Sequential Bayesian Non-Rigid Structure from Motion

Antonio Agudo
Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, Spain

Abstract

This thesis[1] *addresses the problem of recovering simultaneously camera motion and the 3D reconstruction of deformable objects from monocular video. We propose several methods to solve this problem in a sequential fashion, frame-by-frame estimation, as the data arrives. Deformable structures appear constantly in our everyday life, from human non-rigid motion (*e.g., *a smiling face or performing different expressions) to general objects such as flags, clothes, sails, banners, etc. More speculatively, in the medical field such as a beating heart or a bending abdomen, where the problem is particularly challenging. Our research seeks a physics-based method to perform 3D shape recovery in a wide variety of objects with different types of deformation from inextensibility to extensibility, without having to rely on learning data. In addition, our methods can perform also under realistic real-world assumptions allowing large amounts of missing data and measurement noise, they can run in real time at frame rate and can be used from sparse to dense shapes even for strong deformations. This dissertation presents our contributions in the field of deformable shape and camera motion recovery from a sequence of monocular images. In more detail, we present a novel algorithm where both motion and deformation are ruled by physical dynamic models. An important advantage of this method is that it does not require prior knowledge over material properties since they can be factorized out. We also present a generic estimation framework, eliminating the need of rigid priors, which is normally necessary when physics-based models are used. Finally, we show how the sequential estimation is possible for dense shapes, combining low-rank shape models with temporal and spatial smoothness priors. One of the main advantages of our models is the ability to include physical priors, if they are available. In contrast, we show how to solve the problem when this knowledge fails.*

[1] In this document we present an abstract of the PhD titled *Sequential Bayesian Non-Rigid Structure from Motion*, which was presented in the University of Zaragoza (2015).

1. Introduction

The combined inference of 3D scene structure and camera motion from monocular image sequences, or rigid Structure *f*rom Motion (S*f*M), is one of the most active areas in computer vision with applications in many domains. In the last decade, a great variety of methods have been proposed to simultaneously recover the reconstruction of a 3D object and the camera motion from video sequences. For this purpose, it is normally necessary a small collection of images acquired by cameras from different viewpoints, or by a single moving camera. In this work, we are interested in this last case, where the sole input is the camera image sequence gathered by a monocular camera.

In order to solve the S*f*M problem, the proposed methods have used the fundamental assumption that the observed scene is rigid. This rigidity prior has proved to be a powerful constraint to solve the problem, allowing practical and robust solutions. While S*f*M is now considered to be a mature field, these methods cannot be applied to structures undergoing *non-rigid deformations*. For these cases, recovering 3D structure of a deformable object and camera motion from a *monocular image sequence* is an ill-posed problem since many different 3D shapes can have the same image measurements producing severe ambiguities. This problem is known as Non-Rigid Structure *f*rom Motion (NRS*f*M). In addition to the inherent ambiguities, artifacts in the real measurements such as noise and missing data make the task even more challenging. Solving this problem is primarily motivated by the sport and movie industry, augmented reality applications and more speculatively, in medical imaging.

In this thesis, we assume an additional condition and our research focuses specifically on recovering both camera motion and non-rigid shape from monocular video in a *sequential fashion*. While our methods are available to recover the 3D reconstruction of a generic deformable object, ranging from inextensibility to extensibility deformations, additionally it can be used to on-line and real-time performance applications. Note that the on-line estimation based only on the measurements up to the current frame remains a challenging problem, especially when no training data is available.

2. Rigid Structure from Motion

In this section, we first present the S*f*M problem for the rigid case. S*f*M can be defined as the problem of recovering simultaneously the camera motion and 3D geometry of the scene from a monocular video sequence acquired by a single camera that can be calibrated or uncalibrated. For this purpose, the sole input information are the 2D image measurements of a set of points observed in the image plane. Since this problem is ill-posed for the projection stage, proposed S*f*M methods have constrained the problem assuming that the camera observes a rigid scene (see rigid case in Fig. 1). The rigidity prior is enough to make the problem well posed. With this assumption, when two or more views of a scene are available, the 3D geometry can be recovered via triangulation [33]. Additionally, this problem was also extended to uncalibrated cameras, where the calibration must be also be inferred [26].

One of the most influential works in rigid S*f*M was the rigid factorization proposed in [37, 61]. This method needs at least three images of a rigid object that has been observed by a moving camera. An additional assumption is that the input images have been acquired using an orthographic camera, a model that simplifies greatly the projection equation becomes linear and without the need of internal calibration parameters. Note that this camera model is an approximation of the more realistic perspective camera model, but being valid when the range of depths of the points in the structure is small compared to their distance to the camera. Later, the factorization method was extended to multiple independently moving objects [22] and also to the projective camera case [57].

In a similar way, the Simultaneous Localization And Mapping (SLAM) problem has emerged in mobile robotics using several sensors such as lasers or vision cameras. In this problem, given a mobile sensor moving along an unknown trajectory in an unknown environment, SLAM is able to simultaneously estimate both the 3D geometry of the environment –denoted as the map– and the sensor location. Only recently, vision cameras have been massively used by the robotic community as the main SLAM sensor. In this case, the monocular SLAM problem is particularly challenging since the sole input information are the 2D image projections of a 3D structure at frame rate. While the motivation of S*f*M and SLAM has historically been very different, both problems are roughly the same, although in the classical SLAM problem the estimation of the moving robot is normally in real-time since it is continuously observes and maps its unknown environment. Two methodologies have been proposed to solve the SLAM problem, using filtering techniques such as the Extended Kalman Filter (EKF) [20, 23, 24] and optimization techniques such as Bundle Adjustment (BA). In the first one, the information from the current frame is integrated into a multidimensional probability distribution that summarizes the information gathered for all previous frames along the sequence. In the second one, batch optimization over selected frames, using a sliding window [46] or spatially distributed keyframes [36, 56], from the live stream is performed. Again, this problem was also extended to uncalibrated cameras, where the calibration must be also inferred [19].

In recent years, real-time solutions based on S*f*M and SLAM have made significant progress solving the problem for a sparse set of salient points [36, 41] and even providing per-pixel dense reconstructions from video sequences acquired with a hand-held camera [43, 45] or with a micro aerial vehicle [65]. With the advent of new cheap RGB-D sensors that provides simultaneously depth and image intensity data, S*f*M and SLAM techniques have been also adapted to use such data [44]. Other successful examples are large scale reconstruction, where a large database of images available on Internet is used to recover very large reconstructions such as the Coliseum in Rome or the Notre Dame cathedral in Paris [2].

While rigid reconstruction from monocular video is now a well understood problem with many applications in a wide variety of areas such as robotics and movie industry, these methods cannot be applied to structures undergoing non-rigid deformations. For these cases, recovering 3D structure of a deformable object from a monocular image sequence is an ill-posed problem since many different 3D shapes can have the same image measurements producing severe ambiguities.

3. Non-Rigid Structure from Motion

For the non-rigid case, the shape of an object changes over time and the gathered images by the camera are different every time, as a result of a rigid motion of the camera and a non-rigid motion of the object (see non-rigid case in Fig. 1). This makes the problem equivalent to recover the 3D reconstruction from a single image, which is an ill-posed problem, since only one view per 3D configuration is available. As the object is non-rigid, many different shapes can have very similar image measurements (see Fig. 2) producing severe ambiguities.

The key insight to solve the problem is the assumption that objects do not arbitrarily deform their shape, since the deformations are produced by the effect of acting forces and according to their material properties. This observation has been exploited for many works in computer vision to constrain the possible range of solutions by adding prior information in order to make the problem well posed. This a priori knowledge includes constraints on both shape and camera motion.

The seminal work was proposed by [17], where the time-varying shape configuration is coded by means of a linear subspace of a set of deformation modes. This low-rank

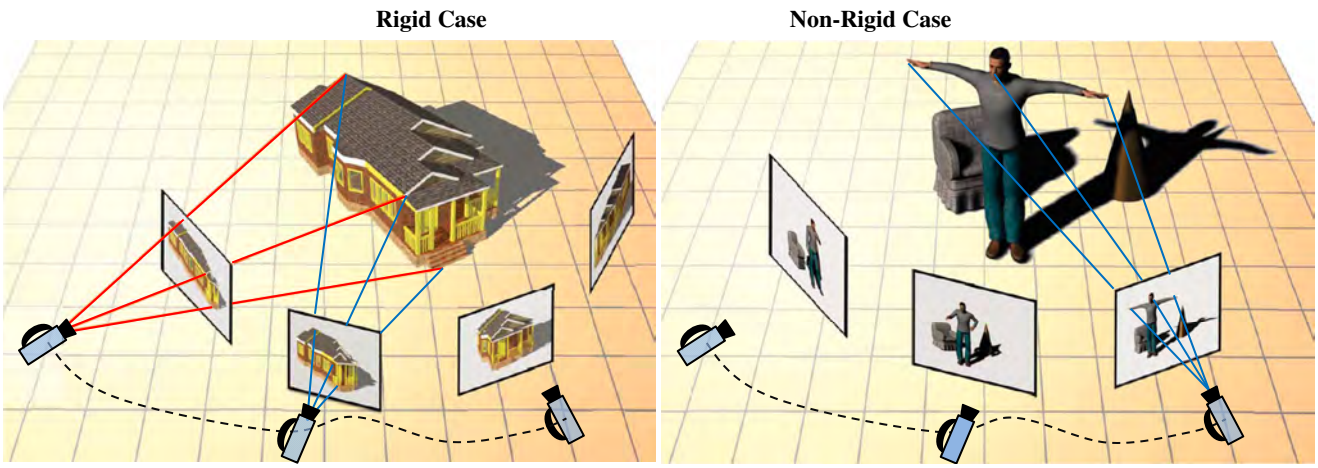

Figure 1. **Rigid and Non-Rigid Structure from Motion. Left:** A moving single camera observes a rigid object. When the object is rigid, the rigidity prior is enough to make the problem well posed. **Right:** A moving single camera observes a time-varying shape. In this case, the problem is ill-posed since it is equivalent to find a 3D geometry from single image.

shape model has proven successful in the 3D reconstruction of many real-world deformable objects, where both shape basis and the coefficients are unknown. A factorization-based algorithm was proposed to express the measurement matrix as a product of two factors, the motion matrix that includes the camera motion and the time-varying coefficients, and the shape matrix to encode the deformation modes. However, this problem has to enforce non-linear constraints on camera motion that make the estimation problem difficult. In order to exactly enforce the constraints, different optimization schemes have been proposed adding additional priors such as rigid component [25], temporal [1, 16, 62] and spatial smoothness [62], allowing them to be robust to missing data and noise.

While the low-rank shape prior has been widely used in NRS*f*M, this model can only code small linear deformations, since small values of rank on the space basis could be insufficient to represent the variation of real-world objects. Stronger deformations and non-linear patterns could require a relatively large number of rank on the shape basis, where the extra degrees of freedom will be unconstrained by the data and end up fitting noise. To solve these problems, other approaches have been proposed to target more complex non-rigid deformations [15, 28, 51, 52].

[14] proposed that the low-rank constraint can be applied in the temporal evolution of each 3D point in the space, instead of in the spatial configuration as shape basis. To do this, they independently code each 3D point position at each instant by means of a linear combination of trajectory basis. In fact, this model is just a dual representation of the low-rank shape basis model, with the same compaction power. However, its great advantage is that the trajectory basis can be pre-defined in advance, using basis representations for temporal signals such as the Discrete Cosine

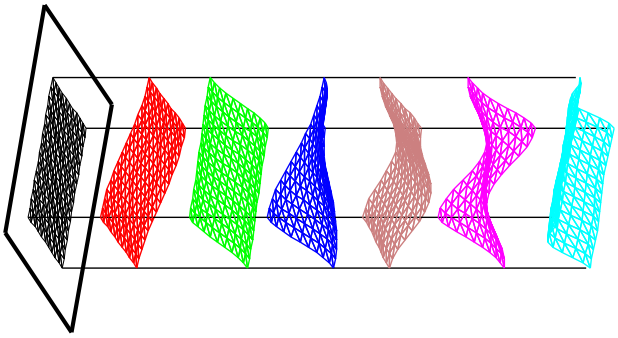

Figure 2. **Ambiguity for reconstructing deformable surfaces from monocular video.** We display the fact that there are many different 3D shape configurations that correspond to similar image observations.

Transform (DCT), reducing the number of parameters to estimate. Moreover, since each point is modeled independently, these methods can handle a wide range of motions, from non-rigid to articulated motions. A shared limitation with low-rank shape models is that they also need to specify the number of trajectories in the basis, being the reconstructions very sensitive to the choice of rank. The ambiguities of this model were analyzed by [48], observing that when the camera motion does not lie in the same subspace of the deformation motion, the ambiguities are reduced. This means that the deformable object has to be observed by a camera with strong motions, a need that reduces its applicability in real scenarios. More recently, [31] used the trajectory representation to impose temporal smoothness on each 3D point trajectory.

Alternative models, such as piecewise models, have been proposed to encode more accurately strong local deformations. These methods split the points into regions denoted as patches, that are modeled independently. The main dif-

ference between piecewise models, it is the model chosen for each path, which can be planar [18, 64], locally rigid [60] or quadratic [27]. A very important drawback of these methods [27, 64] is that they need a manual division of the surface into patches. This requirement was avoided in [52] where a formulation to automate the best division of the surface into local patches was proposed. Despite this progress, these methods rely on common features between reconstructed patches to enforce global consistency or [60] needs too many points to enforce the rigid local constraint that is difficult to hold in practice. Moreover, these methods require a post-precessing step in order to stitch all the local reconstructions into a single smooth surface, increasing their complexity.

More recently, following dense approaches to multi-view stereo [29] and variational techniques to perform real-time dense reconstructions of rigid scenes [45], NRS*f*M methods have been extended to the dense case. [30] propose the first variational approach to NRS*f*M to produce per-pixel dense vivid reconstructions, combining a low-rank shape model with local smoothness priors. [53] propose to estimate both segmentation and reconstruction for all feature points in every frame using piecewise models.

On the other hand, mechanical priors have also been proposed to constrain the deformations of non-rigid objects. Early approaches used deformable superquadrics [40], balloons [21] or spring meshes [35], although these approaches were only valid to code relatively small deformations. In [38, 39] were proposed more closely a Finite Element Method (FEM), modeling the surface as a thin-plate with acting forces, using as input data volume images, such as tomography. In a similar way, FEM models were proposed to accurately represent specific materials [63, 66] with known material properties. To tackle the high dimensionality of these physics-based models, a low-rank representation was proposed by applying modal analysis over a known structure discretized in 3D finite elements [42, 49]. This method was then applied to image segmentation [50], medical imaging [42] and deformable 2D motion tracking [55, 58], requiring again the material properties. Later, more accurate representations were achieved using non-linear FEM for large deformations of beam [34] and for 3D solid structures [32, 63]. However, their applicability was limited to very specific geometries for which the material properties were also known. While physic-based models have proved effective in computer vision, they were discarded for their high computational cost, the requirement of exact material properties and their high dimensionality of unknowns.

In spite of all this tremendous progress, NRS*f*M methods remain behind their rigid counterparts when it comes to real-time performance. The reason behind this is that they are typically limited to batch operation where all frames in the sequence are processed at once, after their acquisition, preventing them from on-line and real-time performance. Only recently, NRS*f*M has been extended to sequential processing [47, 59]. However, they remain slow or do not scale to the use of a large number of points. Furthermore, these methods do not compute the tracking and data association on-the-fly, that it is assumed known.

It is our aim contribution to propose several sequential algorithms to recover the camera motion and the time-varying shape from monocular video as the data arrives, which breaks free from the standard requirement of most of state-of-the-art methods of processing all frames at once. For all methods, we assume not to know any learning prior, and our formulations can be used for a wide range of deformations from inextensibility to extensibility. We use piecewise physic-based models (*elastic problem resolved by finite element*) with unknown material properties, that are able to handle strong deformations at low computational cost. In addition, we show how the non-rigid shape estimation is possible in real time, and propose a novel formulation suitable for the dense case.

4. Applications

Recovering simultaneously non-rigid 3D shape and camera motion have too many applications in many different domains ranging from several industries to medical imaging. Next, we discuss a few of them.

4.1. Movie industry: augmented reality

The movie industry has recently shown great interest in methods that allow recovering the 3D shape and motion of deformable objects, mainly for augmented reality applications. In these cases, a virtual object is inserted on the scene following the recovered camera motion (see Fig. 3). The camera location is especially crucial for the virtual object, since it has to follow a realistic trajectory in the final film. In a similar way, different motions, such as facial expressions, can be recovered reducing considerably the work of graphic artists to animate the virtual object. In both cases, the algorithms propose in this work could be used.

4.2. Sport industry: sailing

In this case, sailors are normally interested in measuring shape changes in their own sails, or even studying the sails of their opponents. Note that these measurements can be used to control their sail boats in real time, adjusting the shape of the sail to get more speed and even helping to improve the design. While these measurements are normally made using classical sensors, such as strain gages, they require a fine calibration and are necessary too many sensors to measure all deformation shape. In this context, vision sensors overcome other contact sensors that normally change the behavior of the sail shape. In Fig. 4 we show as

Figure 3. **Motion capture system applied to films. Left:** Several selected frames used to recover facial expressions by means of artificial markers. **Right:** A multi-camera system is used to track the motion of the face or object. The recovered motion is augmented with virtual objects performing a realistic trajectory. Images copyright Weta Digital.

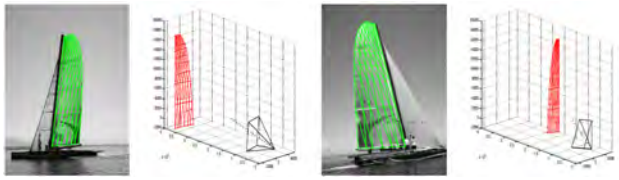

Figure 4. **Non-Rigid shape recovery in sailing.** Another application of our methods is to simultaneously reconstruct a sail and recover the camera pose from monocular video. Image from [54].

similar methods to ours have been proposed to recover the shape of a sail.

4.3. Experimental industry

Another potential application is the on-line characterization of materials using image-based measurements. Again, it is necessary to measure the shape changes, *i.e.*, the 3D displacements, to finally recover the material properties of the object using inverse analysis (see Fig. 5). This problem is particularly challenging when the characterization has to be performed *in-vivo*, being necessary an estimation in real time and using natural landmarks.

In addition, these measurements can be also used to analyze the behavior of several components, such as a plane wing, that deforms during a flight. In this case, it is necessary to compare the predicted values in the design stage with the observed values, helping to improve the simulation software.

4.4. Marketing industry

Similar to movie industry, entertainment and marketing industry is normally interested in reproducing cloth deformations, such as jeans or dresses. These methods could be used to recover the 3D deformations of real clothes in real time and to use the reconstructed shapes in video games, animation movies or publicity.

In a similar way, our methods could be used in intelligence gathering, that normally requires an automated reading of banners or t-shirts, being necessary unwarping the surfaces. This is especially relevant where an estimation in real time is mandatory, for instance, to rebroadcast news at street level.

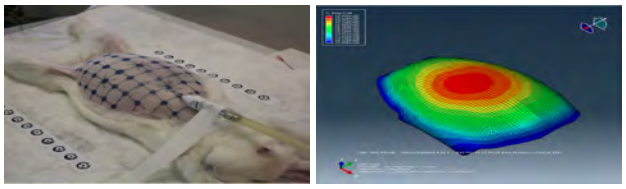

Figure 5. **Non-Rigid shape recovery in biomechanical experimentation.** In this application, the deformable shape has to be recovered to use the experimental displacement field into inverse analysis. To do this, a 3D model of the object with unknown material properties is fitted comparing the predicted response with the observed values. **Left:** Animal surface with artificial markers. **Right:** 3D model.

4.5. Medical Imaging

Finally, these methods can be applied in the medical field, where the current trend is to make surgery more automatic and less invasive. There are several medical scenarios where only a monocular camera is available to analyze the deformable tissues such as laparoscopy, gastroscopy, colonoscopy and bronchoscopy (see some examples in Fig. 6). In this context, the task for simultaneously recovering 3D shape and camera motion is particularly challenging since the resulting images are normally of poor quality and the observed deformations are very large. In addition, for real interventions the estimation has to be provided on-line and in real time. The proposed methods in this work can be used to have a full 3D configuration of the observed tissue from the images or even getting augmented views, making the surgeons' work much easier. Additionally, the camera trajectory recovery allows us the automation of tasks, such as the control of tools.

5. Contributions

In this work, we push monocular NRS*f*M and non-rigid SLAM forward towards real-time operation by proposing several new on-line algorithms to recover the 3D non-rigid shape of strongly deforming surfaces and camera motion under realistic real-word conditions. For this purpose, we exploit physics-based models for both the camera motion and the time-varying shape without the need for a pre-trained model, allowing us to apply our methods even on real videos where this data is not available. While physics-based models have been discarded for their high computational complexity, the requirement of exact physical material properties and high dimensionality of unknowns. In this work, we show as these methods can be used when the material properties are unknown, as they can be applied in

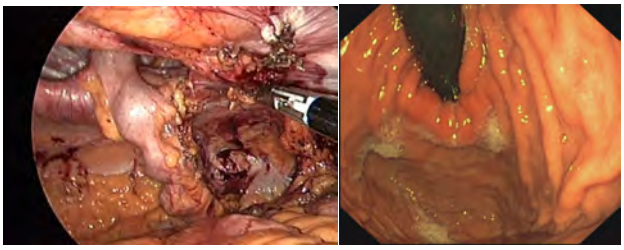

Figure 6. **Medical imaging application.** Our algorithms can be applied to assist surgeons by providing them reconstructions in real time at frame rate to measure and obtain augmented views during the surgery. **Left:** Laparoscopy image. **Right:** Gastroscopy image.

on-line and real-time applications and finally, as they can be used to recover dense objects in a sequential manner.

In detail, the main contributions of this work are:

- We contribute an algorithm that extends the classical rigid SLAM to the non-rigid domain. For this purpose, we represent the object's surface mechanics by means of Navier's equations, which are solved using the finite element method. In addition, most of material properties can be factorized out and do not have to be known in advance, avoiding the needed for a strong knowledge of the mechanic model. Our method can combine both rigid and non-rigid points under a unique formalism. With this approach, we simultaneously recover the full camera trajectory and the deformable shape over time just from the sole input of the image sequence gathered by the camera. While most of state-of-the-art methods use the 2D tracking data as input, our method automatically establishes correspondences between consecutive frames, solving the data association on-the-fly. One of the main advantages of this approach is the ability to handle from isometric to extensible deformations just by tunning, without additional constraints. These results have been published in [4, 13].

- We present a novel rank analysis of the FEM system to avoid imposing the boundary conditions of the FEM system. To achieve this, we propose to approximate the compliance matrix by means of a generalization of the inverse stiffness matrix, enforcing a six rank deficiency that corresponds to the six rigid body motions of an object in the 3D space. In addition, we present a 3D FEM formulation that provides better conditioned matrices and reduces the computational cost since the resulting linear system has a lower dimension. With these ingredients, to the best of our knowledge we present the first approach to simultaneously estimate both camera pose and the 3D reconstruction of deformable objects from monocular images in real time at frame rate without requiring any rigid prior, as we have experimentally demonstrated. These results have been published in [5, 6, 12].

- We propose to reduce the high dimensionality of unknowns in physics-based models by means of a linear subspace of mode shapes that encodes the modes of deformation. To do this, the force balance equation is solved using modal analysis via a simple eigenvalue problem. We incorporate a new classification of mode shapes obtaining three practical mode families: rigid, bending and stretching deformations, instead of the two proposed by state-of-the-art methods. With this new classification differentiating the type of the deformation, we can efficiently obtain mode shapes with bending deformations even for very dense shapes. However, when stretching deformations are needed, the standard solution for the eigenvalue problem can become prohibitive in terms of computational and memory requirements. For these cases, we propose two methods for efficiently obtaining the mode basis, especially for stretching modes; the first one is a frequency-based method, and the second one is a coarse-to-fine approach. Both methods drastically reduce the computational cost remaining a good quality of the solution. These results have been published in [3, 7, 8].

- We propose two sequential algorithms for very different scenarios ranging from sparse to dense objects. To do this, we employ both temporal and spatial smoothness priors using sequential bundle adjustment and expectation maximization algorithms over a sliding window of image frames to optimize the model parameters. The non-rigid shape is modeled as a linear combination of mode shapes obtained by modal analysis, with time-varying weights that define the shape at each frame and are estimated on-the-fly. Our systems exhibit a good trade-off between accuracy and computational budget, and they can work under realistic real-world assumptions such as dealing with structured occlusions and handling non-isometric deformations. As both methods estimate a small number of parameters per frame, they could potentially achieve real-time performance at frame rate. These results have been published in [3, 7, 8].

6. Conclusion

Taken together, the proposed methods in this work push monocular NRS*f*M and non-rigid SLAM forward towards real-time operation. We believe that our methods represent a significant step towards the challenge of real-time estimation of non-rigid objects from monocular video, when the measurements are a sparse set of points or even per pixel.

In the remainder of this section, we discuss their limitations and propose research directions to improve them.

There are various ways in which our methods could be improved. We briefly cite the most interesting extensions.

One way of extending our non-rigid monocular SLAM system in [4, 13] is to explore the use of new feature descriptors to establish correspondences between frames. At the moment, we use cross correlation, that has proven to be sufficient for many deformations in different objects, including the challenging laparoscopic sequence. However, when the elasticity of the object is very high, the current patch of a deformed feature can be very different with respect to the original template. Since a potential application of this work is to process medical images, such as bronchoscopy, we believe that the use the new feature descriptors is key to establish correspondences and avoid that many deformed points are annotated as outliers.

Regarding [5, 6, 12], while we have experimentally demonstrated our performance in real time at frame rate for small maps, around forty points, this method could be extended to handle bigger maps. To do this, we propose to re-think the compliance matrix computation. Since the deformation is smooth, we could obtain it every several frames instead of re-computing every frame. Another option is directly updated the compliance matrix, using the previous matrix and the current stiffness matrix. In addition, since some points disappear out of view and are annotated as missing data, other new points appear and should be tracked. Hence, successfully incorporating the new features into an existing model is another potential direction of this research.

Finally, the modal space based method that we have presented in [3, 7, 8] can be extended to the articulated case and to multiple objects. To this end, new mechanical matrices are necessary to code the articulated motion, mainly for human body estimation. In this case, a combination of different deformable, articulated and rigid parts could be considered. In addition to this, our shape basis could be updated on-the-fly according as new measurements are available.

Some of these problems have been explored in [9, 10, 11].

Acknowledgment: The work of this thesis has been developed together with the groups: Robotics, Perception and Real Time (RoPeRT); and Applied Mechanics and Bioengineering (AMB) of the University of Zaragoza. The work has been partially supported by the Spanish Ministry of Science and Innovation under projects: 1) Robust 3D real-time vision. Application to augmented reality in endoscopic surgery (DPI2009-07130), 2) SVMap: Semantic Visual Mapping for Rigid and Non-Rigid Scenes (DIP2012-32168), 3) Modelado biomecánico del tejido musculoesquelético abdominal (DPI2011-15551-E); 4) the European project POPCORN: System based on 3D plenoptic imaging for dynamic topographical corneal characterization (SME-2013 066634); and by a scholarship FPU12/04886 of the Spanish MECD. Some results have been product of the several collaborations and visits to the following laboratories: Vision Group at the Queen Mary University of London (QMUL), Vision and Imaging Science at the University College London (UCL), Laboratory of Mathematics in Imaging at the Harvard University (HU) in Boston, and the Institut de Robòtica i Informàtica Industrial (CSIC-UPC) in Barcelona.

References

[1] H. Aanæs and F. Kahl. Estimation of deformable structure and motion. In *Workshop on Vision and Modelling of Dynamic Scenes*, 2002.

[2] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building Rome in a day. In *IEEE International Conference on Computer Vision*, pages 72–79, 2009.

[3] A. Agudo, L. Agapito, B. Calvo, and J. M. M. Montiel. Good vibrations: A modal analysis approach for sequential non-rigid structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1558–1565, 2014.

[4] A. Agudo, B. Calvo, and J. M. M. Montiel. FEM models to code non-rigid EKF monocular SLAM. In *IEEE International Workshop on Dynamic Shape Capture and Analysis*, pages 1586–1593, 2011.

[5] A. Agudo, B. Calvo, and J. M. M. Montiel. 3D reconstruction of non-rigid surfaces in real-time using wedge elements. In *Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, pages 113–122, 2012.

[6] A. Agudo, B. Calvo, and J. M. M. Montiel. Finite element based sequential bayesian non-rigid structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1418–1425, 2012.

[7] A. Agudo, J. M. M. Montiel, L. Agapito, and B. Calvo. Online dense non-rigid 3D shape and camera motion recovery. In *British Machine Vision Conference*, 2014.

[8] A. Agudo, J. M. M. Montiel, L. Agapito, and B. Calvo. Modal space: A physics-based model for sequential estimation of time-varying shape from monocular video. *Journal of Mathematical Imaging and Vision, to appear*, 2016.

[9] A. Agudo, J. M. M. Montiel, B. Calvo, and F. Moreno-Noguer. Mode-shape interpretation: Re-thinking modal space for recovering deformable shapes. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1–8, 2016.

[10] A. Agudo and F. Moreno-Noguer. Learning shape, motion and elastic models in force space. In *IEEE International Conference on Computer Vision*, pages 756–764, 2015.

[11] A. Agudo and F. Moreno-Noguer. Simultaneous pose and non-rigid shape with particle dynamics. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2179–2187, 2015.

[12] A. Agudo, F. Moreno-Noguer, B. Calvo, and J. M. M. Montiel. Real-time 3D reconstruction of non-rigid shapes from

single moving camera. *Computer Vision and Image Understanding, to appear*, 2016.

[13] A. Agudo, F. Moreno-Noguer, B. Calvo, and J. M. M. Montiel. Sequential non-rigid structure from motion using physical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(5):979–994, 2016.

[14] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Non-rigid structure from motion in trajectory space. In *Neural Information Processing Systems*, pages 41–48, 2008.

[15] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Trajectory space: A dual representation for nonrigid structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1442–1456, 2011.

[16] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd. Coarse-to-fine low-rank structure-from-motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[17] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 690–696, 2000.

[18] A. Chhatkuli, D. Pizarro, and A. Bartoli. Non-rigid shape-from-motion for isometric surfaces using infinitesimal planarity. In *British Machine Vision Conference*, 2014.

[19] J. Civera, D. R. Bueno, A. J. Davison, and J. M. M. Montiel. Camera self-calibration for sequential bayesian structure from motion. In *IEEE International Conference on Robotics and Automation*, pages 403–408, 2009.

[20] J. Civera, A. J. Davison, and J. M. M. Montiel. Inverse depth parametrization for monocular SLAM. *TRO*, 24(5):932–945, 2008.

[21] L. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2D and 3D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, 1993.

[22] J. Costeira and T. Kanade. A multibody factorization method for independent moving objects. *International Journal on Computer Vision*, 29(3):159–179, 1998.

[23] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *IEEE International Conference on Computer Vision*, pages 1403–1410, 2003.

[24] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, 2007.

[25] A. Del Bue, X. Llado, and L. Agapito. Non-rigid metric shape and motion recovery from uncalibrated images using priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1191–1198, 2006.

[26] O. D. Faugeras, Q. Luong, and S. Maybank. Camera self-calibration: Theory and experiments. In *European Conference on Computer Vision*, pages 321–334, 1992.

[27] J. Fayad, L. Agapito, and A. Del Bue. Piecewise quadratic reconstruction of non-rigid surfaces from monocular sequences. In *European Conference on Computer Vision*, pages 297–310, 2010.

[28] J. Fayad, A. Del Bue, L. Agapito, and P. M. Q. Aguiar. Non-rigid structure from motion using quadratic deformation models. In *British Machine Vision Conference*, 2009.

[29] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010.

[30] R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1272–1279, 2013.

[31] P. F. U. Gotardo and A. M. Martinez. Non-rigid structure from motion with complementary rank-3 spaces. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3065 – 3072, 2011.

[32] N. Haouchine, J. Dequidt, M. O. Berger, and S. Cotin. Single view augmentation of 3D elastic objects. In *International Symposium on Mixed and Augmented Reality*, pages 229–236, 2014.

[33] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. University Press, Cambridge, 2004.

[34] S. Ilić and P. Fua. Non-linear beam model for tracking large deformation. In *IEEE International Conference on Computer Vision*, pages 1–8, 2007.

[35] Y. Kita. Elastic-model driven analysis of several views of a deformable cylindrical object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1150–1162, 1996.

[36] G. Klein and D. W. Murray. Parallel tracking and mapping for small AR workspaces. In *International Symposium on Mixed and Augmented Reality*, pages 225–234, 2007.

[37] L. Kontsevich, M. Kontsevich, and A. Shen. Two algorithms for reconstructing shapes. *Optoelectronics, Instrumentation and Data Processing*, 5:76–81, 1987.

[38] T. McInerney and D. Terzopoulos. A finite element model for 3D shape recognition and nonrigid motion tracking. In *IEEE International Conference on Computer Vision*, pages 518–523, 1993.

[39] T. McInerney and D. Terzopoulos. A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4D image analysis. *Computational Medical Imaging and Graphics*, 19(1):69–83, 1995.

[40] D. Metaxas and D. Terzopoulos. Constrained deformable superquadrics and nonrigid motion tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):580–591, 1993.

[41] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd. Generic and real-time structure from motion using local bundle adjustment. *Image Vision Computing*, 27(8):1178–1193, 2009.

[42] C. Nastar and N. Ayache. Fast segmentation, tracking and analysis of deformable objects. In *IEEE International Conference on Computer Vision*, pages 275–279, 1993.

[43] R. Newcome and A. J. Davison. Live dense reconstruction with a single moving camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1498–1505, 2010.

[44] R. Newcome, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *International Symposium on Mixed and Augmented Reality*, pages 127–136, 2011.

[45] R. Newcome, S. Lovegrove, and A. J. Davison. DTAM: Dense tracking and mapping in real-time. In *IEEE International Conference on Computer Vision*, pages 2320–2327, 2011.

[46] D. Nister, O. Naroditsky, and J. Bergen. Visual odometry. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–659, 2004.

[47] M. Paladini, A. Bartoli, and L. Agapito. Sequential non rigid structure from motion with the 3D implicit low rank shape model. In *European Conference on Computer Vision*, pages 15–28, 2010.

[48] H. S. Park, T. Shiratori, I. Matthews, and Y. Sheikh. 3D reconstruction of a moving point from a series of 2D projections. In *European Conference on Computer Vision*, pages 158–171, 2010.

[49] A. Pentland and B. Horowitz. Recovery of nonrigid motion and structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):730–742, 1991.

[50] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(21):715–729, 1991.

[51] V. Rabaud and S. Belongie. Re-thinking non-rigid structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[52] C. Russell, J. Fayad, and L. Agapito. Energy based multiple model fitting for non-rigid structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3009–3016, 2011.

[53] C. Russell, R. Yu, and L. Agapito. Video pop-up: Monocular 3D reconstruction of dynamic scenes. In *European Conference on Computer Vision*, pages 583–598, 2014.

[54] J. Sánchez-Riera, J. Ostlund, P. Fua, and F. Moreno-Noguer. Simultaneous pose, correspondence and non-rigid shape. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1189–1196, 2010.

[55] S. Sclaroff and A. Pentland. Modal matching for correspondence and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(6):545–561, 1995.

[56] H. Strasdat, J. M. M. Montiel, and A. J. Davison. Scale drift-aware large scale monocular SLAM. In *Robotics: Science and Systems*, 2010.

[57] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure from motion. In *European Conference on Computer Vision*, pages 709–720, 1996.

[58] H. Tao and T. S. Huang. Connected vibrations: A modal analysis approach for non-rigid motion tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 735–740, 1998.

[59] L. Tao, B. J. Matuszewski, and S. J. Mein. Non-rigid structure from motino with incremental shape prior. In *IEEE International Conference on Image Processing*, pages 1753–1756, 2012.

[60] J. Taylor, A. D. Jepson, and K. N. Kutulakos. Non-rigid structure from locally-rigid motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2761–2768, 2010.

[61] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization approach. *International Journal on Computer Vision*, 9(2):137–154, 1992.

[62] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):878–892, 2008.

[63] L. V. Tsap, D. B. Goldof, and S. Sarkar. Nonrigid motion analysis based on dynamic refinement of finite element models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):526–543, 2000.

[64] A. Varol, M. Salzmann, E. Tola, and P. Fua. Template-free monocular reconstruction of deformable surfaces. In *IEEE International Conference on Computer Vision*, pages 1811–1818, 2009.

[65] A. Wendel, M. Maurer, G. Graber, T. Pock, and H. Bischof. Dense reconstruction on-the-fly. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1450–1457, 2012.

[66] A. Young and L. Axel. Non-rigid wall motion using MR tagging. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 399–404, 1993.