論 文 の 内 容 の 要 旨

論文題目　Three -Dimensional Image Mosaicking Using Multiple Projection Planes
（複数投影面を利用した 3 次元画像モザイキング）

氏　名　　全 在 春

Environment visualization in a virtual space using synthesizing photo-realistic views is a common representation method on Internet websites, mobile systems, or so on. So far, most of these websites have used still images for environment visualization, and they can be monotonous, due to the fixed angle and viewpoint used. Recently, panoramas have appeared that give a more impressive visualization. Although the panorama surrounding a viewpoint enables the user to pan and zoom inside the environment, the viewpoint of the panorama remains fixed. Considering the popularity of still/video cameras, a more impressive form of environment visualization, which allows the user to view the scenes from arbitrary viewpoints and angles, is required. Three-dimensional Geographic Information System (3D GIS) data meet the requirement well.

There are two main approaches for creating the 3D GIS data: (i) detailed 3D reconstruction of a scene, and (ii) mosaicking of sequential images. For 3D reconstruction, a laser and a Charge Coupled Device (CCD) combined with a Global Positioning System (GPS) and Inertial Measurement Unite (IMU) capability have been widely used. This type of system can acquire detailed 3D surfaces of various objects along with any texture. In order to create the 3D reconstruction from 3D point data sets taken from laser system, the process is composed of alignment of 3D point-clouds data sets, 3D surface reconstruction from the aligned point clouds data, and texture mapping. Sequential stereo images have also been used for 3D reconstruction. Although cost of the stereo camera system is cheaper than the laser system, stereo matching process is additionally required. In order to create the 3D reconstruction from sequential stereo images taken from stereo camera system, the process is composed of sparsely well-distributed optical flow detection, self-calibration, each pixel matching, alignment of

3D point-clouds data sets, 3D surface reconstruction, and texture mapping. Unfortunately, it is difficult to apply these detailed 3D surfaces to current Internet or mobile systems, because of their limited real-time transmission speeds.

The image mosaicking technique is considered to be another efficient approach to reduce the cost and the difficulty of creating suitable environment visualization on websites. Generally speaking, there are two types of image mosaicking: (i) with prepared 3D data such as Digital Elevation Model (DEM) and (ii) without prepared 3D data. Because we don't have prepared 3D data in the case of environment visualization, this paper focuses on the second type.

The image mosaicking techniques in the second type fall mainly into two categories: (i) manifold (panoramic and spherical) mosaicking from a rotating camera and (ii) single mosaicking from a moving camera. The viewpoint of manifold mosaicking is fixed as stated above. Several single mosaickings have been developed that can create image mosaics using sequential images taken from a linear moving video camera, but when images are captured using a tilted camera, the result is curled. Zomet et al. (2000) dealt with this problem by warping trapezoids into rectangles, while maintaining other image feature invariants. Meanwhile, Zhu et al. (2004) proposed the creation of parallel-perspective stereo mosaics using an airborne video camera. First, an algorithm calculated the relative position between two consecutive frames for all the pairs in a sequential image. Center strips were then extracted from each frame and placed in their relative positions to create an image mosaic. The two algorithms based on the one projection planes are not effective to an image sequence taken from a side-looking camera along a road in an urban. If an image sequence is taken while a camera is turning at the intersection of two streets, the mosaicked image will represent roadside objects far away from reality. Zomet *et al*. (2004) developed crossed-slits projection to solve such problem. Rom *et al*. (2004) proposed several user-specified slits as an application of the crossed-slits projection technique. Even though the crossed-slits projection can create an image mosaic from the image sequence of images, the image motion of each frame is limited to less than a single pixel to create an image mosaic keeping the original image resolution. In addition, it is difficult to calculate camera orientation accurately to create well-aligned crossed-slits images. Since the distance between urban objects and a moving camera is very close so that the image motion is generally over 10 pixels, the algorithm of the crossed-slits projection is not effective. More importantly, the image mosaics based on the crossed-slits projection don't provide the 3D feeling since road-side objects look standing on a straight street.

Solving this problem is critical for the algorithms. Far-range areas, such as the intersection are also common in city street areas. Certain objects appear repeatedly around such areas, and this phenomenon is called the "ghost effect". This problem has to be dealt with in environment visualization. In addition, seam-line detection is important in mosaicking two neighbored image frames.

The goal of this thesis is to give 3D feeling for environment visualization on Internet websites or mobile systems. The created 3D image mosaics are textured to 3D vector data generated from a side-looking camera along a road in a city or town area. The 3D feeling can be obtained through showing 3D vectors and textured image slits. 3D vectors and textured image slits give imagination of the global abstract and detail part of objects, respectively. The proposed method is combination of sparsely well-distributed optical flow detection for video frames, camera orientation approximation for a moving video camera, multiple projection planes with a 3D surface geometry, an expanded crossed-slits projection around the far-range areas to suppress the "ghost effect", and visually optimum seam-line detection around boundaries among image-slits in an image mosaic for creating seamless image mosaics.

To reliably and robustly detect sparsely well-distributed optical flows for video frames, contour matching algorithm using epipolar geometry is used in this research. In order to save the computational cost of contour matching, a hierarchical strategy based on an image pyramid is adopted. Sparsely well-distributed optical flows are detected through evaluation of optical flows in an image.

The optical flow detection is followed by calculation of the exterior parameters of a moving video camera using the optical flows. The exterior parameters of the second frame are first approximated by using coplanarity condition. The exterior parameters from the third frame to the last frame are then approximated by using triplet based on bundle adjustment. The 3D coordinate of the detected optical flows is calculated by using collinearity condition with the approximated exterior parameters at the previous and current frames.

The proposed method uses a roadside scene acquired by a side-looking video camera as a continuous set of textured vertical planar surfaces named "multiple projection planes". The scene geometry is approximated to multiple vertical planes using sparsely distributed optical flows. These vertical planes are concatenated to create an approximate model on which the images could be back-projected as textures and then blended together.

If the multiple projection planes are created around the far-range area in the same way around the close-range area, then the ghost effect will occur. To suppress ghost effect, the far-range areas are detected by using the distance between the image frame and the 3D coordinate of the detected optical flows. The crossed-slits projection is expanded to deal with the ghost effect. The ghost effects are suppressed by projecting the part of image frames onto 3D multiple planes utilizing vectors passing the focal point of frames and a virtual focal slit. The virtual focal slit is calculated by utilizing the first and last frames of the far-range areas.

It is important to point out here that achieving a well-aligned image mosaic from overlapping images is a challenging process. Since visually pleasing borderlines are usually when the pixel differences created from overlapping images are as small as possible, the dynamic programming algorithm can be utilized to find the optimal path. An obstacle to this is that the determined minimum cost path is likely to be the shorter path, while the human visual system is more sensitive to the higher pixel difference, regardless of the length of the seam-line. To overcome this obstacle, this research proposes an algorithm that avoids the path with the large pixel difference by using an adaptive cost-conversion method, and the adequacy of this algorithm with a cost-converting function is explained analytically. This research also suggests a figure of merit, which is the summation of the fixed number of the biggest pixel differences, as an evaluation of seam-lines.

The proposed method is applied to sequences of real images taken from a circling train equipped with a side-looking video camera. The effectiveness of image mosaics in 3D space created by using the proposed method is demonstrated through the software developement in this research.

Since the textured projection plane of each frame consists of four 3D coordinates and a part of the image, the results obtained by using the proposed method can be the form of MPEG-4 data. One of the requirements of MPEG-4 composition for streaming of 3D worlds will be the Virtual Reality Modeling Language (VRML) that has made viewing 3D content on Internet websites possible. Therefore, these results as next generation navigation data can be widely applied to 3D virtual visualization and games on websites, cellular phones, and PDAs.