Robotic navigation through visual sensors

Dr. Ji Hui, Department of Mathematics

Introduction

Autonomous navigation of a mobile robot guided by visual sensors in a natural unstructured environment has always been an interesting but challenging problem. A machine vision system that successfully guides the navigation of the robot not only must supply the robot with "eyes", but also must interpret what the eyes "see". Thus a number of local and global capabilities must be designed for such autonomous systems operating in real environment. Local capabilities enable the system to interact with its immediate surroundings, liking understanding its own motion, recognizing obstacles and avoiding other moving objects. Global capabilities enable the system to deal with larger spaces, like building the memory of an area and finding particular places. My main research focuses on the interpretation of visual information within video data (local capability) and spatiotemporal representations of the outside world for visual navigation (visual homing). Some basic questions I am interested in are: Where is the camera? Where are the moving objects? What is the 3D view of the surrounding environment? How to identify the same object under different view-points?

Camera and rigid motion

Autonomous navigation requires a mapping of the environment which tells where you and other objects are. By observing the scene with a video camera, we could estimate the motion of the camera and the moving objects, as well as the 3D structure of the surrounding static environment. The current popular approach to accomplish such a task in a complex environment is much like in Engineering, a module one, which breaks the problem into multiple modules in a feed-forward fashion. However these modules generally depend on each other in a chicken-and-egg dilemma and the traditional alternative iteration methodology is fragile and impractical in practice. Inspired by biological vision system where many lateral loops exist among modules with feed-back mechanism, we propose a holistic framework for estimating camera motion and understanding the 3D structure of the environment from video data. Our holistic approach differs from the traditional alternative approach in the fact that the modules in our approach are dynamic and their complexities are adapted to the availability of reliable information. A preliminary framework operating on video data have been developed which links: camera motion estimation, 3D scene reconstruction and moving objects detection. In particular, a novel camera motion estimation technique is developed to provide accurate camera motion by utilizing all available image frames within the video simultaneously. This approach is based on a new multi-frame subspace constraint on planar surfaces. The availability of camera motion estimation with high accuracy gives a solid foundation for building the

Dr. Ji Hui

Dr. Ji Hui obtained his Bachelor's and Master's degree in Mathematics from Nanjing University in China and NUS in Singapore respectively. He then received his Ph.D. degree in Computer Science from University of Maryland at College Park. Since then he joined Department of Mathematics at NUS in July 2006 as an assistant professor.

Research interests

- Computer vision and visual navigation
- Human vision and visual psychophysics
- Applied harmonic analysis and wavelets.

Contact:
Tel: (65) 65106-8845
Office: S14-02-26
Email: matjh@nus.edu.sg

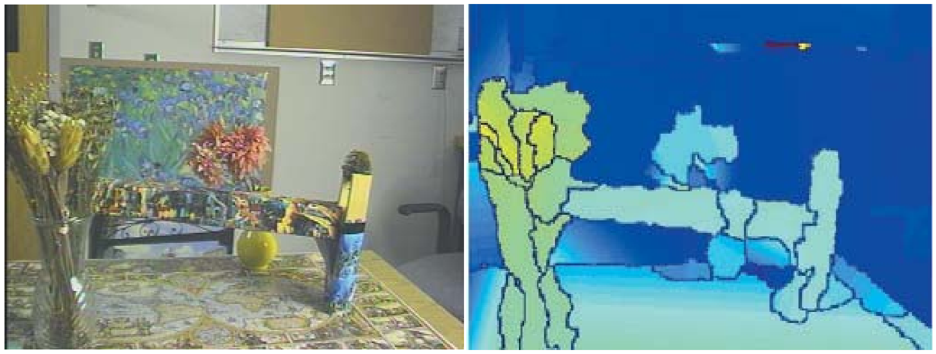

Fig. 1. One key image frame in the video and its corresponding 3D depth map

interactive loops among other modules in our integrated system. The modules benefits most from this interaction include the module of refining 3D structure reconstruction and the module of object segmentation in the scene. Fig. 1. demonstrates the structure segmentation and 3D reconstruction from a video sequence. Many other applications could also benefit from this framework. Fig. 2. illustrates its

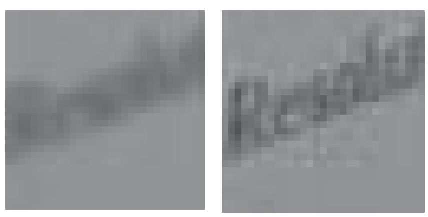

Fig. 2. Some region in a low-resolution image and its high-resolution enhancement from a video sequence

application on super-resolution imaging by providing high quality image frame alignment.

Texture analysis and recognition

Visual homing refers to the capability of the robot to return to important places in their environment by store an abstract image of the surroundings while at the goal, computing a home direction from a matching between this "snapshot" and the currently perceived abstract image. Texture description is one key source for constructing such an abstract image; the other is geometrical description. The ideal texture descriptor for visual homing should be invariant to environmental changes: changes in view-point, illumination and underlying surface geometry. Current available texture descriptors are either based on statistical quantity which are sensitive to the illumination changes, or based on local image features which only keep local invariance to view changes. We develop an efficient framework which is able to combines global spatial variance and local robust measurement based on multi-fractal spectrum theory. The multi-fractal spectrum is very powerful on capturing the irregularity of the image measurements distribution. By fusing multiple multi-fractal spectrum fro the distributions of multiple image feature measurements, we are able to construct a powerful texture descriptor with rich discriminative information and with robust invariance to environmental changes. This texture descriptor is invariant under the bi-Lipschitz map both in spatial domain and in illumination domain. Such strong invariance includes the invariance to view-point changes and to non-rigid deformation of the texture surface, as

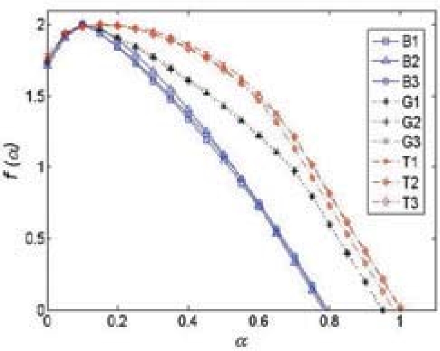

Fig. 3 The MFS texture signature from image intensity for three outdoor textures with different views and different illumination conditions. The MFS curves with the same color belong to the same texture.

well as to illumination changes. Fig. 3. shows the invariance of our texture descriptor to various environment and good discrimination for different textures. Our texture descriptor could also see its application in texture classification and retrieval.

Summary and Conclusions

Visual navigation has been and will continue to be a rich source for basic research questions to machine vision. The research into visual motion and navigational competence both in real or virtual environment could see its wide range of applications. The potential for applications in autonomous systems operating in real environment has been illustrated in DARPA Grand Challenge 2005 where four unmanned ground vehicles successfully complete 132 miles race in the desert within ten hours. With regards to virtual environment, vision-based human-machine interaction could bring the naturalness between machine and human to a new level. This can happen only if an adequate model of the scene from cameras could be automatically extracted.

References

1. Y. Aloimonos (Ed.), Visual navigation: from biological systems to unmanned ground vehicles, Lawrance Erlbaum Associates, 1997.

2. H. Ji and C. Fermuller, A 3D shape constraint on video, PAMI, 2006.

3. H. Ji and C. Fermuller, Wavelet-based super-resolution reconstruction: Theory and Algorithm, ECCV, 2006.

4. Y. Xu, H. Ji and C. Fermuller, A projective invariant for textures, CVPR, 2006.

5. H. Ji and C. Fermuller, Integration of motion fields through shape, CVPR, 2005.