of the useful information contained in the other frames in the sequence can be integrated into this frame. After the fusion operation is completed, resample this enhanced frame to get the image with a uniform resolution.

To implement image fusion, registration should be done at first. This is easy to do with multi-scale images calculated by wavelet transform. In view of the fact that these frames are captured in succession, the candidate area is limited in a local region and is roughly located at a larger scale. It can therefore be refined at a subpixel position using matching methods [Jia, 1999]. Tracking a point through the sequence using matching on the second scale between two adjoining frames, we can find the finest position and then replace the original coefficients at the first scale with its. Smoothness is adopted as a global constraint to ensure the correct of certain point tracking due to noise, nonlinear distortion and occlusion. Additionally, because the pixels located in the higher resolution region will gain few improvements, only those pixels fall into lower resolution region need processing.

This algorithm is able to precisely find the corresponding area without considering the restriction of nonlinear distortion. Once the corresponding location has been determined, the images can be fused at the first scale of wavelet transform.

Image fusion algorithm can be described as the following iteration process:

$$f^i(x,y) = \delta_c(x,y) f_c(x,y) + (1-\delta_c(x,y)) f^{i-1}(x,y) \qquad (10)$$

where $f^i(x,y)$ is the corresponding region in the *ith* iteration, $f_c(x,y)$ the corresponding region in the processing frame, $\delta_c(x,y)$ a discriminate function at location (x,y) of the *ith* frame being fused calculated in advance and defined by

$$\delta_c(x,y) = \begin{cases} 1 & \text{it is the furthest to the image centre than other corresponding areas} \\ 0 & \textit{Otherwise} \end{cases} \qquad (11)$$

Enhancement from an image sequence can be summarized as follows:

1. Enhance each frame in the sequence using the method proposed in Section 3.
2. Track the pixels in the key frame through the sequence and calculate $\delta_c(x,y)$ for each frame.
3. Select an unprocessed frame in the sequence and mark it as processed.
4. Fuse I_r and I_i based on Equation (10).
5. Repeat Step 3 to 4 until no frame in the sequence is marked unprocessed.
6. Recover a fine image of the surrounding scene from the final *Ir* through inverse wavelet transform.

5.Experimental result

We first tested our algorithm with simulated images. An image is degraded in term of omnidirectional projection. Then, it is restored using bilinear interpolation and proposed extrema extrapolation respectively and their PSNRs are computed. Fig 3 is the result of simulation experiments. The lower image in the right column is obtained from bilinear interpolation and the upper image is its enlarged portion (PSNR=15.9119). The middle column is the result of proposed method (PSNR=16.8005). It can be seen from the contrast that our schemes are superior to bilinear interpolation.

Fig.4 illustrates the omnidirectional camera developed in our Lab and one of the omnidirectional images captured by the camera. A sequence with 10 frames of 768×576 pixels is acquired for testing. The 5th frame is selected as the reference frame. An image of the same size will be reconstructed from the selected frames to demonstrate the performance of the proposed approach.

We processed the expanding cylinder image with several wavelets and the results show that it is better using bi-orthogonal and symmetrical wavelets than irregular wavelets. If there is no edge information to rectify the extrapolation of high frequency scales, false high frequency redundancy will result, especially at the edges.

The expanded cylindrical image of 5th frame using bilinear interpolation and a zoomed part of it are shown in Fig 5 and Fig.6 respectively. Fig.7 is the same part using the method proposed in section 3. Fig. 8 is the final result of the part after sequence fusing. Compared with Fig. 6 and Fig.7, it is clear that the high frequency components are dramatically increased, especially in the lower resolution regions of the original image and the enhanced image maintains maximal consistence to the original

Figure 3. **Simulation experiments results**

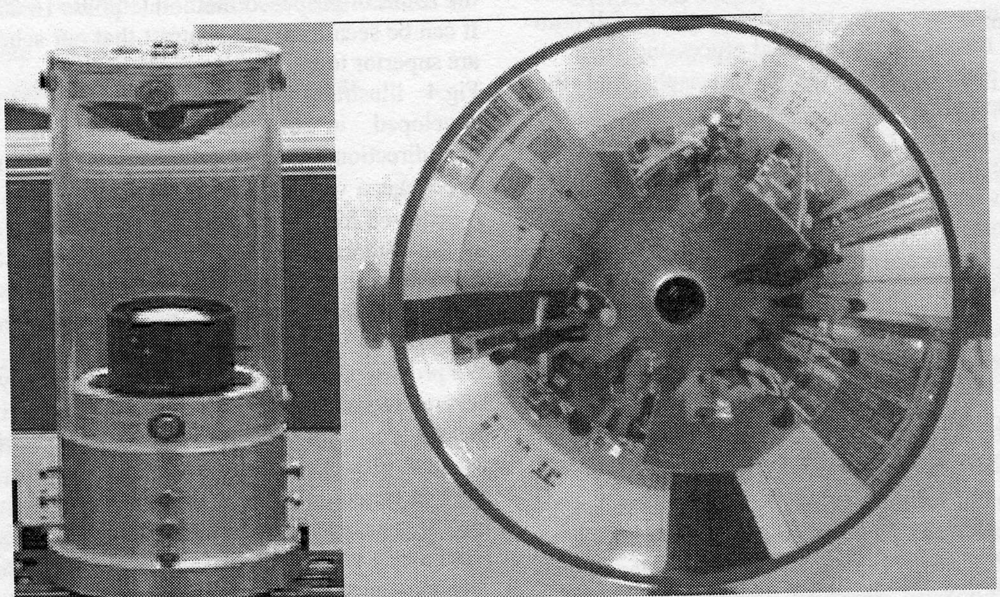

Figure 4. **Omnidirectional camera and omnidirectional image**

Figure 5. **5^{th} frame expanded using bilinear interpolation**

Figure 6. The result of bilinear interpolation

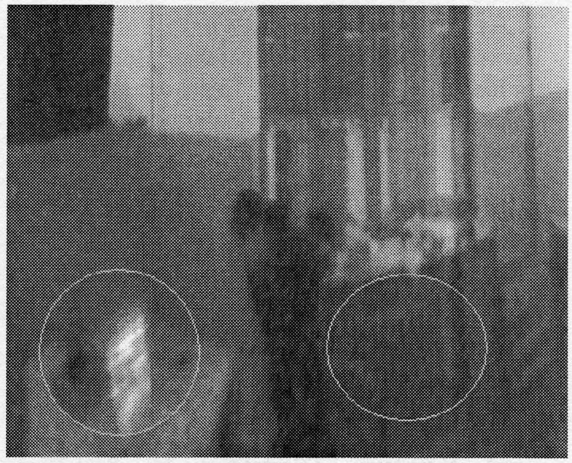

Figure 7. The result of single frame enhanced

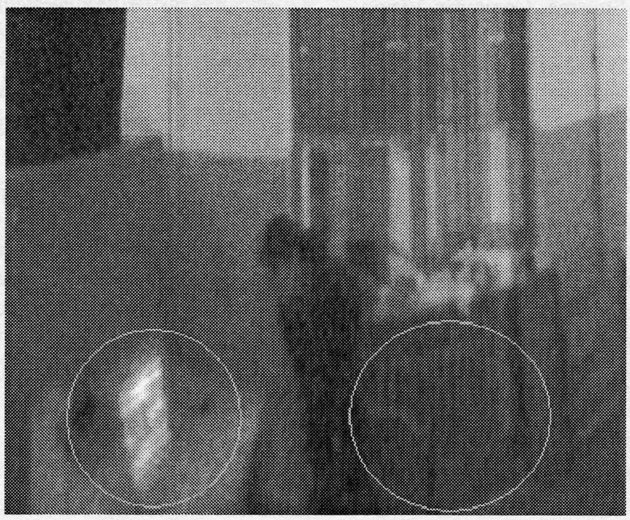

Figure 8. The final result after fusing operation

6. Conclusions

This paper proposes a wavelet-based approach for resolution enhancement of omnidirectional images. There exists a unique corresponds between incident rays and pixels in the expanded cylinderical image, the enhancement operation can be manipulated on the expanded image without introducing into irrelevant ambiguities. Wavelet transform can be selected as appropriate tools to implement these operations. Combined with image fusion from a sequence, extrapolation of high frequency scales under edge information constraints can reconstruct a uniform and resolution enhancement image based on the selected reference frame.

A sequence images captured while dramatically changing the orientation of the omnidirectional camera will lead to a remarkable resolution improvement. Provided the limited relative movement, if the orientation of the camera change from one to its opposite, the enhancement effect reaches its best. A long sequence always produces a better result but it takes more time either.

The future work is to use some adaptive strategies for selecting optimal constraint control parameters, including the regularization parameter λ and a. For the real-time applications, a more efficient method should be substituted for the proposed iterative method. If there are several moving objectives in the scene, optical flow must be introduced into the fusion operation to resolve correspondent among pixels in the frames in order to maintain the local and global constraints.

7. Acknowledgements

We would like to thank Professor Hong Zhang for his valuable comments. This work is partially supported by Chinese Natural Science Foundation K60075005.

References

[Nayar, 1997] S.Nayar. Catadioptric Omiirectional Camera. In CVPR, pages 482-488,Puerto Rico,June,1997.

[Kang and Weiss, 1999] S.B.Kang and R.Weiss. Characterization of Errors in Compositing Panoramic Images. CVIU, 73(2), 269-280, Feb, 1999.

[Onoe *et at.*,1998] Y.Onoe,K.Yamazawa, H.Takemura and N.Yokoya: Telepresence by Real-Time View-Dependent Image Generation From Omnidirectional Video Streams. Computer Vision and Image Understanding, Vol 71, pages 588--592, 1998.

[Boult, 1999] T. Boult. Frame-Rate Omnidirectional Surveillance and Tracking of Camouflaged and Occlude Targets. In Second Workshop of Visual Surveillance at CVPR, pages 48-58, 1999

[Gaspar *et at.*, 2000] J.Gaspar, N.Winters and J.Santos-Victor. Vision-Based Navigation and Environmental Representations with an Omnidirectional Camera. In IEEE Transactions on Robotics and Automation, Vol 16(6) , pages 890--898, 2000.

[Hunt, 1995] B.R.Hunt. Super-Resolution of Images:Algorithms,Principles,Performance. Int. J. Imaging Sys. And Tech., Vol 6, pages 297-304,1995.

[Elad and Feuer, 1999] M.Elad and A.Feuer. Super-Resolution Reconstruction of Image Sequences. IEEE Trans. PAMI 21(9), 1999

[Freeman and Pasztor, 1999] W.Freeman and E.Pasztor. Learning Low-Level Vision. In ICCV'99, 1999

[Schultz and Stevenson, 1996] R.Schultz and R.Stevenson. Extraction of High Resolution Frames from Video Sequences. Trans. On Im. Proc.,5(6), pages 996-1011,1996

[Tekalp, 1995] A.Murat Tekalp. Digital Video Processing. Prentice Hall, PTR, pages 331-345, 1995

[Sementilli *et at.*, 1993] P.J.Sementilli,M.S.Nadar and B.R.Hunt. Poisson MAP Super-Resolution Estimator with Smoothness Constraint. Neural and stochastic methods in image and signal processing. Vol 2032 of Proc. Of the SPIE, pages 2-13,San Diego, CA, July ,1993.

[Lorette *et at.*, 1997] A.Lorette ,H.Shekarforoush and J.Zerubia. Super-Resolution with Adaptive Regularization. Proc of the IEEE International Conference on Image Processing, Vol I,pages 169-172,Santa Barbara,CA,Oct,1997

[Cyetkovic and Vetterli, 1995] Z.Cvetkovic and M.Vetterli. Discrete-Time Wavelet Extrema Representation:Properties and Consistent Reconstruction,IEEE Trans. on Sinal Processing, March 1995.

[Aizawa *et at.*, 1991] K.Aizawa,T.Komatsu and T.Saito. Acquisition of Very High Resolution Images Using Stereo Cameras. Visual communications and Image Processing of Proc. Of the SPIE, Vol 1605,pages 318-328,1991

[Ollis *et at.*, 1999] Mark Ollis, Herman Herman and Sanjiv Singh. Analysis and Design of Panoramic Stereo Vision Using Equi-Angular Pixel Cameras. The Robotics InstituteCarnegie Mellon University, CMU-RI-TR-99-04, Pittsburgh, PA 15213 January 1999.

[Borman and Stevenson, 1998] Sean Borman and Robert Stevenson. Spatial Resolution Enhancement of Low-Resolution Image Sequences A Comprehensive Review with Directions for Future Research. University of Notre Dame,Notre Dame,IN 46556, July 8,1998

[Mallat and Zhong, 1992] S.Mallat and S.Zhong. Characterization of Signals from Multiscle Edges. IEEE Trans,PAMI Vol 14, pages 710-732,July,1992.

[Jia, 1999] Yunde Jia: Machine Vision. Science Press, pages 247- 249,1999.

[Baker and Nayar, 1998] Simon Baker and Shree K.Nayar. A Theory of Single-Viewpoint Catadioptric Image Formation. Proceedings of the 6th International Conference on Computer Vision.,Bombay, January, 1998.