

3D Motion Reconstruction for Real-World Camera Motion

Yingying Zhu^{1,2}, Mark Cox² and Simon Lucey²

¹University of Queensland, Australia

²Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

zy.2@itee.uq.edu.au, Mark.Cox@csiro.au, Simon.Lucey@csiro.au,

Abstract

This paper addresses the problem of 3D motion reconstruction from a series of 2D projections under low reconstructibility. Reconstructibility defines the accuracy of a 3D reconstruction from 2D projections given a particular trajectory basis, 3D point trajectory, and 3D camera center trajectory. Reconstructibility accuracy is inherently related to the correlation between point and camera trajectories. Poor correlation leads to good reconstruction, high correlation leads to poor reconstruction. Unfortunately, in most real-world situations involving non-rigid objects (e.g. bodies), camera and point motions are highly correlated (i.e., slow and smooth) resulting in poor reconstructibility. In this paper, we propose a novel approach for 3D motion reconstruction of non-rigid body motion in the presence of real-world camera motion. Specifically we: (i) propose the inclusion of a small number of keyframes in the video sequence from which 3D coordinates are inferred/estimated to circumvent ambiguities between point and camera motion, and (ii) employ a L_1 penalty term to enforce a sparsity constraint on the trajectory basis coefficients so as to ensure our reconstructions are consistent with the natural compressibility of human motion. We demonstrate impressive 3D motion reconstruction for 2D projection sequences with hitherto low reconstructibility.

1. Introduction

Recent work in [2, 6] has demonstrated the benefit of employing a trajectory basis when solving certain non-rigid structure from motion problems. Non-rigid structure from motion (NRSFM) refers to the task of recovering the time varying 3D coordinates of points on a deforming object from their 2D locations in an image sequence. Trajectory basis approaches to NRSFM are especially useful for non-rigid objects whose 3D coordinates motion are compressible in the trajectory domain, as it places natural constraints on the object's motion (e.g., periodic, slow and smooth, etc.). The term compressible is borrowed from the com-

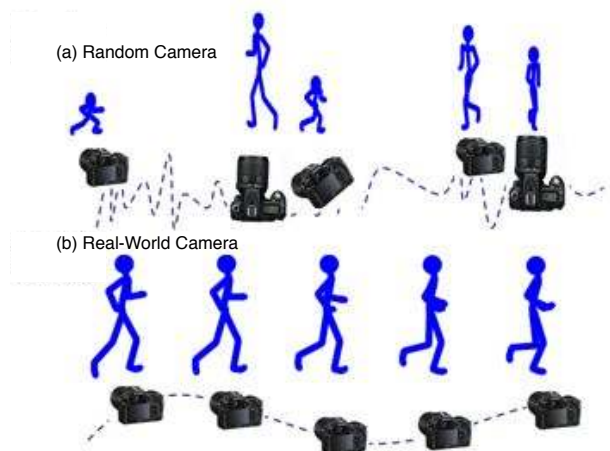


Figure 1. 2D projections for different camera trajectories. (a) 2D projections for a random camera, (b) 2D projections for a real-world camera that is moving slowly/smoothly. It has been well noted in literature [6] that for trajectory basis NRSFM methods (a) will have good 3D motion reconstruction whereas (b) will have poor motion reconstruction. In this paper we propose an approach that incorporates randomly sampled 3D keyframes within a trajectory basis framework in order to gain good 3D motion reconstruction for (b).

pressed sensing community [5] to denote a signal that is well approximated by a sparse vector.

The Problem: Park et al. [6] recently characterized theoretically the accuracy of a 3D reconstruction from 2D projections given a particular trajectory basis, 3D point trajectory, and 3D camera center trajectory. Reconstructibility accuracy is inherently related to the correlation between point and camera trajectories. Poor correlation leads to good reconstruction, high correlation leads to poor reconstruction. Impressive empirical results have been depicted in [2, 6] for NRSFM trajectory basis problems where: (i) camera motion is continuous but much faster than the object's, and (ii) the camera motion is random (see top row Figure 1). Unfortunately, in most real-world situations (see bottom row

Figure 1) involving non-rigid objects (e.g. bodies), camera and point motions are highly correlated (e.g., slow and smooth) resulting in poor reconstructibility in many practical scenarios.

Contributions: Recent work [11, 13] has demonstrated how static NRSFM problems for the human body can be overcome through the application of rigid-body constraints on the torso & hips along with assumptions about joint length symmetry. Although giving impressive results on real-world images the approach is extremely labour intensive as: (i) it requires human labeling to obtain 2D joint locations, and (ii) further human labeling to resolve depth ambiguities. This makes the approach of [11, 13] largely unsuitable for NRSFM problems involving 2D video sequences of points containing thousands of image frames.

In this paper we propose a novel approach for overcoming the current limitations to trajectory basis NRSFM methods [2, 6] when applied to the human body given that we have an extremely small number of 3D keyframes generated through [11, 13]. An elegant aspect of our proposed approach is that it synergetically combines the shortcomings of [2, 6] and [11, 13] resulting in an algorithm that is able to obtain 3D motion reconstruction of 2D image sequences with an extremely small number of hand labeled keyframes. The main contributions of this paper are,

- Investigate the reconstructibility of 3D human motion from a small number of randomly chosen 3D keyframes. Based on the compressibility of human motion in the discrete cosine transform (DCT) trajectory basis we argue for, and empirically justify the employment of an L_1 penalty term. (Section 3)
- We then propose a joint objective function that combines full sequence 2D projections and randomly chosen 3D keyframes into a single framework using a trajectory basis. (Section 3)
- We demonstrate empirically how the inclusion of 3D keyframes with an L_1 penalty term on full sequence 2D projections gives a dramatic improvement in reconstructibility over Park et. al’s [6] method using real-world camera motion. (Section 4)

Related Work: Factorization approaches, first proposed for recovering rigid structure by Tomasi and Kanade in [8], were extended to handle non-rigidity in the seminal paper by Bregler et al. [4]. In this notable work the authors realized that the only way to make the non-rigid structure from motion problem tractable is to impose some constraint on the object being analyzed (e.g. assuming we are looking at a body). Bregler et al. imposed this constraint as a linear shape basis. A number of approaches that develop the use of shape basis have subsequently been proposed, including [9, 14, 10]. A fundamental criticism, however, of

all these approaches is the specificity, or more critically the poor generalization properties, of the shape basis. For example, the shape basis of a “person walking” will not be the same as that of a “person bending down to pick something up”. Akhter et al. [2] recently proposed an approach to non-rigid structure from motion that is shape agnostic, and instead places a constraint on how individual points are allowed to move through space over time. The authors frame this work in a manner similar to Bregler et al. through the introduction of a trajectory basis. This seminal work forms the basis of our proposed approach.

Recently, Wei and Chai [12] proposed an approach for obtaining 3D motion from 2D video sequences based on the employment of a small number of discontinuous 3D keyframes. Although similar in spirit, the method presented in this paper differs substantially to Wei and Chai’s in that we do not rely on a complex, non-linear, computationally costly physics based model for solving the 3D motion. Instead, our method relies on a trajectory basis for solving this missing motion information through the optimization of an elegant convex objective function.

2. Trajectory Basis NRSFM

We will quickly review trajectory basis NRSFM methods [2, 6]. Let a point at frame f in 3D $\mathbf{X}_f = [x_f, y_f, z_f]^T$ be imaged as $\mathbf{x}_f = \mathbf{P}_f \mathbf{X}_f$ where $\mathbf{P}_f \in \mathcal{R}^{2 \times 3}$ is a weak-perspective camera matrix. By taking into account all F time instants the 3D point trajectory \mathbf{X} can be formulated as,

$$\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_F \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1 & & \\ & \ddots & \\ & & \mathbf{P}_F \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_F \end{bmatrix}, \text{ or } \mathbf{x} = \mathbf{P}\mathbf{X} \quad (1)$$

where $\mathbf{x} \in \mathcal{R}^{2F}$, $\mathbf{X} \in \mathcal{R}^{3F}$ and $\mathbf{P} \in \mathcal{R}^{2F \times 3F}$. This 3D point trajectory can be approximated as,

$$\mathbf{X} \approx \sum_{j=1}^{3K} \Theta_j \beta_j, \text{ or } \mathbf{X} \approx \Theta \beta \quad (2)$$

where $\Theta_j \in \mathcal{R}^{3F}$ is the j th trajectory basis vector, $\Theta \in \mathcal{R}^{3F \times 3K}$ is the trajectory basis matrix and $\beta \in \mathcal{R}^{3K}$ is the trajectory coefficient vector. This trajectory basis attempts to model inherent redundancies/constraints on how real-world non-rigid objects move (i.e., motion in 3D needs to be smooth and continuous). The DCT basis has been reported [2, 6] to fulfil this role well for many types of non-rigid motion (e.g., human motion).

A key concept employed in all variations of trajectory basis NRSFM [2, 6] is that it is possible to solve,

$$\arg \min_{\beta} \|\mathbf{x} - \mathbf{P}\Theta\beta\|^2 \quad (3)$$

if $3K < 2F$. This typically takes the form of choosing the first K of F smooth DCT basis vectors for the x , y and z components. For the purposes of this paper we shall assume we know the concatenation of camera matrices \mathbf{P} a priori. More details on the estimation of \mathbf{P} shall be provided in Section 4. Given a solution to β it is then possible to obtain an approximation to the full 3D motion matrix \mathbf{X} through Equation 2.

Reconstructibility: Park et al. [6] pointed out recently a problem in Equation 3. Specifically, an adversary could choose a concatenated matrix \mathbf{P} of camera/projection matrices that would give poor reconstructions of β . Park et al. characterized¹ this situation theoretically by characterizing camera motion as $\mathbf{C} = \Theta_C \beta_C + \Theta_C^\perp \beta_C^\perp$ and point motion as $\mathbf{X} = \Theta_X \beta_X + \Theta_X^\perp \beta_X^\perp$ where \perp refers to the null space of Θ and β for the camera and point. Reconstructibility can be defined by the measure:

$$\eta = \frac{\|\Theta_C^\perp \beta_C^\perp\|^2}{\|\Theta_X^\perp \beta_X^\perp\|^2} \quad (4)$$

as η approaches infinity the solution to Equation 3 will tend towards its true solution. Unfortunately, in practice for non-rigid objects moving smoothly/slowly in 3D space this requires the camera to be moving extremely quickly around the object or randomly. This requirement drastically minimizes the usefulness of trajectory basis NSRFM in real-world problems.

3. Adding 3D Body Keyframes

Recently, Wei and Chai [13] proposed an approach for solving non-rigid structure from motion problem specifically for bodies. This work differs substantially to the work reviewed in Section 2 as it requires no temporal continuity of the 2D projections or constraints on camera motion. Instead, their approach takes advantage of assumed, and empirically validated, rigid constraints in the human body’s torso and hip. Their approach is notable in comparison to previous literature [1, 7, 3, 4] in the area as it: (i) makes no assumptions about bone lengths or camera scale, (ii) is not limited/constrained to modeling shapes previously seen in a train set, and (iii) can handle missing body points. Subsequent work by Valmadre and Lucey [11] demonstrated how this initial approach could be dramatically sped up using a deterministic method, and placed new bounds on how many frames/images are required. One of the major drawbacks to the approach, however, is the requirement for interaction from a human user to resolve depth ambiguities. The approach is useful for obtaining 3D structure from a small

¹We should note that the work of Park et al. [6] concerning reconstructibility was derived for the case of perspective geometry. In our work we shall be only assuming weak perspective, where the measure of reconstructibility still obviously holds.

number of image frames, but is largely unsuitable for obtaining 3D motion from a image sequence containing thousands of frames.

Compressibility of 3D Human Motion: An interesting research question now arises as a result of [11, 13] since we are able to obtain 3D structure from a video sequence, but only for a small discontinuous subset of all frames in the sequence. An example of this situation can be seen in Figure 2(a) for one dimension from a single point (left radius) of a human motion sequence taken from the CMU Mocap database. The blue line in Figure 2(a) denotes the full motion trajectory, and the red dots denote an example of the random subset of keyframes available to us. We denote the randomly chosen 3D keyframes as $\mathbf{X}' \in \mathcal{R}^{3M}$ where $M \ll F$ is the number of randomly chosen keyframes and F is the total number in the full 3D motion trajectory \mathbf{X} .

In this section we argue that through the application of a trajectory basis along with access to a random subset of 3D keyframes \mathbf{X}' a good reconstruction of the full 3D trajectory \mathbf{X} can be obtained. Our argument shall center upon the solution to the following objective function:

$$\arg \min_{\beta} \|\mathbf{X}' - \Theta' \beta\|^2 + \lambda \|\beta\|_p^p \quad (5)$$

where $\Theta' \in \mathcal{R}^{3M \times 3K}$ is a submatrix of the full trajectory basis Θ relating to the M random keyframes. The second term denotes a necessary p -norm regularization term to overcome problems associated with the first term of Equation 5 being under-determined. The λ term controls the influence of the regularization term, and is tuned along with K through cross-validation. Once the trajectory coefficient vector β is estimated from Equation 5 it becomes possible to obtain an estimate \mathbf{X} based through Equation 2.

Unfortunately, as one can see in Figure 2(b) the reconstruction of \mathbf{X} when using a traditional $p = 2$ regularization term (i.e. ridge regression) gives poor reconstruction. This poor reconstruction for $p = 2$ stems from a well understood result in the field of compressed sensing [5]. Specifically, that if the coefficient vector β is compressible when estimated from the full trajectory \mathbf{X} , then this compressible constraint needs to be enforced when estimating β from random samples². A common approach for enforcing sparsity/compressibility is through the employment of a $p = 1$ regularization term (i.e. lasso regression). An example of the $p = 1$ reconstruction can be seen in Figure 2(c) giving an almost perfect reconstruction. A central aspect of the objective function in Equation 5 is that the motion of the camera, unlike trajectory basis NSRFM methods, has no bearing on our ability to estimate β .

²It should be noted that the sampling strategy employed in compressed sensing problems is important with respect to the accuracy of reconstruction. All the work in this paper assumes that keyframes are sampled randomly. Additional, work is required on what the optimal strategy for sampling human motion is with respect to reconstruction.

Empirically, we have found in our experiments on human motion that we need at least 2 – 3 key-frames per second (on average assuming random sampling) to obtain useful 3D trajectory estimation. The next section of this paper shall investigate how 2D point image projections can reduce the number of required key-frames further. Specific details on the number of random samples required for good reconstruction are discussed in the experimental portion of this paper (Section 4).

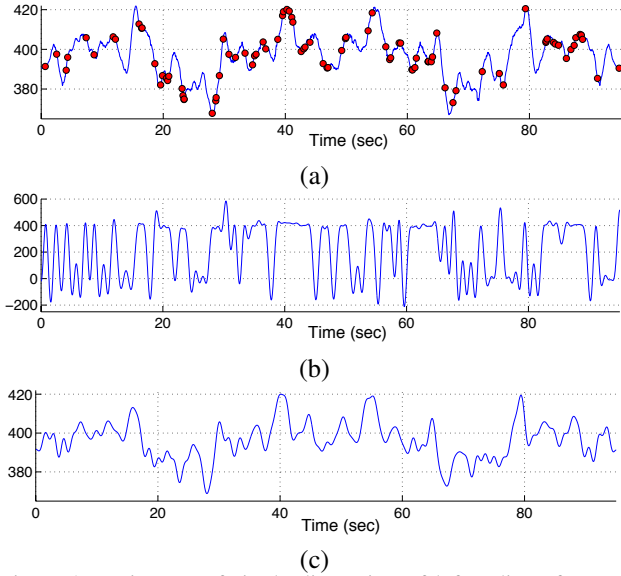


Figure 2. Trajectory of single dimension of left radius of example human motion sequence: (a) the original trajectory (red points are randomly selected key frames), (b) the trajectory reconstructed with a $p = 2$ regularization term, (c) the trajectory reconstructed with a $p = 1$ regularization term.

Reconstructibility & 3D Keyframes: A drawback to the objective function described in Equation 5 is that it does not take into account additional information that could be provided from other tools in computer vision (e.g., 2D point tracking). Conversely, a drawback to the trajectory basis NRSFM methods, based solely on 2D point projections (described in Section 2) is low reconstructibility. Specifically, when camera motion is correlated with point motion (e.g., if cameras and points have similarly slow/smooth trajectories) reconstructibility is poor. In this section we propose an objective function that overcomes the shortcomings of both these methods:

$$\arg \min_{\beta} (1-\gamma) \| \mathbf{x} - \mathbf{P}\Theta\beta \|^2 + \gamma \| \mathbf{X}' - \Theta'\beta \|^2 + \lambda \| \beta \|^p \quad (6)$$

where γ is a weighting factor that controls the influence of the 3D keyframes in relation to the 2D image projections. Inspecting Equation 6 one should note that when $\gamma = 1$ the problem becomes the compressed sensing problem described earlier in Equation 5. When $\gamma = 0$ the problem be-

comes the trajectory basis NRSFM problem described earlier in Equation 3 (with the exception of the regularization term).

Estimating Camera Matrices from 2D Points: In the original work of Akhter et al. [2] a method for solving for \mathbf{P} and β simultaneously was proposed. In our experiments, however, we found much better empirical performance could be obtained if \mathbf{P} was calculated from a rigid substructure of the body. As suggested by Wei and Chai [13] we employed the human torso (due to its approximate rigidity) to estimate the concatenation of camera matrices \mathbf{P} from the 2D image projections through canonical structure from motion [8]. Due to the inherent ambiguities associated with canonical structure from motion a \mathbf{P} was employed that was consistent with known random key frames \mathbf{X}' so as to ensure consistency in the objective function described in Equation 6.

4. Experiments

CMU Mocap:

In this section, we show results of our proposed method on a subset of the CMU motion capture (Mocap) dataset. We used the data in Akhter et al.’s paper [2] such as “Yoga” and “Drinking” and we added more clips such as “Running” from CMU Mocap. In Figure 3 we present the root mean square (RMS) error of 3D motion reconstruction for (a) random, and (b) smooth slow moving camera trajectories (see Figure 1 for a visualization of these two camera trajectories). Camera motions were added artificially for (a) and (b) so as to allow direct comparisons between the same motion sequences. Results are presented as a function of average 3D key frames per second (kfps) employed assuming a random sampling strategy. Complete 2D image projection sequences are assumed. Reconstruction results were obtained using Equation 6 for: (i) 2D only points ($\gamma = 1, p = 2$), (ii) 3D only points ($\gamma = 0, p = 2$), (iii) 3D and 2D points (γ tuned, $p = 2$), (iv) 3D with L_1 constraint ($\gamma = 0, p = 1$), and (v) 3D and 2D points with L_1 constraint (γ tuned, $p = 1$). The two parameters in Eq. 6 were chosen through a cross-validation procedure using a validation set stemming from CMU MoCap. We found stable ranges that (λ 0.05-0.15, γ 0.2-0.3) gave best performance. Further, we found that performance gracefully degraded from these optimal values indicating that good performance could still be obtained even when optimal values were not chosen.

Good performance is obtained in Figure 3(a) using 2D only information, whereas for real-world camera motion in Figure 3(b) performance is quite poor. This result is consistent with the problem of reconstructibility in trajectory basis NRSFM outlined in Section 2. A visualization of this poor performance, for real-world camera motion, can be seen in Figure 4(b). Another interesting result shown

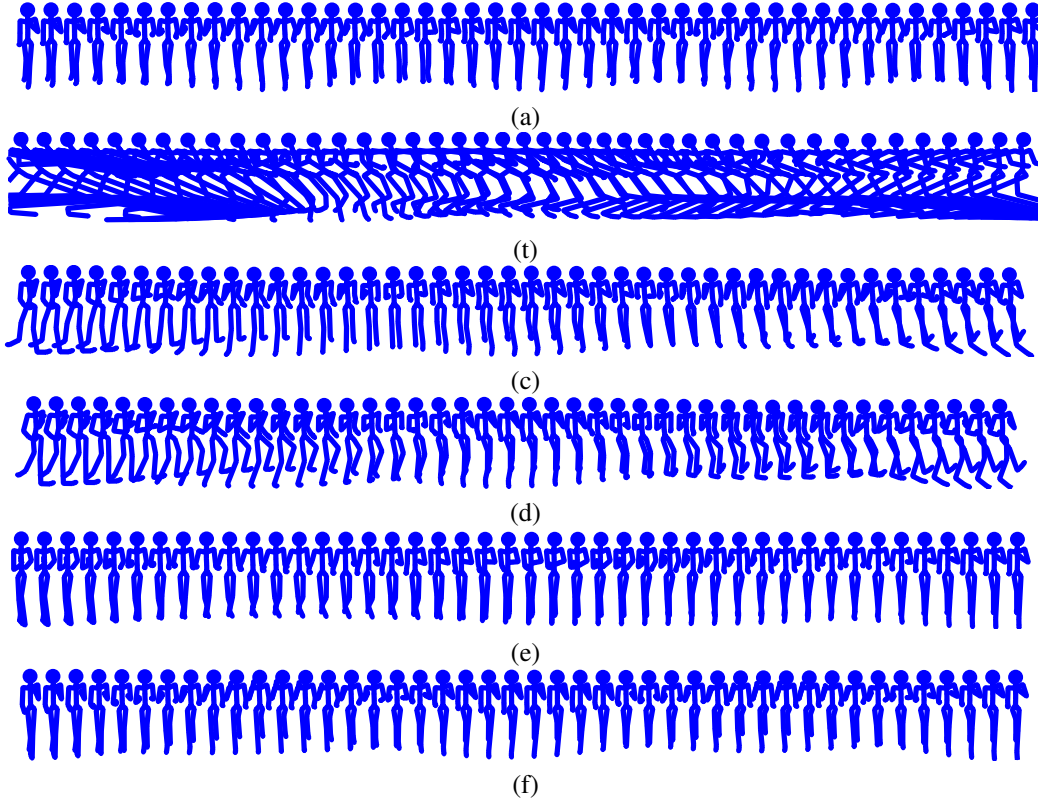


Figure 4. Visualizations of 3D motion reconstructions of an example sequence from CMU Mocap Subject 2-3 shows running: (a) the original motion, (b) 2D only, (c) 2D+3D+ L_1 (0.5 kfps), (d) 2D+3D+ L_1 (1.0 kfps), (e) 2D+3D+ L_1 (2.0 kfps), (f) 2D+3D+ L_1 (3.0 kfps).

in Figure 3 is that all the methods (except 2D only) converge after 3kfps. That’s because the DCT basis itself acts as a subspace regularizer. This subspace regularizer starts becoming dominant (over the other regularizers e.g. L_1) at a critical density of random 3D keypoints (3kfps, or a keyframe every 0.3 seconds). However it is more important how these results differ before this critical point. Interestingly, however, almost identical performance to the 2D only result for a random camera in Figure 3(a) is obtained in figure 3(b) when one employs 3D keyframes with 2D frames and the L_1 constraint at a very modest 0.5 kfps (i.e., hand labeling an image frame on average every 2 seconds). A visualization of this reconstruction can be seen in Figure 4(c). As suggested by Figure 3(b), results for 3D keyframes with 2D frames and the L_1 constraint keep improving as the kfps is increased. Visualizations can be seen in Figure 4 for (d) 1.0 kfps, (e) 2.0 kfps, and (f) 3.0 kfps.

As expected, there is no difference between (a) random and (b) real-world camera trajectories in Figure 3 for the 3D only approaches (with and without the L_1 constraint). The L_1 constraint does improve performance in the 3D only scenario, but only for specific kfps. There is, however, a substantial advantage in employing this same L_1 constraint when using 3D keyframes and 2D frames as depicted in 3(b). This result reinforces our argument in Section 3 for

a joint objective function that combines 3D keyframes, 2D frames and an L_1 constraint to encourage compressibility.

Real-World Results:

In Figure 5(a) we see consecutive image frames of a real-world sequence from the movie *Run Lola Run* (The images are reversed in time so that motions are consecutive from right to left). For this experiment we had access to 2D points for almost all frames in the image sequence (labeled in red). For the purposes of demonstration the 2D points were hand labeled. Future work shall explore automated approaches. We did not explore the role of noise on the 2D point tracks. We applied [11] to generate 3D keyframes from the random 2D keyframes in (a) through a human labeler (in order to resolve depth ambiguities) Figure 5(b) show a result of consecutive reconstructed 3D frames by our method for each frame in (a). In Figure 5(c) we generate 3D keyframes using the trajectory basis NRSFM method of [6]. This reconstruction was obtained using all 2D points in the sequence, without the need for a human labeler. One can see for this real-world example the trajectory basis NRSFM method in (c) obtains quite a poor reconstruction, whereas our approach in (b) obtains a realistic solution. The poor reconstruction in (c) is consistent with the problem of reconstructibility discussed in earlier sections, due to the slow movement of

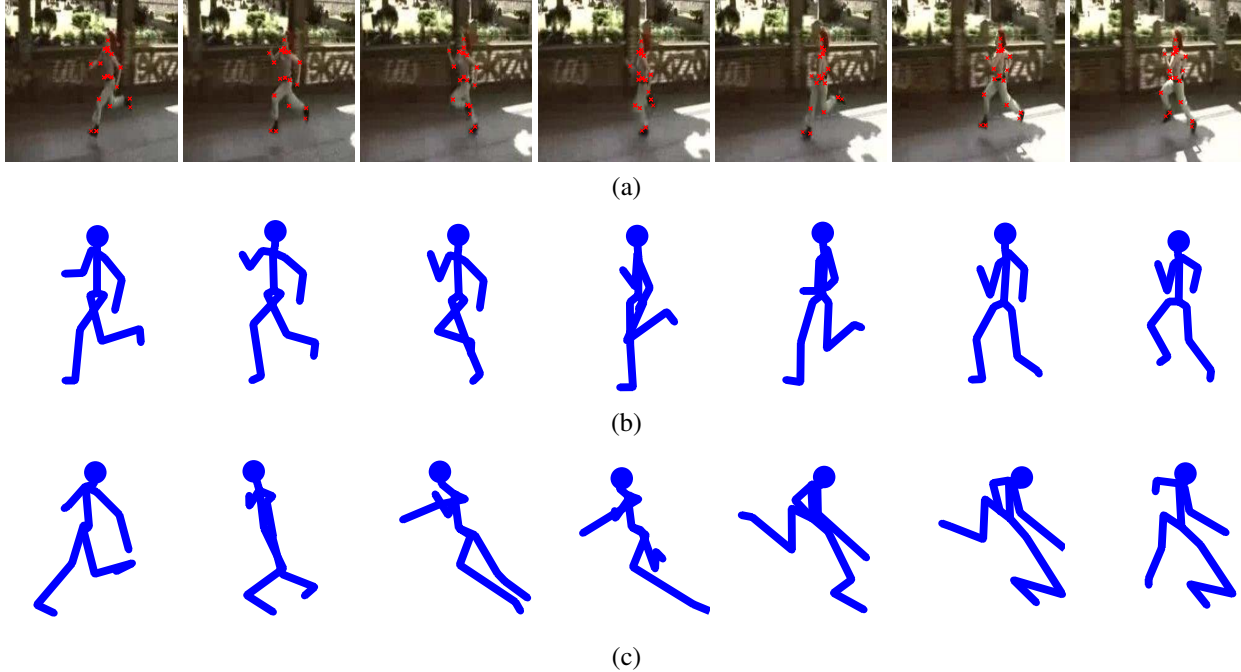


Figure 5. 3D frame reconstruction from the running sequence in the movie *Run Lola Run*: (a) we see consecutive frames with 2D joint locations superimposed in red, (b) depicts 3D reconstruction using our method, and (c) depicts 3D reconstruction using the trajectory basis NRSFM method of [6]. Due to reconstructibility issues concerning the slow moving camera in (a), the performance in (c) is quite poor whereas the result in (b) is invariant to camera motion.

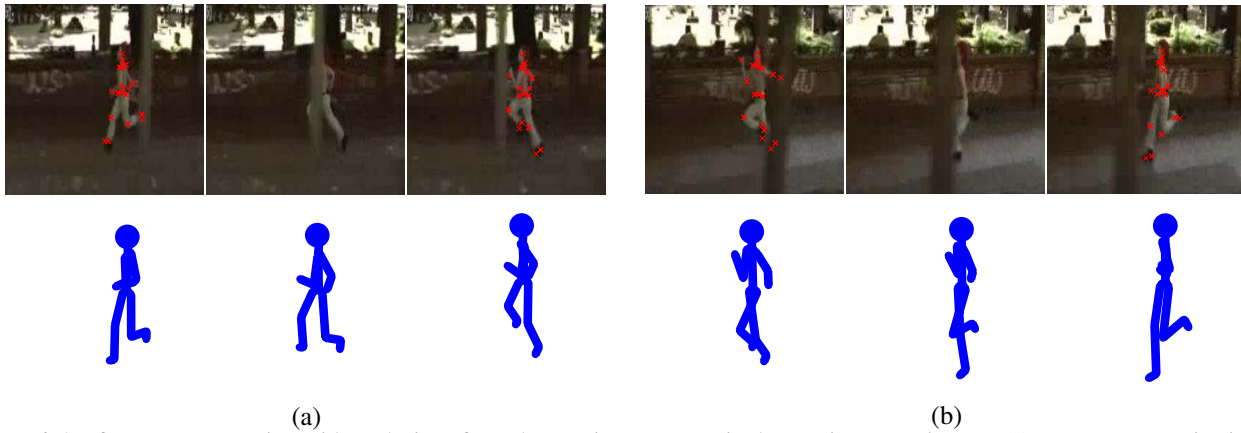


Figure 6. 3D frame reconstruction with occlusions from the running sequence in the movie *Run Lola Run*: (a) we see consecutive image frames with 2D joint locations superimposed in red and occluded image frame without 2D joint locations and 3D frame reconstruction using our method, (b) another consecutive image frames with 2D joint locations superimposed in red and occluded image frame without 2D joint locations and 3D frame reconstruction using our method.

the camera in the image sequence.

In Figure 6 (a) and (b), we see two sequence of consecutive image frames including occlusions and the reconstructed results by our proposed methods. We could not label the 2D points correctly because of the occlusions in the two middle images. And this kind of occlusion always exist in real world video. However, Figure 6 (a) and (b) show that our method can reconstruct the 3D pose even when some 2D points are lost due to occlusion. That's because L_1 penalty

can reconstruct compressible signal from incomplete samples [5].

In Figure 7 one can see the results for obtaining a reconstruction of the full 3D motion of the image sequence from *Run Lola Run* (We used 5 seconds of data from the movie "Run Lola Run". Due to limited space we depicted only a subset of this sequence in Figure 7). In (a) we see the reconstruction results for our proposed joint objective function in Equation 6, which employs 3D keyframes, 2D

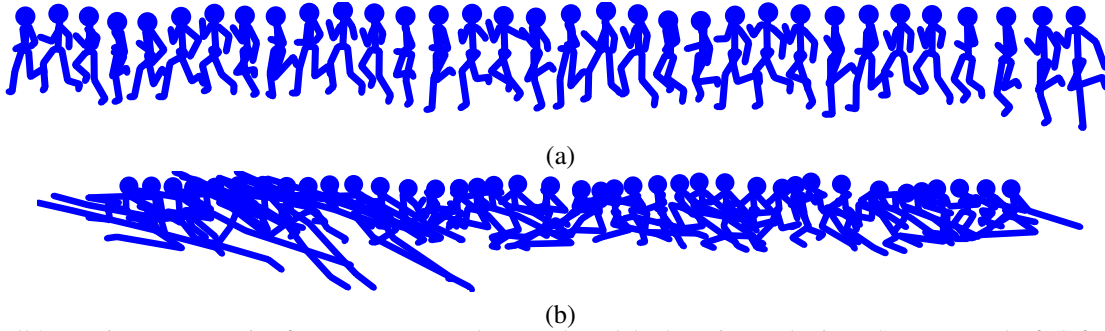
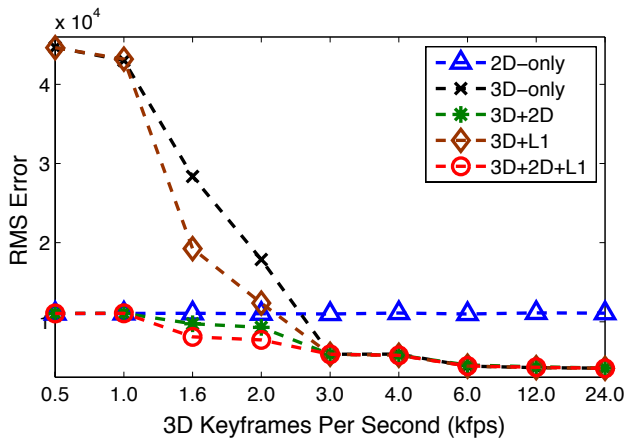
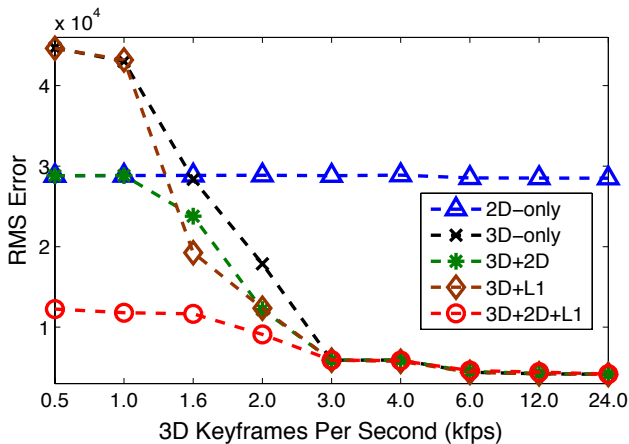


Figure 7. Full 3D motion reconstruction for: (a) our proposed approach, and (b) the trajectory basis NRSFM approach of [6] for the running sequence in the movie *Run Lola Run*.



(a)



(b)

Figure 3. Root mean square (RMS) motion reconstruction error for: (a) random, and (b) slow/smooth camera motion.

frames, and a L_1 penalty. As a comparison we also depict in (b) the 3D motion reconstruction using trajectory basis NRSFM. As expected our proposed approach in (a) dramatically outperforms the trajectory basis NRSFM approach in (b) due to its poor reconstructibility.

5. Discussion and Conclusions

In this paper we have proposed a novel approach for reconstructing 3D motion information purely from 2D image point projections. Specifically, we propose that trajectory basis NRSFM approaches to this problem have inherent problems with respect to reconstructibility. We note that good performance for 3D motion trajectories can be obtained from a small discontinuous subset of 3D keyframes by employing redundancies associated with human motion associated with its compressibility in a trajectory basis (i.e., DCT basis). An important component of this result is that it is, unlike trajectory basis NRSFM, invariant to camera trajectory. Based on this insight we propose a novel approach for combining 3D randomly sampled keyframes, 2D frames, and an L_1 penalty term to encourage compressibility. Based on the approach of [11, 13] for obtaining random 3D keyframes through human labeling, we show impressive full 3D motion reconstruction on a real-world image sequence where traditional trajectory based NRSFM fails.

References

- [1] A. Agarwal and B. Triggs. 3D Human pose from silhouettes by relevance vector regression. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2004. 3
- [2] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Nonrigid structure from motion in trajectory space. In *NIPS*, 2008. 1, 2, 4
- [3] C. Barron and I. A. Kakadiaris. Estimating anthropometry and pose from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000. 3
- [4] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000. 2, 3
- [5] E. J. Candes and M. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, March 2008. 1, 3, 6
- [6] 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 (ECCV)*, 2010. 1, 2, 3, 5, 6, 7

- [7] C. J. Taylor. Reconstruction of articulated objects from point correspondences in a single uncalibrated image. *IEEE Conference on Computer Vision and Pattern Recognition*, 1:677–684, June 2000. [3](#)
- [8] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9:137–154, November 1992. [2](#), [4](#)
- [9] L. Torresani, A. Hertzmann, and C. Bregler. Learning non-rigid 3D shape from 2D motion. In *NIPS*, 2005. [2](#)
- [10] L. Torresani, A. Hertzmann, and C. Bregler. Non-rigid structure from motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):878–892, May 2008. [2](#)
- [11] J. Valmadre and S. Lucey. Deterministic 3d human pose estimation using rigid structure. In *European Conference on Computer Vision (ECCV)*, 2010. [2](#), [3](#), [5](#), [7](#)
- [12] X. Wei and J. Chai. Videomocap: Modeling physically realistic human motion from monocular video sequences. *ACM Transactions on Graphics (SIGGRAPH)*, 29(4), 2010. [2](#)
- [13] X. K. Wei and J. Chai. Modeling 3D human poses from uncalibrated monocular images. *IEEE International Conference on Computer Vision*, 2009. [2](#), [3](#), [4](#), [7](#)
- [14] J. Xiao, J. Chai, and T. Kanade. A closed form solution to non-rigid shape and motion recovery. *International Journal of Computer Vision*, 67:233–246, 2006. [2](#)