Recognition of 3D Compressed Images and its Traffic Monitoring Applications

Nicole Sharlene Love

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: With the digital image network for traffic monitoring, a large number of surveillance TV cameras are connected to control centers through a hierarchical network. Compressed image data and recognition results are transmitted in the network. In conventional approaches each control center receives compressed image data along with preliminary recognition results from lower level control centers or surveillance cameras. Each center needs to decompress the image data for further processing and if necessary send the compressed image data and recognition results to the upper-level control center.

Motivation: To increase the cost-efficiency of the digital image network, we propose to eliminate the decompression required at each control center by developing a recognition method which works in the compressed image domain without decompression. Conventional image compression methods such as Discrete Cosine Transformation is based on spatial frequency which makes it difficult to carry out recognition processes in the compressed domain. The compression method we use will retain attributes relevant to image compression and recognition.

Previous Work: A contour-based compression method has been developed including color and intensity attributes [3]. Preliminary experiments indicate that a 256x256 color image can be compressed from 24 bits/pixel to 0.1-0.3 bits/pixel. A computationally efficient stereo vision algorithm which produces an edge depth map has also been developed [1]. The algorithm uses three images to reduce the number of false matches and finds the depth with sub pixel accuracy. The stereo vision and contour-based compression algorithms are combined to produce 3D compressed image data.

Approach: This approach uses attributes which are relevant both for compression and recognition. The system consists of three modules: 3D image acquisition, compression, and recognition. The 3D compression algorithm begins with a 3D image acquisition. 3D information is used to distinguish objects in a scene. Conventional 2D compression algorithms are not conducive to detection of partially occluded objects, but the 3D information along with color allows the system to detect partially occluded objects.

Stereo vision algorithms have used two or more cameras. The two camera approach produces correspondence errors. Our system uses three cameras. The third camera is used to reduce the number of missed correspondences. Additionally, our system [1] uses feature correlation rather than area correlation. Feature correlation is faster due to the decrease in data and the decrease in calculations in the algorithm. The 3D image acquisition is based on a trinocular vision algorithm [1] which produces an edge depth map of the center image. The images are acquired simultaneously from all three cameras. The cameras are equally spaced, with their optical axes aligned. The first step is to generate the vertical edge gradient for each image. Edge points in the left and right images are matched using the center camera to help eliminate false correspondences.

The goal of the compression algorithm is two-fold; (1) to compress the data as much as possible without losing any of the relevant information, and (2) to provide a representation of the data which is conducive to object recognition. The decision was made to use a contour based lossy compression method based on Mizuki's algorithm [3]. Mizuki proposed a 2D compression algorithm, which has been extended to 3D in this research. The algorithm produces a high compression ratio and enhances the recognition component of the system.

Many image processing schemes use the edge information as a guide in object recognition, for this reason a contour based algorithm was used [2]. The contours provide a skeleton of the contents of the image which is used for recognition of objects as well as compression. The algorithm focuses on retaining information relevant to recognition, such as contour, color, and distance attributes. A description of the components are as follows:

Depth Map The depth map contains the distance of edge pixels from the center camera.

Contour Coding The contours are determined by tracing edges with similar color and distance information. The contours are then coded by using the start location and the directional codes for subsequent points in the contour. Short contours are eliminated.

Mean Coding The mean RGB value of pixels between contours is coded.

Distance Coding The mean distance value of contour pixels is coded.

Color Extraction RGB values for pixels between contours is determined and a line approximation is used to represent the information.

Color Coding The endpoints of the line approximating the RGB values for pixels between contours is coded.

Binary Block Matching The edge maps from two consecutive images are used to determine the motion vectors of each n x n block.

Error Coding The reconstructed image from the motion vectors is compared with the second image to produce an error which is encoded.

The algorithm combines contour, color, and distance attributes to produce an image coding system which can be used for recognition.

The transportation industry has always been concerned with the efficient and safe movement of traffic. The data gathered from traffic monitoring is used to make real-time traffic control decisions which affect the movement and safety of traffic. This research is geared towards highway monitoring to demonstrate the feasibility and benefits of recognition of 3D compressed images. The 3D data will provide the necessary information to eliminate the problems of shadows and overlapping vehicles. Three applications (vehicle counting, vehicle speed detection and vehicle matching) will demonstrate the feasibility of the system. All three applications require the detection of vehicles in the image. Vehicle counting and vehicle speed detection require the tracking of vehicles in subsequent scenes. Vehicle speed detection utilizes motion estimates to determine the vehicle speeds. The following sections discuss the significant components for recognition: detection of vehicles in a scene, tracking of vehicles in subsequent scenes, and motion estimations of vehicles.

Impact: Performing recognition of 3D compressed images can reduce the processing time by limiting the number of images searched and/or by limiting the search region within an image based on contour attributes. This method can be used in many different applications where the frequent use of recognition of images is required.

Future Work: Develop applications which demonstrate the benefits of combining image compression and recognition.

Research Support: US Department of Transportation

References:

[1] J. Bergendahl. A Computationally Efficient Stereo Vision Algorithm for Adaptive Cruise Control, *M.S. E.E. Thesis*, MIT, Cambridge, May 1997.

[2] I. Masaki. Industrial Vision Systems Based on Application-Specific IC Chips, *IEICE Transactions*, Vol. E 74, No. 6, June 1991.

[3] M. M. Mizuki. Edge Based Video Image Compression for Low Bit Rate Applications. *M.S. E.E. Thesis*, MIT, Cambridge, September 1996.