Chapter 21

3-D Video based Disparity Estimation and Object Segmentation

Tao Gao
Electronic Information Products Supervision and Inspection Institute of Hebei Province, China

ABSTRACT
Stereo video object segmentation is a critical technology of the new generation of video coding, video retrieval and other emerging interactive multimedia fields. Determinations of distinctive depth of a frame features have become more popular in everyday life for automation industries like machine vision and computer vision technologies. This paper examines the evaluation of depth cues through dense of two frame stereo correspondence method. Experimental results show that the method can segment the stationary and moving objects with better accuracy and robustness. The contributions have higher accuracy in matching and reducing time of convergence.

1. INTRODUCTION
As human brain can deal with the weak difference between the left and right eye images, we can sense the external 3-D world; and this ability is called stereoscopic vision. Stereo image sequence is a three-dimensional visual image form, it uses left and right groups of images to describe a group of scenes, and the human eyes apperceive the 3-D depth information by addressing the relative position between the two images. The 3-D reconstruction of video scenes by stereo analysis is an important advancement in machine vision and computer vision. Most of the automation work requires modeling of 3-D structure of environment. Compared with traditional two-dimensional images, stereo images are more “realism” and the description of the scene is more natural. Currently, 3-D vision system has been widely applied to stereo video communications (Raymond & Clifton, 2000), robot vision (Nishihara & Poggio, 1984), aviation navigation (Stefano, Marchioni, Mattoccia, & Neri, 2004), and other fields. With ever-growing material and...
spiritual needs, the stereo images will gradually replace the traditional single vision images, and will be more used in television, online shopping, remote medical diagnosis and other civilian areas.

To store or transmit the stereo images, a more efficient compression encoding program must be developed. Stereo image sequence compression method was put forward in the late 1980s; after more than 25 years of development, people have developed several comparatively mature algorithms. However, to the practical application point, there is still not a unified coding standard.

The video object based stereo video coding method is to separate the video object from the scenes and to extract its borders, texture, movement and other parameters, then to code these parameters to achieve the purpose of coding the whole image (Aizawa & Huang, 1995). This method uses the hidden 3-D depth information, through the creation of 3-D objects and coding model to improve the coding efficiency and to reduce the influence of the block. It provides a more natural scene interpretation. However, this approach requires sophisticated image analysis process, such as: object segmentation, object modeling, and all these are not ripe at present so it can only be applied to a single background image with simple motion; its widely using depends on better solving some of these key technologies.

This paper presents a redundant wavelet transform based stereo video object segmentation algorithm. First, we use the redundant wavelet transform to extract the feature points of stereo video images, then according to the feature points we do the disparity estimation, to form a disparity map. The stationary objects are segmented from the stereo images by the disparity map. For the moving objects, we use a redundant wavelet transform based moving object extraction algorithm to segment the moving target from the redundant wavelet domain. Experimental results show that our algorithm can segment video objects from stereo video images, including stationary objects and moving objects with good results, highlighted details, and simple calculation process; all these can help to the subsequent coding operation.

2. STEREO VISION GEOMETRIC THEORIES

Three-dimensional camera system generally can be classified as three-dimensional parallel camera system and three-dimensional clustering camera system. If the two cameras are installed with parallel optical axis, it is called three-dimensional parallel camera system; if the two optical axes intersect at the objects, it is called three-dimensional clustering camera system. The best way is by solving the correspondence algorithm from left image to right image as well as right to left image also to be considered. The input is a set of video frames, typically the left image, is used as reference, to select the best depth estimate for each pixel. For each depth estimate of the reference depth map, the number of other depths maps that it occludes or passes is computed. This process is repeated for each depth estimate of all other depth maps with respect to reference image plane. The most stable solution is defined as the minimum depth from the identified discontinuity through dense two frame stereo matching method. This paper considers the 3D parallel camera system. Optical axis of two cameras parallel to each other and are perpendicular with the baseline. The space is determined by the plane known as the heart polarization plane. With a space (X, Y, Z), X is the horizontal direction and Y is the vertical direction, Z is the depth direction. The camera lens in the image plane is at point (x, y), where x parallel to the X, y parallel to the Y. In parallel camera system, the point in space around the image plane of the projector is at the same general location of the y coordinate, and the two set up their corresponding image brightness of the same point, that disparity search may be mainly concentrated in the horizontal direction, thus speeding up the searching and matching process.

Let \( f \) be the focal length, \( s \) is the image plane size, \( x_{11}, x_{12}, x_{21}, x_{22} \) are the positions of points \( p_1 \) and \( p_2 \) in planar image around the projection image plane to the right position between the left and right images. The disparity definition for the