Research Statement

Ashok Veeraraghavan
Email: vashok@umiacs.umd.edu
Homepage: http://www.cfar.umd.edu/users/vashok/

1 Introduction

I am interested in envisioning, designing and developing computer vision systems that are better able to understand, interpret and interact with us and the world around us. In this regard, my research has focused on 3 distinct subareas. Firstly, I have designed and built **computational Fourier cameras** that are able to overcome some of the limitations (depth of field, focus, depth extraction) of traditional cameras. Secondly, I have developed novel representations for appropriate visual data specifically in surveillance scenarios. This has resulted in algorithms for efficient **multi-video compression** and **target tracking for single and multiple cameras**. Finally, I have also developed appropriate inference algorithms for several vision tasks such as **gait based person identification**, **simultaneous tracking and recognition**, **human activity analysis**, **activity based video mining** and **structure from motion**. This research thrust will have broad impact on several domains including **consumer photography**, **surveillance**, **visual biometrics** such as face and gait, **markerless motion capture** for **medical applications**, wide area visual **sensor networks** and **computational photography**.

2 Research Thrusts

My research focus will be devoted to these interconnected areas: design of novel computational cameras that can alleviate some of the limitations of normal perspective cameras and development of sound inference algorithms for processing the video data and extracting the information content present in the captured videos.

2.1 Computational and Fourier Cameras

The visual sensors must contain all the information that is required in order to reliably extract required environment information. Unfortunately, traditional cameras are projective devices that maps lines in the 3D world to points on the image. The many-to-one nature of this mapping makes inference algorithms for many vision tasks either too intractable or very unreliable in the presence of noise. One way to alleviate this problem is to design visual sensors that capture light-fields as opposed to images. The light-field is the radiance as a function of position and directions. For scenes composed of lambertian objects this leads to an overcomplete representation of the visual information contained in the scene. Therefore inference algorithms for traditionally hard problems in vision like depth extraction, stereo and separating texture and depth edges etc. become simple from captured light-fields.

Towards this end, my research has been involved in the design of novel computational cameras, that go above and beyond the limitations of standard digital cameras. In particular, I have designed a **heterodyne light-field camera** (**SIGGRAPH 2007, Eurographics 2008**) that overcomes the depth-of-field and viewpoint based limitations of normal cameras. This design used sound principles from signal processing and communication theory especially, the **Fourier domain analysis of light-fields** in order to **modulate** the incoming light-field so that this 4-D light field can be captured on a 2D sensor. This is the one of the first devices that appropriately modulates a 4-D signal (light-field) in such a manner that it is captured at **Nyquist rate** on a 2-D sensor. This work also led to a theoretical understanding of the effect of **non-refractive modulation of plenoptic fields (CVPR 2008)** and has led to several other designs both from us and from other researchers.

2.2 Representations for Visual Data

As the number of interconnected cameras continue increasing, it becomes essential to develop efficient representations of the video data in a form that is amenable for both storage and inference. I have developed compression algorithms that use projective geometry to efficiently **compress multi-view videos (ICASSP 2008)** using a representation that exploits the inter-view redundancy in the form of a **homography**. In several surveillance and traffic monitoring scenarios, the only information that we seek to transmit to the central node may be the actual moving targets, their positions and appearance while the actual video is not as important. My research has concentrated on solving many of these problems especially in the context of domain knowledge assisted systems. I have developed **single and multiple camera tracking algorithms (PAMI 2007, PREMI 2005, ICASSP 2005)** for a variety of subjects including bees in a hive, humans and faces. Specifically, I have developed a **simultaneous tracking and behavior analysis** approach (**PAMI 2007**) where the behaviors of the actors are explicitly modeled as a **hierarchical Markov model** and the behavior and position estimation is performed by filtering using **Monte Carlo** methods such as **particle filtering**.

2.3 Visual Inference: Activity Recognition and Shape Sequence Matching

In order to efficiently interact with its surroundings a machine vision system must be able to understand the geometric structure of its environment (Occlusions, 3D structure, obstacles, moving objects etc.), recognize objects and understand the actions and behaviors of the subjects. My research has resulted in algorithms for **gait based person identification** (**CVPR 2004, PAMI 2005, CVPR 2006**), simultaneous tracking and recognition (**PAMI 2007**), **human activity analysis**, activity based **video mining** (**CVPR 2007**), model based **face recognition (PReMI 2005)** and **structure from motion** estimation in the presence of visual and non-visual sensors (**ICCV 2007**).

I have worked on an in-depth analytical characterization of the **manifold** on which shape sequences lie and this study has led to several approaches for **modeling and comparing shape sequences (PAMI 2005, CVPR 2004)**. Methods for comparing shape sequences are popular and have applications in action recognition, gait recognition, medical imaging and other related fields. I have also extended the approach for shape sequence characterization in order to study the **'function space of time warping'** specifically in the context of activity recognition (**CVPR 2006**).

My research has also led to several significant contributions to activity recognition and mining from videos. I have studied the relative importance of shape and dynamical cues (**CVPR 2004**) using shape-dynamical models. I have developed viewpoint and execution rate **invariant dynamical models** for activity recognition and have proposed a method for activity based **unsupervised mining** of video sequences (**CVPR 2007**) based on an accurate **manifold characterization** of the observation sequence using **subspace angles** between the observability spaces. These algorithms have significant impact in applications such as surveillance, traffic monitoring, house-hold monitoring of the elderly etc.

3 FUTURE RESEARCH PLANS

3.1 Short Term Research Plans

Computational Cameras: Several exciting vision applications can be tackled by appropriate use of the heterodyne-light field camera. We have already developed inference algorithms for depth extraction, all-in-focus image, separating texture and depth edges and also reconstruction of a texture mapped 3D surface for lambertian scenes from a single exposure of the light-field camera. We have also been able to build an extremely light-weight hand-held heterodyne light field camera that has immense potential for consumer photography, by providing the photographer with post-exposure control of focus and aperture size.

Representations: The homography based multi-video coding is inherently amenable for distributed visual sensor networks. I will work on extending the compression scheme to be robust to complex real-world phenomena like object motion, illumination variations, shadows etc. so that the algorithm may be deployed in large scale visual sensor networks like traffic cameras and other distributed surveillance systems. I will integrate the distributed multi-camera tracking algorithm with the homography based multi-video coding method in order to explore the effect of using explicit motion information of the targets into the compression scheme. Such an integrated approach will make this scheme ideally suited for surveillance applications.

Activity Recognition and Shape Sequence Matching: I am currently working on **factorization of an activity sequence** into its inherent covariates, those that are characteristics of the actor and those that are characteristics of the activity. This is made possible because of a careful manifold characterization of the observation space so as to account for the actor and activity based covariates jointly. We project the observations into an appropriate **Stiefel manifold** and distance measures computed in this manifold allow for efficient **style-content** separation. Preliminary experiments are promising. Recently, I have also worked on developing a unifying theory for detection, tracking and recognition problems in **large scale distributed visual sensor networks (PIEEE 2008)**. I will work towards extending this principle to include both activity inference and person identification in large scale visual sensor networks by integrating some of the rich theory developed in distributed function estimation with the geometric constraints obtained while using multiple cameras.

3.2 Long Term Research Vision

Compressive Sensing is a new signal processing paradigm in which signals that are sparse in some basis can be accurately reconstructed from sparse measurements at a rate much lower than the Nyquist rate. Emmanuel Candes (CalTech), Terence Tao (UCLA) and David Donoho (Stanford) independently developed the basic principles of compressed sensing while Richard Baranuik (Rice) has built a single pixel camera which is based on this principle. Since images and light-field data are **inherently sparse** in the **wavelet** and **fourier** domains respectively, it is only natural that the basic principles of compressed sensing and **sparse representations** (pioneered by Bhaskar Rao at UCSD) can be exploited to develop exciting new computational cameras that can go well beyond the resolution, depth of field, blur, sampling, and dynamic range limitations of current imaging devices. My research will study these fundamental limitations and exploit new mathematical paradigms such as compressed sensing in order to design computational cameras that can ameliorate these limitations.

Human Activity Perception and Understanding is critical so that a machine vision system can efficiently interact with its surroundings. My reserach will explore the limits of human action understanding from videos and will specifically concentrate on the study of **invariants** so as to make action perception robust to changes in viewpoint and illumination. My research will also explore the fundamental limits of these invariances and whether appropriate light-field modulation based sensing devices will be able to overcome these limitations. In particular, both non-refractive modulation of plenoptic fields and the signal processing framework of sparse representations will enable us to break these barriers.

A unique characteristic of imaging is the need to be **invariant** to viewpoint (projective invariants) and illumination. I will address questions regarding the limitations imposed due to the need for invariants. An **Information theoretic** understanding of the source of these limitations will allow us to design better inference algorithms and design better sensors and representations that can ameliorate these limitations.

3.3 Research Philosophy

Research is the constant quest of a scientist to steadily improve our understanding of nature and its varying elements. With Engineering Research comes the added responsibility of utilizing our knowledge about the nature of the world to design new systems that have a positive impact on our quality of life. I believe that it is important to work on problems that have significant impact on real-world issues. Equally important is to develop methodologies and principles that are grounded in sound mathematics so that a) objective measures of utility may be obtained and b) inferences and designs made in one particular domain may be easily applied to several domains. Therefore, I will strive to maintain a research focus that provides solid mathematical foundations for design and development of applications that have immense real world impact.

3.4 Research and Collaborations

I also believe that inter-disciplinary research is a very important component and may provide exciting and satisfying challenges. In this regard, I have been fortunate enough to have worked with scientists from several disciplines. I have worked with Prof. Amit K. Roy-Chowdhury (University of California, Riverside, Department of Electrical Engineering) in projects concerning activity analysis and recognition. I had the opportunity to work on the automated behavior analysis of bees with Dr. Mandyam Srinivasan (The University of Queensland, Queensland Brain Institute) who is one of leading researchers in visual neuroscience. I have also worked with Graphics researchers, Dr. Ramesh Raskar (MERL) and Prof. Jack Tumblin (Northwestern University, Department of Computer Science) in designing Light-field cameras. Recently, I spent a couple of weeks visiting Prof. Anuj Srivastava (Florida State University, Department of Statistics). During this time I learnt about the fundamental principles in learning statistics on complex non-linear manifolds and have come away with several applications of these techniques. I continue to maintain active collaborations with all these researchers. I have come to realize that research collaborations serve as a significant catalyst for new and innovative ideas and will always continue to maintain significant research collaborations with these and other researchers and scientists.