On Environment Modeling for Visual Navigation

Zhigang Zhu

Ph.D. Thesis
at
Department of Computer Science and Technology
Tsinghua University, Beijing
(1997)

Table of Contents

Appendices

Bibliography

Extended Summary

Panoramic/omnidirectional representations of image sequences have a wide application scope, including robot navigation, virtual reality, interactive 2D/3D video, content-based video compression, and full-view video surveillance. Scene modeling using image mosaicing and panoramic/omnidirectional vision has attracted great attentions in the fields of computer vision and computer graphics in recent years. Usually, researchers either focus on the analysis/recognition part (e.g., panoramic /omnidirectional vision for robot navigation), or on the synthesis/visualization part (e.g. image mosaicing, panoramic and layered representation in image-based rendering or virtualized reality). This thesis makes a first attempt to systematically bring the two seemingly quite different topics under a single umbrella of "*visual modeling and presentation*".

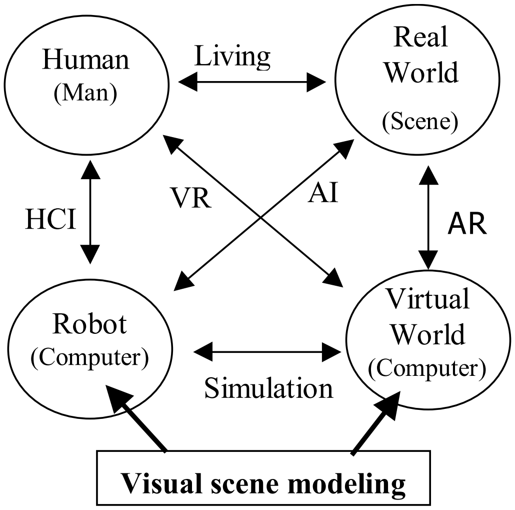


Fig. 1. Interactions diagrams. (HCI: Human-Computer Interaction; VR: Virtual/Virtualized Reality; AI: Artificial Intelligence / Visual Navigation; AR: Augmented Reality)

First, let us have a look at the two topics – robot navigation and virtualized reality -- in a broader perspective of interaction between "being" and "environment" (Fig. 1). We can find a very close resemblance between them: robot navigation is the interaction between a robot (i.e. a digital being) and the real 3D world, while virtual/virtualized reality is the interaction between a person (i.e. a human being) and a virtual/virtualized environment. If we limit our discussion of the "interaction" to visual perception, the central problem that needs to be solved for these two kinds of interactions is visual scene modeling and representation in a computer – either inside the “mind” of a robot or outside the mind of a human being (Fig. 1).

Second, a closer examination of the research efforts of the past ten years shows that techniques and representations for the two applications are surprisingly similar. Graphics people talk about multi-perspective projection for image-based rendering of large-scale scenes, while vision people try to use the concept of spatio-temporal panoramic view images in robot localization and landmark recognition. Vision/robotics people take advantage of the 360-degree view angle of omnidirectional images for map building, road following, obstacle detection in robot navigation, whereas graphics people try to generate omnidirectional image representations for image-based rendering. My own research also shows that we can use the exact same basic methodology for building and the same structures for representing visual scene models for both robot navigation and image-based rendering.

Finally, we can find a class of interesting applications for integrating these two kinds of models: a *human-robot intelligent navigation* (***HRIN***) system, such as a semi-autonomous mobile robot for mail delivery, military surveillance and intelligent transportation. In a HRIN system, the robot will automatically carry out most of the basic tasks such as road following, obstacle detection, and target localization, while a human supervisor will make important decisions or deal with some emergent situations via augmented reality and tele-operation. Thus a unified model that includes both the symbolic environment model for navigation and the photorealistic scene model for visualization is required.

Needless to say, visual navigation of a mobile robot in a natural environment has always been a very interesting but challenging problem. It involves almost every aspect of computer vision research - from visual sensors through robust algorithms to visual representations. The basic requirements of visual navigation include global localization (to decide where to go), road following (to stay on the road) and obstacle detection (to avoid collision). Only after these safety requirements have been satisfied, which have been proven to be not a trivial problem, can the robot pursue other task-oriented goals. It is clear that visual environment modeling is the foundation of these basic issues in visual navigation - and it may extend to most of the real world problems in computer vision. This work presents a systematic approach to visual modeling of a natural scene for robot navigation:

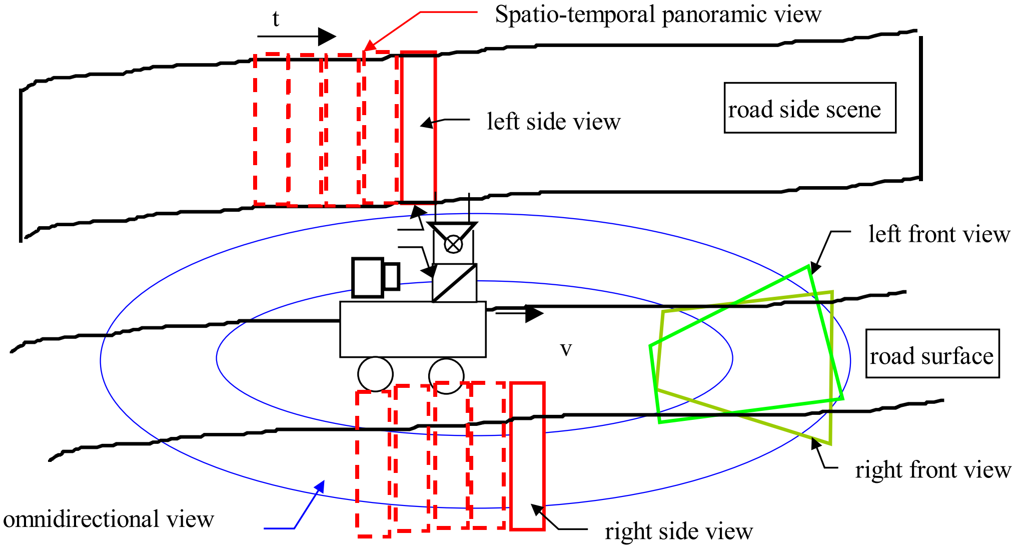
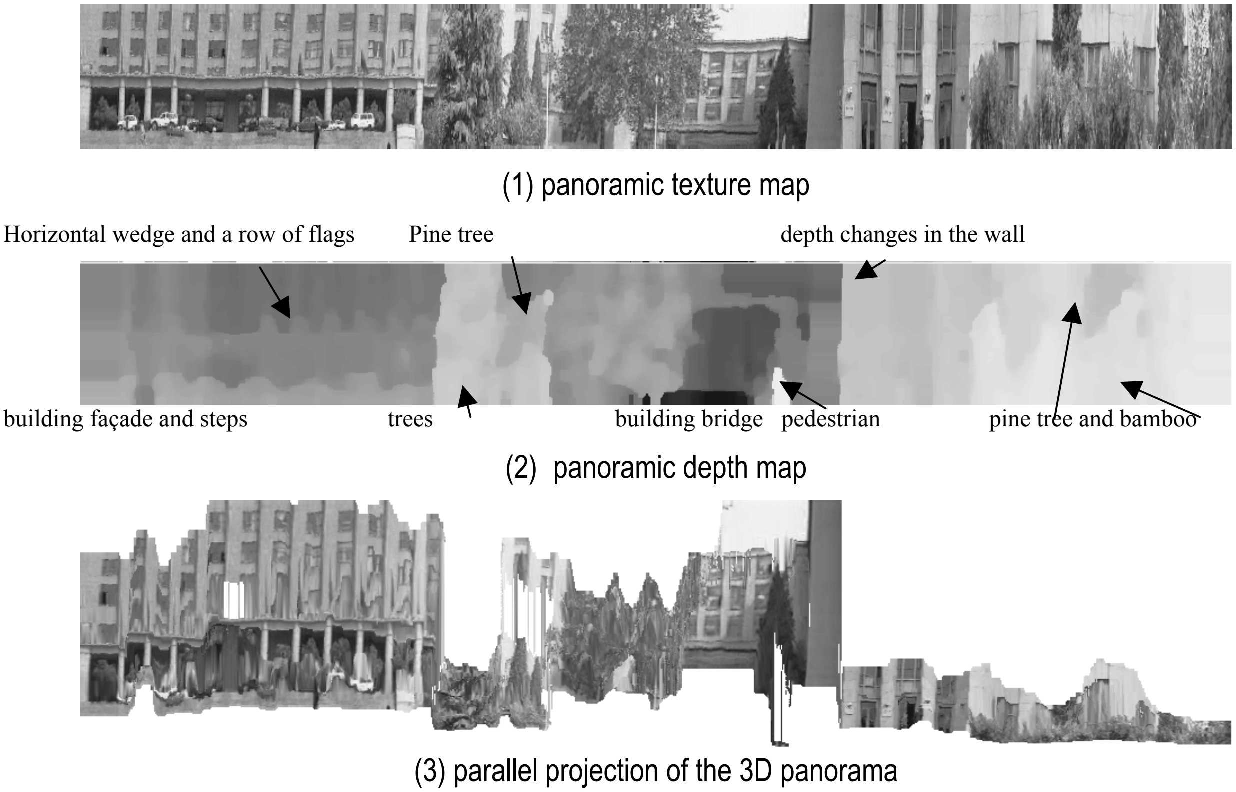


Fig. 2. Full view vision for robot navigation

1. A purposive, multi-scale and full-view visual scene modeling approach is proposed for visual navigation in a natural environment (Chapter 1 - Chapter 5). As a typical instance, an integrated system **POST** is proposed which combines three novel modules (Fig. 2): ***P****anoramic vision for landmark recognition,* ***O****mnidirectional vision for road understanding and* ***ST****ereo vision for obstacle detection*. This approach tries to overcome the drawbacks of traditional visual navigation methods that have mostly depended on local and/or single view visual information. However, the proposed approach is not just a simple combination of the three novel sensors and methods, but rather a systematic integration under the strategy of purposive vision ("the right way for the right work"), and under the philosophy of a systems approach which emphasizes that "the whole is more than sum of its components". Thus, correct sensor design, adequate levels of scene representation and corresponding robust and fast algorithms are specifically explored for each given task while the interconnection among the vision sub-systems are taken into consideration under the overall goal of autonomous navigation. The human-robot cooperation in different navigation modes (autonomous, semi-autonomous and tele-operational) and different levels of vision enhancements (video enhancement, stereo enhancement, view enhancement, information enhancement and virtualized reality) will be discussed.


(1) panoramic texture map

Horizontal wedge and a row of flags
Pine tree
depth changes in the wall
building façade and steps
trees
building bridge
pedestrian
pine tree and bamboo

(2) panoramic depth map

(3) parallel projection of the 3D panorama

Fig. 3. 3D Panoramic representation for landmark selection

2. A two-stage method is presented for *3D panoramic scene modeling for landmark selection* (Chapter 2). As inputs, image sequences are captured by a video camera subject to small but unpredictable fluctuation on a common road surface. First, a 3D image stabilization method is proposed which eliminates fluctuation from the vehicle's smooth motion so that "seamless" panoramic view images (PVIs) and epipolar plane images (EPIs) can be generated. Second, an efficient *panoramic EPI analysis method* is proposed to combine the advantages of both PVIs and EPIs efficiently in two important steps: frequency domain locus orientation detection, and spatio-temporal domain motion boundary localization. The two-stage method not only combines Zheng-Tsuji's PVI method with Bolle-Baker's EPI analysis, resulting in the so-called panoramic EPI method, but also generalizes them to handle image sequences subject to small but unpredictable camera fluctuations. Since camera calibration, image segmentation, feature extraction and matching are completely avoided, all the proposed algorithms are fully automatic and rather general. Finally, a compact representation in the form of a 3D panorama for a large-scale scene is constructed that can be used effectively for generalized landmark selection for robot navigation (Fig. 3). This method will

further be applied in image-based rendering.

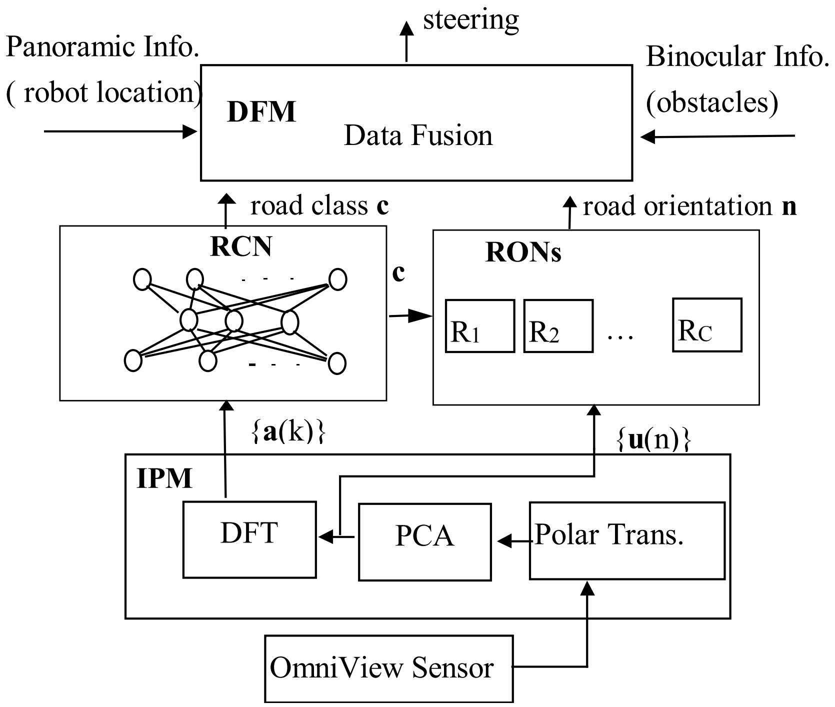


Fig. 4. . ROVINN architecture and the interconnection with other two modules (RCN: Road Classification Network; RON: Road Orientation Network; DFM: Data Fusion Module; IPM: Image Processing Module; DFT: Discrete Fourier Transform; PCA: Principal Component Analysis)

3. A new road following approach, the ***R**oad **O**mni-**V**iew **I**mage **N**eural **N**etworks* (**ROVINN**), has been proposed (Chapter 3). It combines the omnidirectional image sensing technique with neural networks in such a manner that the robot is able to learn recognition and steering knowledge from the omnidirectional road images that in turn guarantee that the robot will never miss the road. The ROVINN approach brings Yagi's COPIS (conic omnidirectional projection image sensor) method to outdoor road scenes and provides an alternative solution different from CMU's ALVINN system. Compact and rotation-invariant image features are extracted by integrating an omnidirectional eigenspace representation with frequency analysis, using principal component analysis (PCA) and Fourier transforms (DFTs). The modular neural networks of the ROVINN estimate road orientations more robustly and efficiently by classifying the roads as a first step, which enables the robot to adapt to various road types automatically.

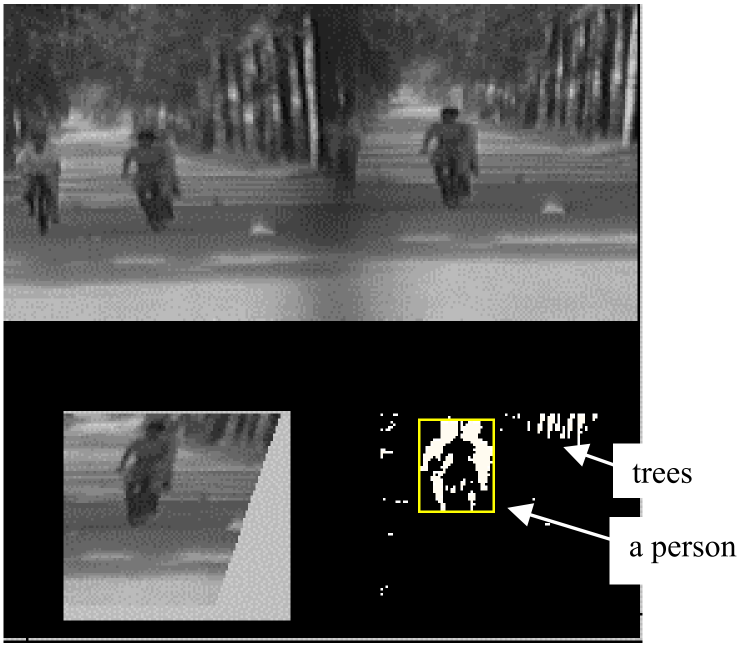


Fig. 5. Image gaze transformation and obstacle detection. Top: Left and right view in a single camera image; Bottom-left: rectified left image by gaze transformation; Bottom-right: obstacle region after zero-disparity gaze control. The difference image shows that the ground images have been registered.

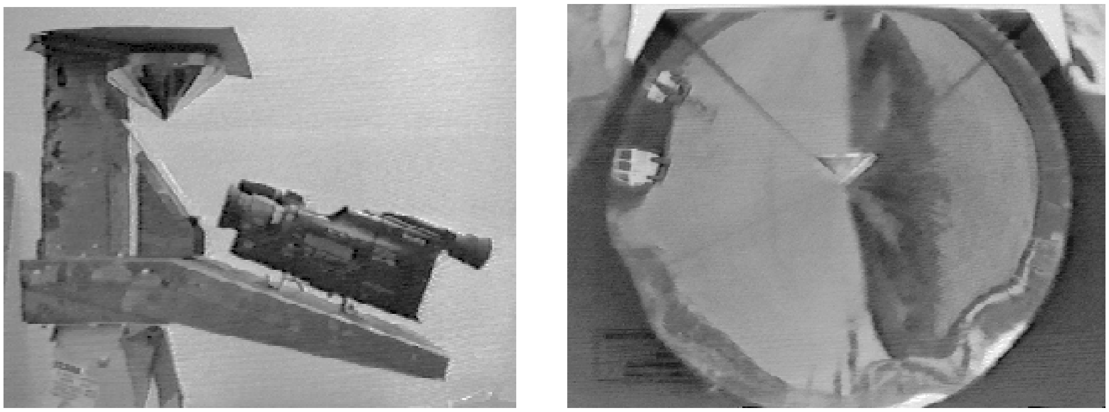
4. A novel method called the *Image Gaze Transformation* is presented for stereo-vision-based road obstacle detection (Chapter 4). Obstacle detection is modeled as a reflexive behavior of detecting anything that is different from a planar road surface. Dynamic gaze transformation algorithms are developed so that the algorithms can work on a rough road surface. The novelty of the (dynamic) gaze transformation method, which resembles gaze control of the human vision, lies in the fact that it brings the road surface to zero disparity so that the feature extraction and matching procedures of traditional stereo vision methods are completely avoided in the proposed obstacle detection algorithms. The progressive processing strategy from yes/no verification, through focus of attention, to 3D measurement based on the reprojection transformation make the hierarchical obstacle detection techniques efficient, fast and robust.

To validate the proposed strategies and methods, we have implemented the following algorithms and systems.

(1) **Design of novel sensors**. An omni-view image sensor is designed and realized (Fig. 6), and its properties for outdoor road understanding are thoroughly studied. A patented single camera binocular vision system with full horizontal field of view is also designed and constructed where left and right views are projected to the up- and bottom halves of a single image (Fig. 7). It has been put into real road application for obstacle detection. An inexpensive and integrated full view smart sensor POST (Panoramic, Omnidirectional and STereo vision sensor) is proposed, which integrates a 360-degree omnidirectional view with a binocular forward view as well as both left and right side views by using a single camera and a set of reflection mirrors (Fig. 8).

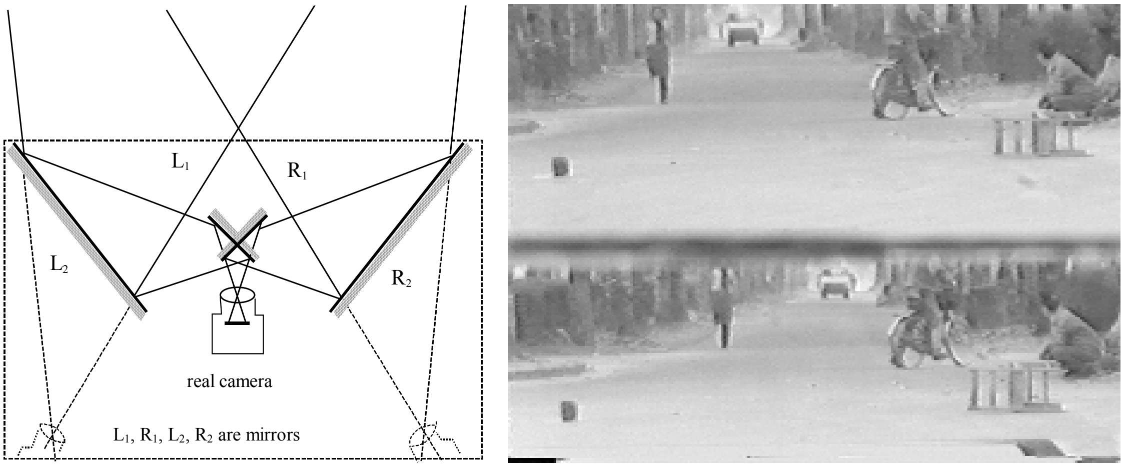
(2) **Real scene experiments.** Experimental results of training and testing the ROVINN using real road images have shown that the proposed method for road following is quite promising. A real-time visual obstacle detection system has been set up and extensively tested on outdoor road scenes.

(3) 3D **Scene modeling system**. In the 3D panoramic scene modeling system (Fig. 9), the algorithms for motion filtering and image stabilization, kinetic occlusion detection and depth layering have been developed, and 3D layered panoramic models have been constructed for many image sequences. These efforts form the basic framework for both global localization using generalized landmark selection, and the synthesis of photo-realistic image-based renderings.



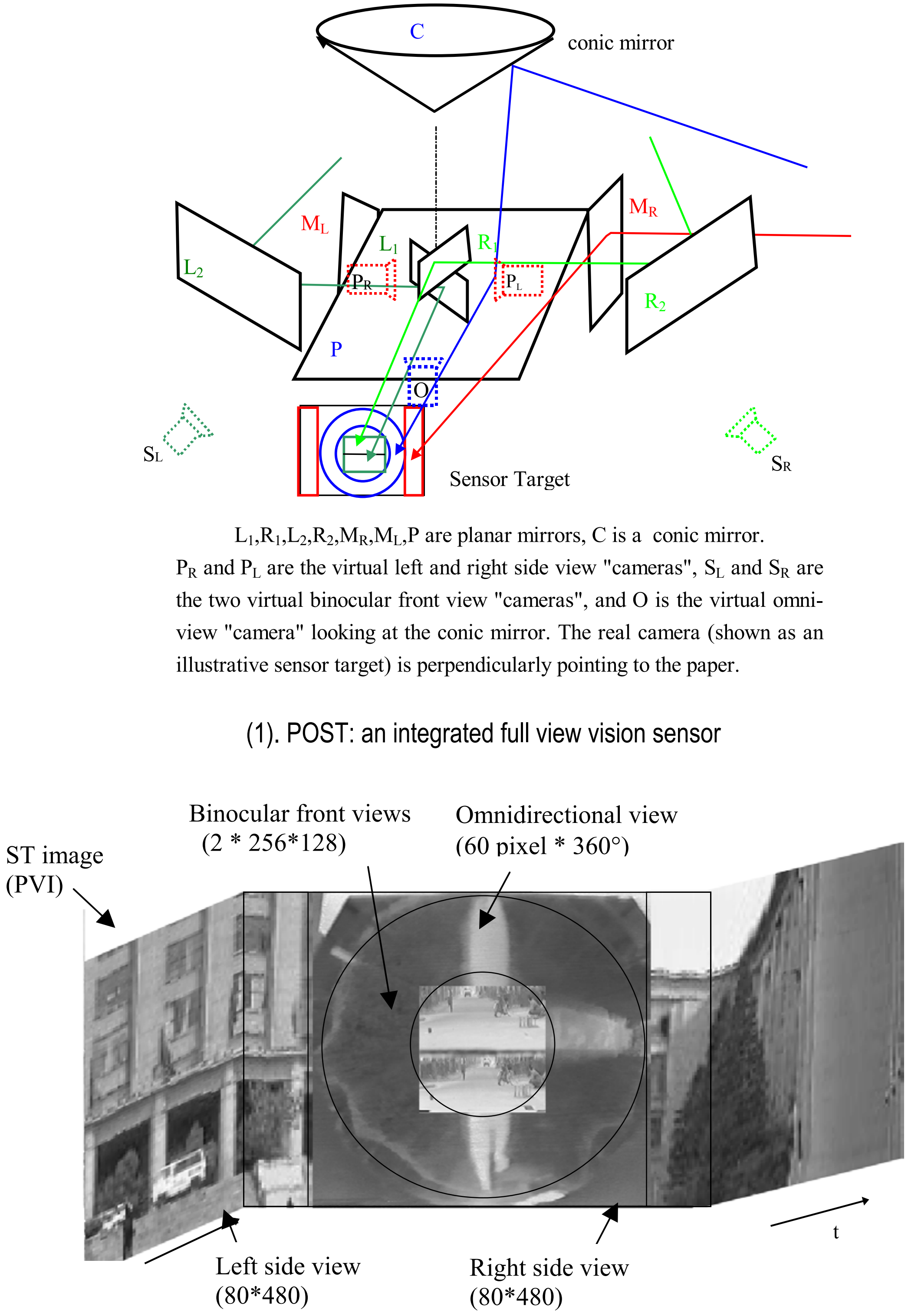
(1). A prototype of the OVI sensor (2). An omnidirectional image

Fig 6. Omni-view image (OVI) sensor



(1) System geometry (top view) (2) A real binocular image pair

Fig. 7. Single camera stereo vision system: left and right views are projected to the top and bottom halves of a single image



L_1,R_1,L_2,R_2,M_R,M_L,P are planar mirrors, C is a conic mirror.
P_R and P_L are the virtual left and right side view "cameras", S_L and S_R are the two virtual binocular front view "cameras", and O is the virtual omni-view "camera" looking at the conic mirror. The real camera (shown as an illustrative sensor target) is perpendicularly pointing to the paper.

(1). POST: an integrated full view vision sensor

(2). A composite image (640x480 image)

Fig 8. Integrated full view vision sensor: POST

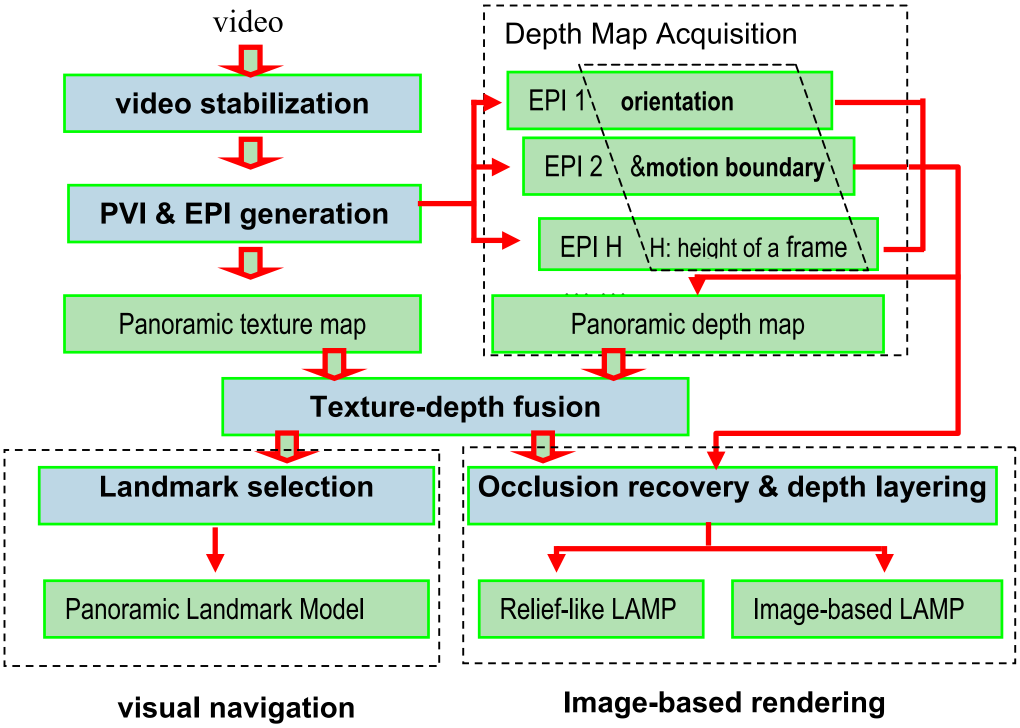


Fig. 9 System diagram of 3D panoramic scene modeling (PVI: Panoramic View image; EPI: Epipolar Plane Image; LAMP: Layered, Adaptive-resolution and Multi-perspective Panorama)