DEFENSE TECHNICAL INFORMATION CENTER

Information for the Defense Community

DTIC® has determined on 25/03/2009 that this Technical Document has the Distribution Statement checked below. The current distribution for this document can be found in the DTIC® Technical Report Database.

☒ **DISTRIBUTION STATEMENT A.** Approved for public release; distribution is unlimited.

☐ **© COPYRIGHTED**; U.S. Government or Federal Rights License. All other rights and uses except those permitted by copyright law are reserved by the copyright owner.

☐ **DISTRIBUTION STATEMENT B.** Distribution authorized to U.S. Government agencies only (fill in reason) (date of determination). Other requests for this document shall be referred to (insert controlling DoD office)

☐ **DISTRIBUTION STATEMENT C.** Distribution authorized to U.S. Government Agencies and their contractors (fill in reason) (date of determination). Other requests for this document shall be referred to (insert controlling DoD office)

☐ **DISTRIBUTION STATEMENT D.** Distribution authorized to the Department of Defense and U.S. DoD contractors only (fill in reason) (date of determination). Other requests shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT E.** Distribution authorized to DoD Components only (fill in reason) (date of determination). Other requests shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT F.** Further dissemination only as directed by (inserting controlling DoD office) (date of determination) or higher DoD authority.

Distribution Statement F is also used when a document does not contain a distribution statement and no distribution statement can be determined.

☐ **DISTRIBUTION STATEMENT X.** Distribution authorized to U.S. Government Agencies and private individuals or enterprises eligible to obtain export-controlled technical data in accordance with DoDD 5230.25; (date of determination). DoD Controlling Office is (insert controlling DoD office).

AFRL-SR-AR-TR-09-0076

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
02/24/2009

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
1/1/06 11/30/09

4. TITLE AND SUBTITLE
Dynamic Vision for Control

5a. CONTRACT NUMBER
FA9550-06-1-0138

5b. GRANT NUMBER
n/a

5c. PROGRAM ELEMENT NUMBER
n/a

6. AUTHOR(S)
Stefano Soatto

5d. PROJECT NUMBER
n/a

5e. TASK NUMBER
n/a

5f. WORK UNIT NUMBER
n/a

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, Los Angeles

8. PERFORMING ORGANIZATION REPORT NUMBER
n/a

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR
875 N Randolph St
Arlington, VA 22203

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

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
FA9550-06-1-0138

12. DISTRIBUTION/AVAILABILITY STATEMENT
public

13. SUPPLEMENTARY NOTES
n/a

14. ABSTRACT

The on-going goal of this pro ject is to develop analytical tools and computational models for vision to be used as a sensor for the purpose of control. Vision, as in remote passive distributed sensing, whether in the visible or other spectra, is a flexible, powerful and cheap sensory modality for unmanned vehicles to interact with complex, unknown, uncertain and dynamic environments. To that end, algorithms and models must be developed to causally infer geometric (shape), photometric (reflectance) and dynamic (motion) properties of ob jects and scenes. In this report we describe progress on all areas, including the following breakthroughs:
1. We have fully developed, after their introduction early into the pro ject, Dynamic Active Appearance Models [15, 16] to describe variations in shape (domain deformation), reflectance (contrast) and motion via non-linear conditionally Gaussian processes that capture complex phenomena such as the motion of faces, flames, foliage. We have further extended these models to take into account occlusions [illegible] [27], which are fundamental in vision. In addition, we have formulated a series of conjectures and developed a research program for the

15. SUBJECT TERMS
Image Analysis, Tracking, Video Analysis, Autonomous Navigation, Guidance, Vision-based Control

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
public	public	public			Stefano Soatto

19b. TELEPHONE NUMBER *(Include area code)*
(310)869-2972

DYNAMIC VISION FOR CONTROL
FA9550-06-1-0138

FINAL REPORT

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

February 5, 2009

Abstract

The on-going goal of this project is to develop analytical tools and computational models for vision to be used as a sensor for the purpose of control. Vision, as in remote passive distributed sensing, whether in the visible or other spectra, is a flexible, powerful and cheap sensory modality for unmanned vehicles to interact with complex, unknown, uncertain and dynamic environments. To that end, algorithms and models must be developed to causally infer geometric (shape), photometric (reflectance) and dynamic (motion) properties of objects and scenes. In this report we describe progress on all areas, including the following breakthroughs:

1. We have fully developed, after their introduction early into the project, *Dynamic Active Appearance Models* [15, 16] to describe variations in shape (domain deformation), reflectance (contrast) and motion via non-linear conditionally Gaussian processes that capture complex phenomena such as the motion of faces, flames, foliage. We have further extended these models to take into account occlusions phenomena [27], which are fundamental in vision.

2. We have developed filtering and identification techniques for a class of (Hammerstein) dynamical models driven by non-Gaussian processes that are particularly well suited to model human motion in video [3, 43, 6]; we have also developed computational and modeling tools to enforce priors on dynamically moving shapes [11, 26] to enable tracking through occlusions or with partial information.

3. We have developed forward diffusion models and numerical schemes for integrating the ensuing partial differential euations for shape estimation from diffusion images (defocus or motion blur) [20].

4. We have proven the observability and identifiability of ego-motion in the presence of visual and inertial measurements [29], and characterized the set of sufficiently exciting inputs. In addition, we have proved that the landscape of local minima in visual motion estimation moving forward can be removed by positivity constraints [47].

5. We have developed a framework for visual tracking of severely deforming objects, by proposing the first ever (infinite-dimensional) observer capable of predicting general (diffeomorphic) deformations, rather than just a finite-dimensional group (unpublished).

6. We have been able to characterize visual invariants to general viewpoint and contrast changes (unpublished).

In addition, we have formulated a series of conjectures and developed a research program for the classification of time series [42, 40], together with the first steps to build dynamic action dictionaries [41, 54]. Other contributions include algorithms for vision-based and lidar-based localization [21], proximity distribution kernels for visual category recognition [32], and wide-sense filtering on Lie groups [10]. Some of this work has been conducted in collaboration with Stan Osher at UCLA [11, 20], René Vidal at Johns Hopkins [56], Anthony Yezzi at Georgia Tech.

Outcomes at-a-glance

This project has resulted in a number of technical achievements, and some breakthroughs, documented in 47 **publications**[1] in the most prestigious conferences and journals in the field of Computer Vision, including a texbook [19].

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. Andrea Vedaldi – awarded the only Outstanding PhD Award from the UCLA Computer Science Department in 2008 – now a Postdoc at Oxford University after turning down a Faculty Position at the Ecole Centrale in Paris, Byung-Woo Hong, now Assistant Professor at Chung-Ang University in Seoul – Korea, Haibin Ling, now Assistant Professor at Temple University, Gregorio Guidi, now an Analist at the Central Bank of Italy, Daniele Fontanelli, a researcher at the University of Pisa, and Alessandro Bissacco now Research Engineer at Google INC., Emmanuel Prados, Researcher at INRIA, Grenoble - France.

Our work has also sparked the attention of several **companies** that have supported or are supporting corollary activities, including Mitsubishi Heavy Industries, Toshiba, Sony and Panasonic of Japan, who supported a staff researcher to visit the Vision Laboratory for a year and resulted in state-of-the art algorithms for human detection in video (Toshiba) and stereo reconstruction (Panasonic), and Mobileye, INC. (Israel) that collaborated with the Vision Lab during the 2005 DARPA Grand Challenge prior to the commencement of this project.

Technical Achievements

Dynamic Active Appearance Models

Images are generated by a complex interplay of photometric (illumination, reflectance), geometric (shape, pose) and dynamic (motion deformation) properties of the scene. These factors are intermingled, and non-uniquely identifiable in an image (or even a sequence of images), so that there are infinitely many combinations of photometric, geometric and dynamic processes that generate the same image sequences. When the goal is *not* to control each factor individually, but instead to perform classification, for instance the detection, recognition and localization of objects or events of interest, one seeks for the simplest possible models that capture the phenomenology of the data. In the past, we have developed models of so-called *dynamic textures* [14, 13] that capture temporally and spatially stationary sequences. This model has been extended further to represent spatially non-stationary, photometrically and temporally stationary sequences, which we have called Dynamic Active Appearance Models (DAAM) [16, 15].

DAAM are considerably more powerful than dynamic textures, which can be used to represent the same data with lower complexity, or to capture more complex phenomena such as a flag waving in the wind or a talking face with greater fidelity at equal complexity. The price we pay for that benefit is an increased complexity of the inference process (filtering and identification). In fact, whereas Dynamic Textures were essentially linear-Gaussian dynamical models, DAAM are intrinsically non-linear. However, the models are *conditionally Gaussian* in the sense that shape deformation can be characterized as a Gaussian

[1] Listed in the reference section as [10, 49, 43, 42, 16, 19, 35, 23, 25, 27, 20, 28, 3, 46, 15, 34, 11, 17, 18, 58, 38, 41, 22, 54, 48, 61, 36, 24, 21, 40, 29, 62, 32, 26, 6, 8, 33, 47, 56, 52, 7, 60, 4, 5, 12, 51, 39, 53]

shape space (the quotient of $\mathbb{R}^2$ modulo the affine group); conditioned on shape, reflectance is represented as an affine shape space ($\mathbb{R}^+$ modulo monotonic continuous transformations); conditioned on reflectance and shape, the temporal evolution of the (finite-dimensional) representation is modeled as the output of a linear-Gaussian dynamical model. So, although non-linear and non-Gaussian, the model is conditionally Gaussian. Unfortunately, filtering and identification of these models is no longer simply performed with standard subspace identification techniques [56], but instead finite-element methods and identification of hybrid systems must be brought to bear.

As we have done for the case of Dynamic Textures in the past, after developing the model and testing its generative power by measuring the matching fidelity of the statistics of second and higher-order of the data, we intend to exploit these models for the purpose of decision, specifically to detect, localize, classify and recognize events of interest in video. Whereas data fidelity and complexity are all that matters for communication purposes (transmission, compression), recognition requires handing nuisance transformations and endowing the models with a metric that enables efficient computation of distances and the proper definition of prior probability measures. This direction of investigation will be pursued in future work.

Dynamical Models of Human Motion

Human motion presents a significant challenge due to the importance of the application domain (security, persistent ISR), the variability in which they can appear in images (different clothing, different pose, different illumination, partial occlusions) and move (different gaits), their complex dynamics (humans are essentially multiple inverted penduli). Our goal is to develop techniques to detect, localize and recognize actions regardless of the individual, and to recognize the individual regardless of the action, clothing, pose, illumination etc. The first step in this program is to extract a time series from the measurements. This can be done in a number of ways, form trivial (consider the video itself as a time series of pixel intensity values) to complex (estimate silhouettes of moving humans, seen from a moving platform [30], including the enforcement of prior knowledge on their shape and deformation [11, 25, 17]). Once that is done, we cannot simply compare two time series as if they were two functions of time, using any number of functional norms, because of the large variability due to nuisance factors (pose, initial condition, speed of execution etc.). Long ago, in [2], we have been the first to propose interpreting such time series as the output of dynamical models, and to perform decisions – such as the recognition of an individual form her gait – by comparing statistics of the models identified from the data. That has worked very well for stationary sequences, for instance segments of walking, running, jumping, limping etc. In the past we have also extended these models to piecewise stationary statistics, which led us to develop techniques to perform filtering and identification of hybrid (jump-linear) models [57, 55, 59]. However, for more complex and transient motions, it was not clear at the beginning of this project how one could factor out nuisance factors such as the speed of execution of an action, or the initial condition, in a computationally efficient manner. During the course of this project, we have developed (Mercer) Kernels between dynamical models that enables one to compare them while discounting the effects of various factors, depending on the application, including initial time, initial condition, distribution of the input sequences, speed of execution etc. [3]. This work required considerable care in order to properly treat the non-Gaussian statisics of the input, which are filtered through non-minimum phase models.

In addition, we have also developed techniques for imposing prior knowledge on the evolving shape of objects [11], and most recently we have developed (infinite-dimensional) models to identify the state of a model tracking a deforming shape. This enables the prediction not just of the motion of an object, but also of its *deformation*, which enables tracking objects through severe occlusions [26], whereas previous schemes could predict pose, but not shape after the target became unoccluded.

Shape From Anisotropic Diffusion

During the course of this project we have completed a comprehensive research program, initiated in previous years, on the exploitation of diffusion cues, from motion blur to defocus to confocal imaging, for the estimation of shape and reflectance properties of scenes. In later investigations we have also extended this to infer independent motions in a scene. This work has been collated in a textbook published by Springer Verlag in December 2006 [19].

In particular, one approach used for shape and motion optimization has proven especially successful, as we describe next. In the presence of unknown motion or unknown optical configuration, the imaging process is essentially a (blind) convolution process. Hence, most prior work on shape and motion inference focused on blind deconvolution, a notoriously ill-posed, numerically ill-conditioned process. In [20], in collaboration with former student P. Favaro (now at Heriot-Watt and the University of Edimburgh) and colleague S. Osher, we have proposed a revolutionary approach based on a forward diffusion with a space-varying stopping time. This has vastly exceeded the state of the art in both accuracy and computational complexity in the estimation of shape from defocus as well as motion blur.

Visual-Inertial Integration

It is generally known that vision and inertial are complementary modalities: The former is slow, global – modulo visbility – but yields only measurements up to a Euclidean reference and an unknown scale; the latter is fast and local, subject to drift. Several approaches have been proposed to integrate the two, but all of them, without exception that we know of, fail to address the two issues that are of most critical important in practical application, that is (a) the calibration between the camera reference frame and the inertial frame (usually assumed known through some delicate metrology), and (b) the dealing with gravity (usually assumed known through the usual cohort of ad-hoc fixes common in the inertial navigation practice). This poses an obstacle to the deployment of vision-inertial measurements, when (a) no accurate calibration is provided, and (b) when small errors in gravity (usually estimated by averaging acceleration over a long time frame) cause long-range divergence of the ensuing filter.

In [29] we have eviscerated the observability and identifiability of a full Euclidean frame from joint vision and inertial measurements. We have been able to show that (a) the camera-imu calibration is *identifiable*, so it does not need to be accurately known ahead of time and can be refined on-line, and (b) gravity is *observable* from joint vision and inertial measurements, so it can be updated on-line while the overall filter is guaranteed to be observable (unlike the case of pure inertial navigation). The observability/identifiability conditions impose that the input sequence be *sufficiently exciting*. We have characterized the sufficiently exciting inputs in terms of a motion sequence, corresponding to an "autocalibrating motion" that can be performed at the beginning of an experiment, without the need for additional instrumentation.

Since the motion usually undergone during steady-state guidance (constant velocity) does *not* correspond to a sufficiently exciting input, in general the gravity and camera-imu calibration states can drift during steady-state motion. Therefore. While in practice this problem can be fixed in practice with suitable gain scheduling techniques, there remains the need to analyze the overall identifiability/observability of a dynamical model whose (non-linear) state-space is partitioned into regions that are observable, and regions where the parameters of the model are not identifiable. Questions we intend to address in the future is whether transients following transitions across such region boundaries are sufficient to absorb the drift (and therefore no specific action is needed in the design of observers), or whether there is an optimal strategy to lock the unobservable states and the unidentifiable parameters when the system goes through regions of state-space, and of inputs, that correspond to non-identifiable, non-observable conditions.

Tracking Deforming Shapes (unpublished)

Tracking deformable objects in video is an important problem in numerous application; we have already described the specific case of humans, but a variety of other objects are of interest. Note that even rigid objects can yield deformable domains when imaged through a projection, and the more complex the three-dimensional shape, the more complex the deformation of the projection. The usual approach to tracking such deformable objects consists in either treating each frame as an independent entity and *segmenting* the object from the background based on pictorial (reflectance) cues, or in tracking a *finite-dimensional* representation of the object corresponding to its coarse motion (e.g. affine). The former approach has nothing "dynamic" to it, and therefore has no predictive power whatsoever (in the presence of missing data, the estimate remains locked to the last available measurement). The latter can extrapolate coarse motion (e.g. centroid and second-moment matrix), but cannot extrapolate deformation. For instance, when a walking person is partially occluded, the approach extrapolates a moving cardboard figure, but not the actual deformation due to individual limb motion [11].

We have proposed what is, to the best of our knowledge, the first approach to design a proper filter (observer) in the infinite-dimensional space of shapes (closed Jordan curves). This is based on endowing the space with a Riemannian (Sobolev) metric, then shooting geodesics from the current best estimate of the state using the exponential map in the infinite-dimensional Lie group of diffeomorphisms, finally correcting the prediction when a new measurement of a curve becomes available. This work has yet to be published, although a technical report has recently been deposited.

The next step in this program consists in making the measurement equation one step closer to the actual data, which are images. So, instead of assuming that we have an intermediate representation that yields a pseudo-measurement (a curve), we intend to measure images directly, thus estending the framework of DEFORMOTION [44] to a proper dynamical system observer framework.

Visual Invariants (unpublished)

It was widely believed that viewpoint invariant image statistics did not exist except for planar object, due to [9]. In [50], we showed that viewpoint invariant statistics always exist, for scenes of arbitrary shape, provided Lambertian reflection. At the same time, a significant avenue of research deals with contrast-invariant image processing, pioneered by Koenderink [31] and Morel [1]. Unfortunately, the statistics that are invariant to viewpoint are not invariant to contrast, and vice-versa. Therefore, we have recently turned our attention to the problem of either identifying viewpoint-and-contrast invariant statistics, or disprove their existence.

We have recently shown that such viewpoint-and-contrast invariants exist, and they are supported on a zero-measure subset of the image domain. They are related to topological constructions derived from the Morse-Smale complex, which we have called *sub-Reeb trees*. While this is already interesting, because it illustrates why it is possible to compress images so efficiently without much perceptual effect, what is remarkable is the fact that this zero-measure object is actually a *sufficient statistic*, meaning that from it one can recover exactly the original image, modulo the action of a domain diffeomorphism (viewpoint change) and contrast transformation (illumination change). This result, which we believe to be of great theoretical significance, has not yet been published, but the main argument and the proofs have been deposited in a technical report [45], coauthored with colleague P. Petersen.

These results pertain to Lambertian scenes and assume that the image is approximated by a Morse function. While this is a fair assumption for many subsets of the image, in the sense that it is possible to find image statistics that are to good approximation piecewise smooth [37], there remains the need to first partition the image into domains where the Morse assumption is satisfied to a reasonable extent. We have recently begun experimenting with various approach for texture segmentation, that would provide a preprocessing step for the construction of a collection of sub-Reb trees. We are studying the theoretical properties that such as "sparse invariant coding" would have in the overall problem of performing decisions from image data, and at the same time we are considering numerical and computational implications that these representations would have, in the sense of enabling the design of more efficient image-based and video-based classification and recognition algorithms.

Acknowledgment/Disclaimer

This work was sponsored (in part) by the Air Force Office of Scientific Research, USAF, under grant/contract number FA9550-06-1-0138. 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] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel. Axioms and fundamental equations of image processing. *Arch. Rational Mechanics*, 123, 1993.

[2] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto. Recognition of human gaits. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages II 52–58, December 2001.

[3] A. Bissacco, A. Chiuso, and S. Soatto. Classification and recognition of dynamical models: the role of phase, independent components, kernels and optimal transport. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(11), 2007.

[4] 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.

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

[6] A. Bissacco and S. Soatto. On the blind identification of time series. In *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Recog.*, June 2007.

[7] A. Bissacco, M.-H. Yang, and S. Soatto. Detecting humans with their pose. In *NIPS*, December 2006.

[8] A. Bissacco, M.-H. Yang, and S. Soatto. Fast pose estimation via multi-dimensional boosting regression. In *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Recog.*, pages 1–8, 2007.

[9] J. B. Burns, R. S. Weiss, and E. M. Riseman. The non-existence of general-case view-invariants. In *Geometric Invariance in Computer Vision*, pages 120–131, 1992.

[10] A. Chiuso, G. Picci, and S. Soatto. *Communications in Information and Systems*, chapter Wide-sense Estimation on the Special Orthogonal Group. 2009.

[11] 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.*, 69(3):335–351, 2006.

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

[13] G. Doretto, A. Chiuso, Y. Wu, and S. Soatto. Dynamic textures. *Intl. J. of Comp. Vis.*, 51(2):91–109, 2003.

[14] G. Doretto and S. Soatto. *The Handbook of Mathematical Models in Computer Vision*, chapter An Overview of Dynamic Textures. N. Paragios, Y. Chen and O. Faugeras (Eds.), Springer Verlag, 2005.

[15] G. Doretto and S. Soatto. Dynamic shape and appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):2006–2019, December 2006.

[16] G. Doretto and S. Soatto. *The Handbook of Texture Analysis*, chapter From Dynamic Texture to Dynamic Shape and Appearance Models: An Overview. Majid Mirmehdi (Ed.), World Scientific, 2007.

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

[18] A. Duci, A. Yezzi, S. Soatto, and K. Rocha. Harmonic embedding for linear shape analysis with application to eulerian thickness estimation. *Journal of Mathematical Imaging and Vision*, 25(3):341–352, October 2006.

[19] P. Favaro and S. Soatto. *3-D Shape Estimation and Image Restoration: Exploiting Defocus and Motion Blur.* Springer Verlag, December 2006.

[20] P. Favaro, S. Soatto, M. Burger, and S. Osher. Shape from defocus via diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007.

[21] D. Fontanelli, L. Ricciato, and S. Soatto. A fast ransac-based registration algorithm for accurate localization in unknown environments using lidar measurements. In *CASE*, 2007.

[22] B. Fulkerson, A. Vedaldi, and S. Soatto. Localizing objects with smart dictionaries. In *Proc. of the Eur. Conf. on Comp. Vis. (ECCV)*, October 2008.

[23] B.-W. Hong, K. Ni, S. Soatto, and T. Chan. Unsupervised multiphase segmentation: a recursive approach. *Computer Vision and Image Understanding*, Accepted 2009.

[24] B.-W. Hong, S. Soatto, K. Ni, and T. F. Chan. The scale of a texture and its application to segmentation. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2008.

[25] B.-W. Hong, S. Soatto, and L. Vese. Enforcing local context into shape statistics. *Advances in Computational Mathematics*, Special issue on Mathematical Imaging 2008.

[26] J. Jackson, A. J. Yezzi, and S. Soatto. Joint priors for variational shape and appearance modeling. In *Beyond Multiple View Geometry, CVPR Workshop*, June 2007.

[27] J. Jackson, A. J. Yezzi, and S. Soatto. Dynamic shape and appearance modeling via moving and deforming layers. *Intl. J. of Comp. Vis.*, accepted 2008.

[28] H. Jin, A. Yezzi, and S. Soatto. Mumford-shah on the move: Region-based segmentation on deforming manifolds with application to the reconstruction of shape and appearance from multiple images. *Journal of Mathematical Imaging and Vision*, accepted 2007.

[29] E. S. Jones, A. Vedaldi, and S. Soatto. Inertial structure from motion and autocalibration. In *Workshop on Dynamical Vision*, October 2007.

[30] T. Ko, S. Soatto, and D. Estrin. Background subtraction on distributions. In *Proc. of the Eur. Conf. on Comp. Vis. (ECCV)*, October 2008.

[31] J.J. Koenderink and A.J. van Doorn. Representation of local geometry in the visual system. *Biol. Cybern.*, 55:367–375, 1987.

[32] H. Ling and S. Soatto. Proximity distribution kernels for geometric context in category recognition. In *Int. Conf. on Comp. Vision*, October 2007.

[33] Y. Lu, P. Favaro, and S. Soatto. Autocalibration and uncalibrated reconstruction of shape from defocus. In *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Recog.*, page June, 2007.

[34] S. Manay, D. Cremers, B.-W. Hong, A. Yezzi, and S. Soatto. Integral invariants for shape matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1602–1618, 2006.

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

[36] J. Meltzer and S. Soatto. Edge descriptors for robust wide-baseline correspondence. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2008.

[37] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure and Applied Mathematics*, 1989.

[38] E. Prados, N. Jindal, and S. Soatto. A mathematical study of the non-local shape from ambient shading problem. In *Proc. of the Intl. Conf. on Scale-Space and Variational Methods (SSVM), Voss – Norway*, June 2009.

[39] 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.

[40] M. Raptis, M. Bustreo, and S. Soatto. Time warping under dynamic constraints. In *Workshop on Dynamical Vision*, October 2007.

[41] M. Raptis, K. Wnuk, and S. Soatto. Flexible dictionaries for action classification. In *Proc. of the Workshop on Machine Learning for Visual Motion Analysis (MLVMA)*, October Marseille, France, 2008.

[42] S. Soatto. *Modeling, Estimation and Control*, chapter On the distance between non-stationary time series. A. Chiuso, A. Ferrante and S. Pinzoni (Eds.), Springer Verlag, 2007.

[43] S. Soatto and A. Bissacco. *Homeland security*, chapter Visual detection and classification of humans, their pose, and their motion. G. Franceschetti (Ed.), Artech House, 2008.

[44] S. Soatto and A. Yezzi. Deformotion: deforming motion, shape average and the joint segmentation and registration of images. In *Proc. of the Eur. Conf. on Computer Vision (ECCV)*, volume 3, pages 32–47, 2002.

[45] G. Sundaramoorthi, S. Soatto, and P. Petersen. On the set of 2-d graph surfaces modulo contrast and domain deformations. In preparation 2008.

[46] G. Unal, A. J. Yezzi, and S. Soatto. A variational approach to problems in calibration of multiple cameras. *IEEE Trans. Pattern Anal. Mach. Intell.*, accepted 2006.

[47] A. Vedaldi, G. Guidi, and S. Soatto. Moving forward in structure from motion. In *Proc. of the IEEE Conf. on Comp. Vis. and Patt. Recog.*, 2007.

[48] A. Vedaldi, G. Guidi, and S. Soatto. Joint data alignment up to (lossy) transformations. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2008.

[49] A. Vedaldi, H. Ling, and S. Soatto. *Recognition in Context*, chapter Knowing a good feature when you see it: ground truth and methodology to evaluate local features for recognition. R, Cipolla, G.-M. Farinella (Eds), 2008.

[50] A. Vedaldi and S. Soatto. Features for recognition: viewpoint invariance for non-planar scenes. In *Proc. of the Intl. Conf. of Comp. Vision*, pages 1474–1481, October 2005.

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

[52] A. Vedaldi and S. Soatto. A rate-distortion approach to joint pattern alignment. In *NIPS*, December 2006.

[53] A. Vedaldi and S. Soatto. Viewpoint induced deformation statistics and the design of viewpoint invariant features: singularities and occlusions. In *Eur. Conf. on Comp. Vision (ECCV)*, pages II–360–373, 2006.

[54] A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In *Proc. of the Eur. Conf. on Comp. Vis. (ECCV)*, October 2008.

[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, and S. Soatto. Application of hybrid system identification in computer vision. In *Proc. of the Eur. Control Conf.*, pages 27–34, 2007.

[57] 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.

[58] R. Vidal, Y. Ma, S. Soatto, and S. Sastry. Two-view multibody structure from motion. *Intl. J. of Computer Vision*, 68(1):7–25, 2006.

[59] 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.

[60] 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*, pages 171–178, June 2006.

[61] K. Wnuk and S. Soatto. Filtering internet image search results towards keyword based category recognition. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2008.

[62] Z. Yi and S. Soatto. Correspondence transfer for the registration of multimodal images. In *Int. Conf. on Comp. Vision*, October 2007.

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).
Byung-Woo Hong, Postdoc, University of California, Los Angeles (partial support.
Gianfranco Doretto, Graduate Student, University of California, Los Angeles (partial support).

Publications

See references above.

Honors & Awards Received

- Associate Editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence (2003–2007).
- Member of the Editorial Board, International Journal of Computer Vision (current).
- Program Co-Chair, IEEE Intl. Conf. on Comp. Vision, 2011 (Barcelona).
- Area Chair, IEEE Intl. Conf. on Comp. Vision, 2009 (Kobe)

Transitions

Dynamic Textures technology transitioned to security and monitoring systems at GE (G. Doretto, GE Global Research); motion estimation techniques employed in 3-D dental scanner (D. Durbin, IOS-3D, INC.).

Patents

None during the period covered by this grant.

AFRL Point of Contact

Prof. William M. McEneaney, Program Manager, Dynamics and Controls AFOSR/NL, Tel. (703) 696-7796, Fax. (703) 696-8449, email william.mceneaney@afosr.af.mil.