MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

8—8

Distortion-Free Navigation of Omni-Directional Images Using Constructive Neural Networks

Frank Ekpar *
Faculty of Engineering
Toyama University

Masaaki Yoneda†
Faculty of Engineering
Toyama University

Hiroyuki Hase‡
Faculty of Engineering
Toyama University

Abstract

In this paper, we present a novel approach to the generation of arbitrary perspective-corrected views from omni-directional images for the purpose of interactive navigation of the images in real time. We use a constructive neural network of suitable complexity to learn the inherent distortion of the omni-directional imaging system using training sets obtained through carefully constructed calibration patterns. Starting with a near-minimal neural network, the topology of the neural network is modified automatically over successive training cycles until a reasonable, near-optimal network is obtained. Our system overcomes the limitations of previous methods by obviating the need to derive the perspective projection equations of the omni-directional imaging system. Additionally, our system is robust enough to correctly approximate an omni-directional imaging system of arbitrary complexity yet elegantly simple enough to permit real-time correction for most omni-directional imaging systems currently available. We demonstrate the practicability of our approach by describing its application to the generation of arbitrary perspective-corrected views from fish-eye images.

1 Introduction

Recently, there has been unprecedented interest in research and application development in the fields of immersive (omni-directional) panoramic imaging, telepresence and virtual reality. Immersive panoramic imaging applications permit a user to immerse himself in a recording of a remote scene with the ability to look in any desired direction in the scene, limited only by the field of view of the imaging system employed. Immersive telepresence applications permit live omni-directional viewing of remote locations while immersive virtual reality generally refers to the synthesis of immersive content. A very popular example of an immersive imaging system is Apple Computer's QuickTime Virtual Reality System. Immersive panoramic video applications [1, 2, 3] that allow a viewer to interactively select an arbitrary portion of the live or recorded video for viewing have been developed. Such applications have the potential to define the future of video and television and to open up new vistas in virtual reality, surveillance and machine vision. Consequently, their importance cannot be over-emphasized. Immersive imaging systems typically utilize wide-angle lenses, such as the fish-eye lens, and mathematical models of wide-angle lenses. Wide-angle lenses, however, introduce significant amounts of non-linear geometric distortion into the images they produce and so for these images to be viewed comfortably by a human being, they need to be de-warped. Although methods for generating perspective-corrected views from fish-eye images already exist, there is (to the best of our knowledge) as yet no practical general-purpose method for generating perspective-corrected views from an arbitrary wide-angle image. Accordingly, it is the goal of this paper to present a practical method for generating perspective-corrected views from an arbitrary wide-angle image without the need for knowledge of the projection equations of the lens with which the image was obtained.

2 Brief Description of Our Approach

Our approach is to consider the formation of a distorted 2D image from the undistorted 3D real world scene by a wide-angle lens as an input-output mapping between an undistorted 2D image plane and the distorted 2D image plane. We then construct a neural network of sufficient complexity to learn the char-

*Address: 3190 Gofuku, Toyama-shi, Toyama, 930-8555 Japan. E-mail: frank@pat.iis.toyama-u.ac.jp
†Address: 3190 Gofuku, Toyama-shi, Toyama, 930-8555 Japan. E-mail: yoneda@pat.iis.toyama-u.ac.jp
‡Address: 3190 Gofuku, Toyama-shi, Toyama, 930-8555 Japan. E-mail: hase@pat.iis.toyama-u.ac.jp

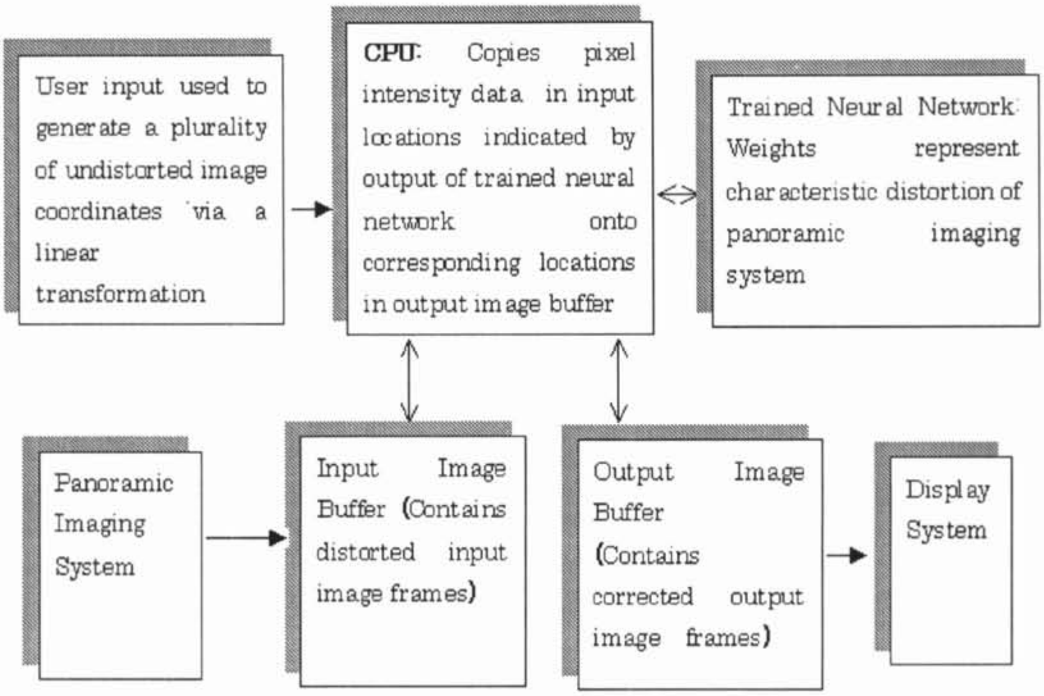


Figure 1: Omni-directional Image Navigation System based on our distortion-correction method. USER INPUT consists of position on input omni-directional image, size of corrected output image frame and zoom level.

acteristic distortion of the wide-angle lens, train the network on sample data and then use the trained network to correct all or selected portions of the wide-angle image. Figure 1 illustrates an application of our method. Here, we exploit the general approximating ability of a multi-layer neural network consisting of perceptron processing units [4]. In presenting an application of our approach in the next section, we show how the problems of constructing a suitable experimental setup for training and validation data gathering and determining the optimum neural network topology are resolved.

3 Application of Our Method

As a demonstration of the effectiveness of our proposal, we have applied the approach outlined above to the construction of perspective-corrected views from a fish-eye image.

3.1 Experimental Setup for Obtaining Training Data Set

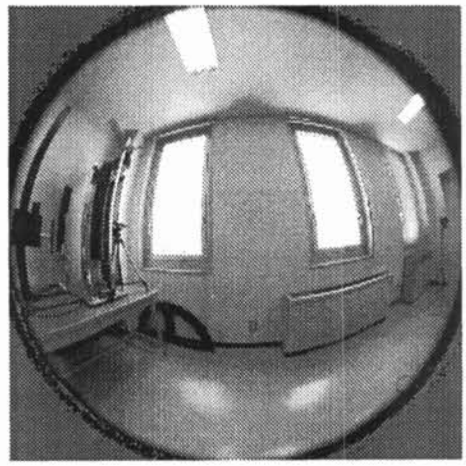
We use a planar grid of equally spaced black dots on a white background as our source of sample data. The positions of the dots are tabulated for use as input data for the network. If a video camera with high enough resolution is used, enough sample points to cover the entire problem domain could be imaged at once. By constraining all sample data points to a plane and aligning this plane with the optical axis of the fish-eye lens, we can overcome the need to deal with 3D data as all data points will have a uniform third component. To align the plane containing the sample data points with the optical axis of the fish-eye lens, we use a method similar to that described in [5]. First, with the iris of the video camera closed to avoid damage to the CCD, a weak coherent light source, such as a low power laser is directed at a video camera to which the fish-eye lens has been attached through a thin plate with a narrow slit in front of the center of the lens. Part of the laser beam is reflected onto the plate with the slit. The alignment of the camera and laser source is then adjusted until the laser beam reflects onto itself. This occurs when the optical axis of the lens coincides with the laser beam. The iris is then opened and a very bright dot appears on the display. The dot now appearing on the display is the optical center of the fish-eye lens. The plane containing the sample points is then aligned to be perpendicular with the optical axis of the lens. Optionally, the center of the sample grid could be made to coincide with the optical center of the lens. The positions of the dots on the image generated by the fish-eye lens are then tabulated for use as reference responses in training the neural network. Linear scaling of the data based on the distance between the sample grid and the fish-eye lens, the resolution of the video camera and other relevant camera data could be carried out.

3.2 Network Training and Dynamic Modification of Network Topology

Our goal is to find a near-optimal network that gives a reasonable approximation to the characteristic distortion of the fish-eye lens. To achieve this goal, one possibility is to start off with a minimal network containing one input layer and one output layer but no hidden layer and to progressively add hidden neurons in such a way that each hidden neuron is connected only to the inputs and its neighbors. This method, cascade correlation, is well documented in [6] and tends to generate rather deep networks. To test our approach, we started off with a near minimal three-layer perception network with the following topology: 2 input neurons, one each for the X and Y coordinates of the undistorted destination image plane; 1 hidden neuron; 2 output neurons, one each for the X and Y coordinates of the distorted source image plane. Since the outputs are independent of each other, it is not strictly necessary to have two output neurons but we have retained this architecture for the convenience of training only once for both the X and Y coordinates.

4 Simulation Results

For the test, we used a Nikon 8mm fish-eye lens attached to a Nikon CoolPix 950 digital camera. A sample grid containing 400 dots was used to gather the training and validation data. The network was then trained using back-propagation with the delta learning rule and applying meiosis (node splitting) until a mean squared error of about 0.08 was obtained. The resultant network had 20 hidden neurons and performed well on real images, as can be seen in Figure 2 below. Sigmoid activation functions



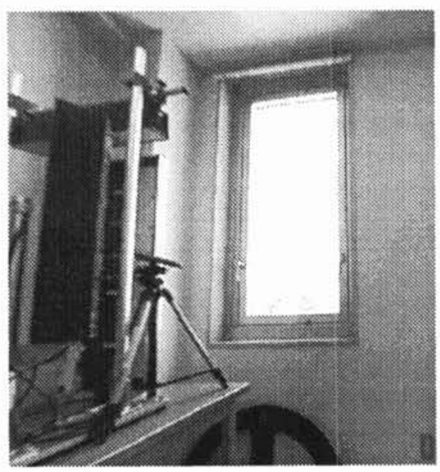


Figure 2: Perspective correction of an arbitrarily selected sub-area of the fish-eye image. To the top of the figure is the distorted input fish-eye image while to the bottom is the corrected view of a selected portion of the input image. Using our method, any arbitrarily selected region of interest can be corrected. The entire input image can also be corrected at once. Our method works for images captured using an arbitrarily complex wide-angle imaging system.

were used in both the output and hidden layers. The learning rate used was 0.1.

5 Conclusion

A practical method of generating perspective-corrected views of omni-directional images using constructive neural networks that obviates the need to use perspective projection equations (which usually apply only to the ideal case) was presented. The method was then applied to the generation of arbitrary perspective-corrected views of a fish-eye image with good results. The size of the network needed to learn the distortion of the panoramic imaging system could be further reduced by partitioning the image and using smaller networks to learn the characteristic distortions of the resulting partitions. The method presented in this paper could be used in surveillance, interactive immersive televison, robot control and virtual reality applications. Further studies on the use of neural networks with lateral connections [7] providing good global and local approximation as well as cascade correlation networks are currently underway.

Acknowledgements

The authors wish to thank Mr. Shinya Yamauchi, Mr. Tadashi Doi and the other employees of Tateyama R&D for their invaluable support of this research effort.

References

[1] Onoe M. and Kuno Y., "Digital Processing of Images Taken by Fish-eye Lens," , 6th International Conference on Pattern Recognition, pp.105-108 (1982).

[2] "Wide-angle Image Dewarping Method and Apparatus," U.S. Patent No. 6,005,611

[3] "Omniview Motionless Camera Orientation System," U.S. Patent No. 5,185,667

[4] Rumelhart D. E., Hinton G.E. and Williams R.J., "Learning Internal Representations by Backpropagating Errors," Nature, Vol. 332, pp. 533-536 (1986).

[5] Chang Y. L., Lebegue X. and Aggarwal J.K., "Calibrating a Mobile Camera's Parameters," Pattern Recognition, Vol. 26, No. 1, pp. 75-88 (1993).

[6] Fahlman S. E. and Lebiere C. "The Cascade Correlation Learning Architecture." In Touretzky D.S., editor, Advances in Neural Information Processing Systems (Denver, 1989); (2), pp. 524-532. Morgan Kaufmann, San Mateo (1990).

[7] Kothari R. and Agyepong K., "Controlling Hidden Layer Capacity through Lateral Connections," Neural Computation, Vol. 9, No. 6, pp. 1381-1402 (1997).