REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-06-0345

1. REPORT DATE *(DD-MM-YYYY)*: 07/28/2006

2. REPORT TYPE: Final

DATES COVERED *(From - To)*: 1 January 2003- 31 December 2005

4. TITLE AND SUBTITLE
Dynamic Vision for Control

5a. CONTRACT NUMBER

5b. GRANT NUMBER
F49620-03-1-0095

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Stefano Soatto

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University Of California, Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90095

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research /NM
4015 Wilson Blvd
Mail Room 713
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The goal of this project was to develop analytical and computational tools to make vision a viable sensor for the purpose of control. In order for unmanned vehicles to interact with a complex, unknown, uncertain and dynamic environment, they must be able to sense its shape (geometry), its reflectance properties (photometry) and its motion relative to the platform, along with the motion of independently moving objects and targets within (dynamics).

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU		

DYNAMIC VISION FOR CONTROL
AFOSR F49620-03-1-0095

FINAL REPORT

Stefano Soatto
University of California at Los Angeles
`http://vision.ucla.edu`

July 27, 2006

Abstract

The goal of this project was to develop analytical and computational tools to make vision a viable sensor for the purpose of control. In order for unmanned vehicles to interact with a complex, unknown, uncertain and dynamic environment, they must be able to sense its shape (geometry), its reflectance properties (photometry) and its motion relative to the platform, along with the motion of independently moving objects and targets within (dynamics). We have recorded progress on all areas, documented in a number of publications in the most prestigious journals, including a number of breakthroughs: (1) We have demonstrated the existence of general-case viewpoint invariants, and used them to design signatures to establish long-range correspondence for persistent tracking and recognition, (2) we have designed the first, and so far the only, algorithms to estimate 3-D shape and non-Lambertian reflectance from images, and proven their optimality, (3) we have proposed non-linear filtering algorithms based on sample-consensus techniques that have proven successful with large percentages of outliers, (4) we have developed stochastic dynamical models of complex visual phenomena (dynamic textures) and designed system identification algorithms to estimate their parameters, (5) we have developed novel metrics to compare dynamical models for the purpose of classification and recognition, or smoke, fire, fog), (6) analyzed the observability of hybrid dynamical models, and designed algebraic-geometric algorithms to estimate their state and identify their parameters, (7) designed the first optimal algorithms to estimate 3-D shape and reflectance from blurred and motion-blurred images. These results have allowed us to (8) further our development of the first ever system to estimate 3-D structure and motion causally and in real time from a monocular sequence of images.

Outcomes at-a-glance

This project has resulted in a number of technical achievements, and some breakthroughs, documented in over 50 **publications**[1] in the most prestigious conferences and journals in the field of Computer Vision. The project also resulted in the publication of **two books**, one published in late 2003 [39], one to be published in November 2006 [30].

Some of the **students and postdocs** involved in the project have found **placement** in prestigious industrial and academic institutions in the US and abroad, including Dr. Daniel Cremers, now Associate Professor at the University of Bonn in Germany, Dr. Paolo Favaro, now Assistant Professor at Heriot-Watt

---

[1] Listed in the reference section as [41, 20, 21, 40, 8, 22, 29, 11, 35, 58, 2, 5, 14, 53, 46, 54, 47, 33, 42, 51, 52, 38, 1, 26, 15, 32, 45, 7, 28, 34, 24, 43, 44, 37, 48, 17, 19, 9, 49, 3, 57, 23, 16, 10, 25, 50, 59, 36, 31, 27, 18, 56, 55]

University in Edimburgh, U.K. after stints at Cambridge University and Siemens Corporate Research in Princeton, Dr. Hailin Jin, now member of the technical staff at Adobe Research, Santa Clara, Dr. Gianfranco Doretto, now member of the technical staff at General Electric (GE) Global Research, in Niskayuna - NY, Siddharth Manay, now member of the technical staff at Lawrence Livermore National Labs in Livermore - CA, Dr. Rene Vidal, now Assistant Professor at Johns Hopkins University, Dr. Alessandro Chiuso, now Associate Professor at the University of Padova - Italy, Dr. Emmanuel Prados, now Researcher at INRIA, Grenoble - France.

This project has resulted in a number of **collaborations** across departments and institutions, including Prof. Anthony Yezzi (Georgiatech), Prof. Sanjoy Mitter (MIT), Prof. Stanley Osher (UCLA), Prof. Judea Pearl (UCLA), Prof. Shankar Sastry (Berkeley), Prof. Olivier Faugeras (INRIA), Prof. Roberto Cipolla (Cambridge), Prof. Sir Mike Brady (Oxford), Mario Sznaier (Penn State), Alan Willsky (MIT).

Our work has also sparked the attention of several **companies** that have supported or are supporting corollary activities, including Honda (Mr. Jason Meltzer and Mr. Alessandro Bissacco were Summer Interns in 2004 and 2005 respectively), Intel (Mr. Paolo Favaro was an intern, and they supported the lab with generous equipment donations), Toshiba (collaborated on a stereo project, will donate $75K to my laboratory and become Gold Affiliate Member of our Department), Panasonic (donated $50K to my laboratory), Sony (donated $50K to my laboratory and became Gold Affiliate Member of our Department), Mobileye, INC. (collaborated on visual analysis of traffic scenes), Aerovironment (collaborated on an STTR) and Robotics Research (collaborated on an STTR).

Technical Achievements

This section summarizes the technical achievements during the course of the project.

Dynamic Textures

One of the goals of this project was to develop dynamical models of visual scenes that support control and decision tasks. For instance, one may want to identify dynamical systems that correspond to visual processes such as smoke, foliage, steam, or a walking/limping person in order to then be able to recognize a novel instance of such a process. In [19, 49, 3, 18] we have shown that (1) one cannot recover the correct (Euclidean) model of a complex scene from visual data alone; however (2) one can develop a statistical model of the dynamics and spatial statistics of the scene. In order for such a model to support control/decision tasks, the model must have compression power (re-create the data with less complexity than the data themselves), as well as predictive power (synthesize novel sequences that match the spatio-temporal statistics of the original data). In [16] we have shown that one can spatially segment an image into regions that have homogeneous dynamic signatures, and therefore detect and localize dynamic processes such as smoke or fire in a scene. This is the first ever scheme proposed for such a task.

In addition, we have shown how to model the spatial statistics within each segment. This is fundamental for a number of reasons: first, because modeling spatially homogeneous statistics allows for a significant reduction in model complexity, and hence allows for significant *compression*. Second, because a generative model of the spatial statistics allows for a metric and probabilistic structure that enables *recognition* and *synthesis*. Therefore, such a model could be used to extrapolate existing dynamic textures in time *and* space. In [17], we have presented a very efficient model of the joint spatio-temporal statistics that is based on existing work on multi-scale autoregressive models (MAR). While work in MAR concentrated on the dynamics across scale (space), we have extended this to space and time, with the result of having very simple and efficient algorithms that model the joint spatio-temporal statistics.

A Dynamical Systems approach has been proposed for modeling the temporal variation of shape and appearance of a sequence of images, for the purpose of recognition as well as matching. This model, while strongly non-linear, is conditionally linear, in the sense that given the spatial statistics (geometry), the photometric statistics are linear (photometry), and given geometry and photometry, the dynamics are linear (linear dynamical system). We have derived novel system identification procedures based on alternating minimization schemes [20, 21].

Motion analysis

The problem of segmenting dynamically homogeneous regions is not limited to dynamic textures, or to the assumption of spatial homogeneity. Therefore, we have investigated the use of other signatures to segment dynamic scenes, for instance the spatial statistics of the gradient field, which yields novel algorithms for optical flow estimation and segmentation, published in [11, 7, 34, 9, 13, 6, 12]. This has been key to the introduction of shape priors for segmentation, described next.

Shape analysis and non-linear shape statistics

Even in the presence of an explicit generative model, segmentation is an ill-posed inverse problem, and therefore it can benefit significantly from non-generic priors, that is prior assumptions on the objects or scenes that are expected. This is very important in applications in ATR as well as in the detection and tracking of complex dynamic objects such as the human body. In [7] we have presented a novel class of geometric invariants for level set segmentation that exploits the temporal dynamics of the sequence.

Shape classification

A proper theory of shape, that can support shape analysis, segmentation, classification etc. is still elusive. Properly defining the space where "shapes" live has been a challenge for decades, and the problem presents a formidable challenge even for closed planar contours. While for finite-dimensional collections of points there are obvious choices for metrics and consequently the definition of a probabilistic structure on the set of shapes is fairly straightforwardly in the realm of statistical shape spaces, for contours, or infinite-dimensional curves, this is yet an open problem.

We have introduced a new framework called *integral invariants* that combines the benefits (invariance, locality) of differential invariants while being based on integral computations as opposed to differential [43]. The result is that we can compare and recognize targets despite occlusions and despite significant amounts of noise, where differential invariants failed.

We have also furthered our research on the design of robust invariant descriptors for shape matching and recognition. Progress includes the design of a new shape feature that enforces local shape knowledge while allowing global deformations. We hope to be able to use it to recognize broad categories of objects, such as vehicles or tools [41, 40].

Shape reconstruction

In [35] we have consolidated all our previous result on multi-view reconstruction for photometrically complex scenes (non-Lambertian), and in [22] we have reported on the handling of occlusions, a fundamental and strongly non-linear phenomenon that is peculiar to visual information processing. Our work on multi-motion segmentation, that is the problem of the partitioning of a dynamic scene into independently moving objects, has alsoappeared in print [57]. The problem of segmentation of independent motion is crucial in applications to autonomous control, where the scene contains significant clutter. Our variational framework for motion segmentation, developed during the course of the past two years, has also appeared om [11].

This task is important for autonomous navigation and in general interaction tasks with unknown and unstructured environments. Towards this goal, we have developed techniques to learn and enforce shape priors in segmentation and shape recognition [9], a novel representation of shape as the solution of a linear partial differential equation with two-point boundary conditions that has a (quasi)-linearity property [23], techniques to segment spatio-temporal image statistics using variational techniques, implemented via numerical solution of partial differential equations in the level set framework [10], techniques to exploiting occlusions in multiview geometry towards reconstruction of scene geometry despite obstacles [25].

In a related line of work, we have developed techniques to reconstruct *both* the geometry and the photometry of complex scenes [50, 59, 36]. The latter is a breakthrough because nobody before could estimate both shape and photometry for non-Lambertian scenes using only passive non-contact sensors.

We have proposed the framework of *stereoscopic segmentation* where multiple images of the same obejcts were jointly processed to extract geometry and photometry. We have added the explicit estimation of illumination, by factoring illumination and reflectance from the radiance tensor field [60, 35, 34, 37]. In addition, we have further expanded and perfected our shape and radiance estimation techniques to include piecewise smooth albedos, with explicit estimation of the discontinuities. This can be thought of as performing functional segmentation a' la Mumford-Shah on an evolving surface, an important theoretical problem in its own right. Its practical impact is extremely important since many man-made objects are made with composition of different material and hence their reflectance is necessarily piecewise constant.

Dynamic human motion modeling

In order to develop models of human gaits, we must first detect, localize, and track human motion from a distance. Some of these tasks call for non-standard system identification problems, such as the filtering and identification of hybrid systems, where continuous dynamics (e.g. a walking gait, or moving water) is alternated with discrete events (e.g. a change in gait, for instance running, or an explosion). We have been involved in developing algorithms and analysis for linear hybrid systems, offering the first ever characterization of observability and identifiability of such a class of models [56, 55, 57].

In [4, 48] we have reported progress on the modeling of the dynamics of human and facial motion with an eye towards classification and decision tasks. We have developed identification and classification algorithms for simple classes of hybrid dynamical systems (piecewise autoregressive) to model human motion while factoring out contact forces or the characteristics of the terrain [2, 5, 1].

Identification of hybrid dynamical models

In [57, 56] we have reported on continuing efforts to develop a theory for filtering and identification for piecewise linear systems (also called switching linear systems, or jump-Markov systems). These are important in order to perform dynamic scene analysis since one wants to simultaneously track objects of interest and classify modes of operation (e.g. a walking person suddenly running or limping; a traffic scene suddenly jamming; a crowd suddenly behaving uncontrollably).

Visual localization, mapping and navigation

In [45, 44] we have reported on a new representation of salient visual feature that exploits changing viewpoint in order to arrive at a representation that is illumination and viewpoint invariant. This is an important development, since our descriptor is so far the only one that can provably guarantee these characteristics. The descriptor, in turn, is used to automatically extract visual landmarks that can then be used for autonomous navigation, as we have shown in [45] in an application to vision-based autonomous localization and mapping.

Dynamic optical compensation for tracking

For moving vehicles, one wants to detect motions that are as large as possible in order to optimize the SNR (large motions are more informative). However, due to finite aperture and finite shutter time, large motions are difficult to estimate. We have addressed these issues in [28, 24, 25], where we have shown that as a side effect of motion blur estimation one can retrieve the 3-D structure and motion of the scene, thus complementing other visual cues such as stereo and motion.

Accommodation

A separate line of work exploited the optical properties of a finite aperture lens to reconstruct shape and radiance [31], and to diffuse occlusions so as to be able to reconstruct scenes despite obstacles that are small relative to the aperture [27]. We are now collating the results obtained during the course of the past 3 years into a book, to be published by Springer Verlag in late 2006 [30].

Robust Filtering

State estimation and prediction for dynamical models is one of the classical problems in Systems Theory, with important applications in control. In Dynamic Vision, 3-D motion and structure estimation can be posed as a filtering problem, but this formalization does not conform to the standard models that have been extensively studies since the sixties. First, the state-space is of very large dimension, typically in the order of a few hundreds to tens of thousands of states (in Dynamic Textures [2,3] the state is 297,600-dimensional). Second, the state space has time-varying dimension (singular perturbations) due to occlusions in the scene (portions of the scene appear and disappear when they enter/exit the field of view or become occluded by objects within), and finally the noise model is very heavily non-Gaussian, both because the state-space and the dynamics are non-linear, and because of the large portion of outliers. We have proposed a filtering framework based on random sample consensus (RANSAC) that has proven successful with a large portion of outliers, up to 85% [18]. The results have been compared with the state of the art, including kernel-correlation-based approach, and the first results indicate a significant improvement both in terms of accuracy, robustness to outliers, and performance (localization error).

Viewpoint Invariance

We have proven that viewpoint invariants always exist, for scenes with arbitrary shape and albedo. We have also derived generic features (invariant statistics) that can be used for matching despite changes in viewpoint [19], including a growth procedure that touches upon variational region-based segmentation techniques that have been so successful on the field [12]. Furthermore, we have performed an empirical study of the natural statistics of image deformations due to changes of viewpoint [14]. This is important because, although one scan construct viewpoint invariants for arbitrary geometries, natural scenes have very peculiar geometric statistics, so instead of designing descriptors that are invariant in the worst-case sense (and therefore lose discriminative power) one can construct statistics that are only invariant to scenes that are likely, or that respect the natural statistics. Our study follows the lines of Mumford's and Geman's work on the study of natural image statistics, but extended to deal with dynamic imagery and the subsequent image deformations.

Classification of Dynamical Models

The problem of performing decisions in the space of dynamical models, for instance in order to recognize events by their dynamic characteristics, or to perform fault detection, or to compute the mean plant and the plant uncertainty from repeated identification trials, is a fundamental problem in system theory and one that has received remarkably little attention during the past four decades. Even for the simple case of linear finite-dimensional systems, no thorough study on the structure of probability measures on such spaces is available, and even metric structures are usually restricted to cord distances (i.e. distances that do not preserve the geometric structure of the space of models) that only take into account the steady-state response, for instance Martin's distance for SISO systems, or subspace angles. We have been after a meaningful notion of distance between models that could take into account, depending on the application, transient behavior, initial conditions, input distribution, as well as steady-state dynamics. Our source of inspiration is the problem of recognizing people from their gait. We are motivated by the fact that humans are able to recognize a familiar person from afar by the way they walk, regardless of pose, clothing etc. In particular, we are interested in recognizing classes of gaits, whether a person is limping, etc. The applications are not limited to visual analysis. Another application of interest is the determination of gait types for soldiers moving on foot in the battlefield from accelerometers mounted on their backpack. Given that humans are collection of controlled inverted penduli, a useful distance should apply to non-minimum-phase models, and given that we are interested in stationary as well as transient behavior, it should apply to marginally stable models. Most common metrics are restricted to the second order statistics of their covariance sequence, which forces the attention on the stable and minimum-phase representation of the plant. We have drawn inspiration from recent work of Smola and coworkers on the definition of Mercer kernels between dynamical models. Unfortunately their definition of distance requires marginalization with respect to the joint density of the inputs of two different plants, which is never available, or it requires that the inputs are independent, in which case the effect of the initial condition is discounted. We have been able to define a kernel distance between dynamical models that can take into account the transient, input distribution and steady-state behavior of models of arbitrary dimension, including non-minimum-phase ones and marginally stable ones. This work reveals interesting connections to optimal transport and independent component analysis, in that in order to eliminate dependency of the distance from the joint density of the inputs of the two models being compared, one has to strongly whiten the inputs (make them spatially and temporally independent), which involves the solution of an independent-component analysis problem, and to define a kernel between arbitrary scalar distributions, which minimizes the Wasserstein distance, an optimal transport problem.

Causal Inference for Event Detection and Analysis

We have begun to explore the potential of Pearl's Causal Analysis framework to represent, and hence recognize, dynamic events of interest. The idea is to factor out the wide variability in spatial and temporal statistics, and only concentrate on causal relationship to describe events of interest such as suspicious behavior of an individual seen in video, or the occurrence of an accident at an intersection monitored with embedded sensors.

Numerical Schemes for Hamilton-Jacobi-Bellman Equations We have developed effective fast-marching algorithms for numerically integrating partial differential equations that arise in shape analysis (shape from shading) and in tensor imaging [46, 47].

Fast Feature Tracking

We have developed a reconfigurable architecture platform for fast feature tracking (up to 300,000 features at 100Hz) [2].

Acknowledgment/Disclaimer

This work was sponsored (in part) by the Air Force Office of Scientific Research, USAF, under grant/contract number F49620-03-1-0095. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed

References

[1] A. Bissacco. Modeling and learning contact dynamics in human motion. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2005.

[2] A. Bissacco and S. Ghiasi. Fast visual feature selection and tracking in a hybrid reconfigurable architecture. In *Proc. of the Workshop on Applications of Computer Vision (ACV)*, June 2006.

[3] A. Bissacco, P. Saisan, and S. Soatto. Image-based modeling andn synthesis of human gaits with subtleties. In *Proc. of the IFAC Symp. on System Identification (SYSID)*, August 2003.

[4] A. Bissacco, P. Saisan, and S. Soatto. Gait recognition using dynamic affine invariants. In *Proc. of the MTNS*, pages TH A8, 1–7, July 2004.

[5] A. Bissacco and S. Soatto. Classifying gaits without contact forces. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2006.

[6] D. Cremers. A variational framework for image segmentation combining motion estimation and shape regularization. In *IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages (1) 53–58, October 2003.

[7] D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: teaching level set to walk. In *Pattern Recognition, Proc. of DAGM, Springer LNCS Vol. 3175*, pages 36–44, Tuebingen 2004.

[8] D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: teaching level sets to walk. *Intl. J. of Comp. Vis.*, in press.

[9] D. Cremers and S. Soatto. A pseudo distance for shape priors in level set segmentation. In *Proc. of the Intl. Workshop on Variational and Level Set Methods*, pages 169–176, October 2003.

[10] D. Cremers and S. Soatto. Variational space-time motion segmentation. In *Intl. Conf. on Comp. Vision*, pages (2) 886–892, October 2003.

[11] D. Cremers and S. Soatto. Motion competition: a variational approach to piecewise parametric motion segmentation. *Intl. J. of Comp. Vision*, pages 249–265, May, 2005.

[12] D. Cremers, N. Sochen, and C. Schnoerr. Towards recognition-based variational segmentation using shape priors and dynamic labeling. In *Intl. Conf. on Scale-Space Theories in Computer Vision*, pages 388–400, June 2003.

[13] D. Cremers and A. Yuille. A generative model based approach to motion segmentation. In *German Conf. on Pattern Recognition (DAGM)*, pages 313–320, September 2003.

[14] F. Cuzzolin. Using bilinear models for view-invariant action and identity recognition. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2006.

[15] G. Doretto. Modeling dynamic scenes with active appearance. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2005.

[16] G. Doretto, D. Cremers, P. Favaro, and S. Soatto. Dynamic texture segmentation. In *Intl. Conf. on Comp. Vision*, pages (2) 1236–1242, October 2003.

[17] G. Doretto, E. Jones, and S. Soatto. Spatially homogeneous dynamic textures. In *Proc. of the Eur. Conf. on Comp. Vision*, pages 591–602, May 2004.

[18] G. Doretto and S. Soatto. Editable dynamic textures. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages II–137–142, June 2003.

[19] G. Doretto and S. Soatto. Towards plenoptic dynamic textures. In *Proc. of the Workshop Textures 2003*, pages 25–30, October 2003.

[20] G. Doretto and S. Soatto. *The Handbook of Mathematical Models in Computer Vision*, chapter Modeling dynamic scenes: an overview of Dynamic Textures. Springer Verlag, 2005.

[21] G. Doretto and S. Soatto. Dynamic shape and appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, in press 2006.

[22] A. Duci, A. Yezzi, S. Mitter, and S. Soatto. Region matching with missing parts. *Image and Vision Computing*, 24(3):271–277, 2006.

[23] A. Duci, A. J. Yezzi, S. K. Mitter, and S. Soatto. Shape representation via harmonic embedding. In *Intl. Conf. on Comp. Vision*, pages 656–662, October 2003.

[24] P. Favaro, M. Burger, and S. Soatto. Scene and motion reconstruction from defocused and motion-blurred images via anisotropic diffusion. In *Proc. of the Eur. Conf. on Comp. Vision*, pages 257–269, May 2004.

[25] P. Favaro, A. Duci, Y. Ma, and S. Soatto. On exploiting occlusions in multiple view geometry. In *Intl. Conf. on Comp. Vision*, pages 479–486, October 2003.

[26] P. Favaro, H. Jin, and R. Cipolla. Tracking with motion blur. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2005.

[27] P. Favaro and S. Soatto. Seeing beyond occlusions (and other marvels of a finite lens aperture). In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages II–579–586, June 2003.

[28] P. Favaro and S. Soatto. A variational approach to scene reconstruction and image segmentation from motion blur cues. In *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog.*, July 2004.

[29] P. Favaro and S. Soatto. A geometric approach to shape from defocus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005.

[30] P. Favaro and S. Soatto. *3-D Shape Estimation and Image Restoration: Exploiting Defocus and Motion Blur.* Springer Verlag, November 2006 (in copyediting).

[31] P. Favaro, S. Soatto, L. Vese, and S. J. Osher. Shape from anisotropic diffusion. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages I–179–186, June 2003.

[32] J. Jackson, A. Yezzi, and S. Soatto. Tracking deformable moving objects under severe occlusions. In *Proc. of the IEEE Conf. on Decision and Control*, December 2004.

[33] J. D. Jackson, A. J. Yezzi, and S. Soatto. Dynamic shape and appearance modeling via moving and deforming layers. In *Proc. of the Workshop on Energy Minimization in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 427–448, 2005.

[34] H. Jin, D. Cremers, A. Yezzi, and S. Soatto. Shedding light in stereoscopic segmentation. In *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog.*, July 2004.

[35] H. Jin, S. Soatto, and A. Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *Intl. J. of Comp. Vis.*, 63(3):175–189, 2005.

[36] H. Jin, S. Soatto, and A. J. Yezzi. Multi-view stereo beyond lambert. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages I–171–178, June 2003.

[37] H. Jin, A. J. Yezzi, and S. Soatto. Region-based segmentation on evolving surfaces with application to 3d shape and radiance estimation. In *Proc. of the Eur. Conf. on Comp. Vision*, pages 114–125, May 2004.

[38] E. Jones and S. Soatto. Layered active appearance models. In *Proc. of the Intl. Conf. of Comp. Vision*, October 2005.

[39] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An invitation to 3D vision, from images to models.* Springer Verlag, 2003.

[40] S. Manay, D. Cremers, B.-W. Hong, A. Yezzi, and S. Soatto. Integral invariants and shape matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, in press 2006.

[41] S. Manay, D. Cremers, B.-W. Hong, A. Yezzi, and S. Soatto. *Statistical analysis of shapes (modeling and simulation in science, engineering and technology)*, chapter Integral Invariants and Shape Matching, pages 137–167. Birkhauser, 2006.

[42] S. Manay, D. Cremers, A. Yezzi, and S. Soatto. One-shot integral invariant shape priors for variational segmentation. In *Proc. of the Workshop on Energy Minimization in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 414–426, 2005.

[43] S. Manay, A. J. Yezzi, B. W. Hong, and S. Soatto. Integral invariant signatures. In *Proc. of the Eur. Conf. on Comp. Vision*, pages 87–94, May 2004.

[44] J. Meltzer, M. Yang, R. Gupta, and S. Soatto. Multiple view feature descriptors from image sequences via kernel principal component analysis. In *Proc. of the Eur. Conf. on Comp. Vision*, pages 215–227, May 2004.

[45] J. Meltzer, M.-S. Yang, R. Gupta, and S. Soatto. Simultaneous localization and mapping using multiple view feature descriptors. In *Proc. of the IEEE Intl. Conf. on Intelligent Robots and Systems (IROS)*, volume 2, pages 1550–1555, Sendai, Japan, October 2004.

[46] E. Prados, C. Lenglet, J.-P. Pons, O. Faugeras, and S. Soatto. Control theory and fast marching techniques for brain connectivity mapping. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2006.

[47] E. Prados and S. Soatto. Fast marching method for hamilton-jacobi-bellman equations and generic shape from shading. In *Symposium on Variational and Level Set Methods (VLSM)*, October 2005.

[48] P. Saisan, A. Bissacco, and A. Chiuso S. Soatto. Modeling and synthesis of facial motion driven by speech. In *Proc. of the Eur. Conf. on Comp. Vision*, pages 456–467, May 2004.

[49] S. Soatto and A. Chiuso. Snippets of identification theory in computer vision. In *Proc. of the IFAC Symp. on System Identification (SYSID)*, August 2003.

[50] S. Soatto, A. J. Yezzi, and H. Jin. Tales of shape and radiance in multiview stereo. In *Intl. Conf. on Comp. Vision*, pages 974–981, October 2003.

[51] A. Vedaldi, H. Jin, P. Favaro, and S. Soatto. Kalmansac: Causal inference of dynamical processes in the presence of outliers. In *Proc. of the Intl. Conf. on Comp. Vision*, October 2005.

[52] A. Vedaldi and S. Soatto. Features for recognition: viewpoint invariance for non-planar scenes. In *Proc. of the Intl. Conf. of Comp. Vision*, October 2005.

[53] A. Vedaldi and S. Soatto. Local features all grown up. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2006.

[54] A. Vedaldi and S. Soatto. Viewpoint induced deformation statitics and the design of viewpoint invariant features: singularities and occlusions. In *Eur. Conf. on Comp. Vision (ECCV)*, 2006.

[55] R. Vidal, A. Chiuso, and S. Soatto. Observability and identifiability of jump-linear systems. In *Proc. of the Intl. Conf. on Decision and Control*, pages 3614–3619, December 2002.

[56] R. Vidal, A. Chiuso, S. Soatto, and S. Sastry. Observability of linear hybrid systems. In *Proc. of the Hybrid Systems Computation and Control*, pages 526–539, October 2003.

[57] R. Vidal, S. Soatto, and S. Sastry. An algebraic geometric approach to the identification of linear hybrid systems. In *IEEE Conf. on Decision and Control*, December 2003.

[58] D. Wang, E. Prados, and S. Soatto. Towards robust and physically plausible shaded stereoscopic segmentation for piecewise constant albedo. In *Proc. of the 3D Cinematography Workshop (3DCine) in conjunction with CVPR 2006*, June 2006.

[59] A. J. Yezzi and S. Soatto. Structure from motion for scenes without features. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages I–525–532, June 2003.

[60] A. J. Yezzi, S. Soatto, H. Jin, A. Tsai, and A. Willsky. *Geometric Level Set Methods in Imaging, Vision and Graphics, S. J. Osher and N. Paragios (Eds.)*, chapter Mumford-Shah from segmentation to stereo, pages 207–228. Springer Verlag, 2003.

Personnel Supported During Duration of Grant

Stefano Soatto, Professor, University of California, Los Angeles.
Jason Meltzer, Graduate Student, University of California, Los Angeles (partial support).
Andrea Vedaldi, Graduate Student, University of California, Los Angeles (partial support).
Eagle Jones, Graduate Student, University of California, Los Angeles (partial support).
Philip Venturelli, Graduate Student, University of California, Los Angeles (partial support).
Daniel Cremers, Postdoctoral Researcher, University of California, Los Angeles. (partial support)
Hailin Jin, Postdoctoral Researcher, University of California, Los Angeles (partial support).
Byung-Woo Hong, Visiting Scholar, University of California, Los Angeles (partial support.
Paolo Favaro, Postdoctoral Researcher, University of California, Los Angeles (partial support).
Gianfranco Doretto, Graduate Student, University of California, Los Angeles (partial support).
Chen Avin, Graduate Student, University of California, Los Angeles (partial support).
Judea Pearl, Professor, University of California, Los Angeles (partial support).

Publications

See references above.

Honors & Awards Received

- Semi-plenary speaker at the IFAC Symposium on System Identification.
- Associate Editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence (2003–).
- Member of the Editorial Board, International Journal of Computer Vision.
- Program Co-Chair, IEEE Computer Vision and Pattern Recognition, 2005
- Outstanding Poster award from the IEEE Computer Society for their presentation on Motion Blur Segmentation associated with the publication [28] at the IEEE Intl. Conf. on Comp. Vis. and Pattern Recognition.
- UAI Best Paper Award, Association of Uncertainty in Artificial Intelligence, Paper titled: Identification of Conditional Interventional Distributions, I. Shipster and J. Pearl.
- Finalist in the Second DARPA Grand Challenge: Team Lead of UCLA/Golem, S. Soatto.
- Member of the Technical Advisory Board, TGAL (terrain-guided automatic landing), ST9 NASA New Millennium Program, 2006.

- During the 3-year period between the approval (2002) and the completion (2005) of this project, the PI went from Asssistant Professor to Full Professor, effective July 1, 2005, at the Henry Samueli School of Engineering and Applied Sciences at UCLA.

Transitions

3-D Reconstruction and Motion Estimation being deployed into micro-UAV by Aerovironment, as part of an STTR Project (contact: Rick Pedigo, Aerovironment). Real-time Structure from Motion will be implemented on autonomous vehicle to compete in third DARPA Grand Challenge (Urban Challenge) in future months.

Patents

US Patent 6,944,327, Apparatus and Method for the Interactive Customization of Eyeglass Frames, September 13, 2005.

AFRL Point of Contact

Dr. Sharon Heise, Program Manager, Dynamics and Control AFOSR/NM, 4015 Wilson Blvd., Rm. 713, Arlington, VA 22203-1954, Voice: 703-696-7796, Fax: 703-696-8450.