1-1-2010

Compressive evaluation in human motion tracking

Yifan Lu
Australian National University

Lei Wang
Australian National University, leiw@uow.edu.au

Richard Hartley
Australian National University

Hongdong Li
Australian National University

Dan Xu
Yunan University

Compressive evaluation in human motion tracking

Abstract

The powerful theory of compressive sensing enables an efficient way to recover sparse or compressible signals from non-adaptive, sub-Nyquist-rate linear measurements. In particular, it has been shown that random projections can well approximate an isometry, provided that the number of linear measurements is no less than twice of the sparsity level of the signal. Inspired by these, we propose a compressive anneal particle filter to exploit sparsity existing in image-based human motion tracking. Instead of performing full signal recovery, we evaluate the observation likelihood directly in the compressive domain of the observed images. Moreover, we introduce a progressive multilevel wavelet decomposition staged at each anneal layer to accelerate the compressive evaluation in a coarse-to-fine fashion. The experiments with the benchmark dataset HumanEvaI show that the tracking process can be significantly accelerated, and the tracking accuracy is well maintained and comparable to the method using original image observations.

Keywords

compressive, evaluation, human, tracking, motion

Disciplines

Engineering | Science and Technology Studies

Publication Details

Lu, Y., Wang, L., Hartley, R., Li, H. & Xu, D. (2010). Compressive evaluation in human motion tracking. 10th Asian Conference on Computer Vision (ACCV) (pp. 1-12). Berlin Heidelberg: Springer-Verlag.

Compressive Evaluation in Human Motion Tracking

Yifan Lu[1], Lei Wang[1], Richard Hartley[1,3], Hongdong Li[1,3], and Dan Xu[2]

[1] School of Engineering, CECS, Australian National University
[2] Department of Computer Science and Engineering, SISE, Yunan University
[3] Canberra Research Labs, National ICT Australia
{Yifan.Lu,Lei.Wang,Richard.Hartley,Hongdong.Li}@anu.edu.au,
danxu@ynu.edu.cn

Abstract. The powerful theory of compressive sensing enables an efficient way to recover sparse or compressible signals from non-adaptive, sub-Nyquist-rate linear measurements. In particular, it has been shown that random projections can well approximate an isometry, provided that the number of linear measurements is no less than twice of the sparsity level of the signal. Inspired by these, we propose a compressive anneal particle filter to exploit sparsity existing in image-based human motion tracking. Instead of performing full signal recovery, we evaluate the observation likelihood directly in the compressive domain of the observed images. Moreover, we introduce a progressive multilevel wavelet decomposition staged at each anneal layer to accelerate the compressive evaluation in a coarse-to-fine fashion. The experiments with the benchmark dataset HumanEvaII show that the tracking process can be significantly accelerated, and the tracking accuracy is well maintained and comparable to the method using original image observations.

1 Introduction

Compressive sensing (CS) acquires and reconstructs compressible signals from a small number of non-adaptive linear random measurements by combining the steps of sampling and compression [1, 2, 3, 4]. It enables the design of new kinds of compressive imaging systems, including a single pixel camera [5] with some attractive features, including simplicity, low power consumption, universality, robustness, and scalability. Recently, there has been a growing interest of compressive sensing in computer vision and it has been successfully applied to face recognition, background subtraction, object tracking and other problems. Wright et al [6] represented the test face image in a linear combination of training face images. Their representation is naturally sparse, involving only a small fraction of the overall training database. Such a problem of classifying among multiple linear regression models can be then solved efficiently via $L1$-minimisation which seeks the sparsest representation and automatically discriminates between the various classes presented in the training set. Cevher et al [7] cast the background subtraction problem as a sparse signal recovery problem and solved by greedy

R. Kimmel, R. Klette, and A. Sugimoto (Eds.): ACCV 2010, Part IV, LNCS 6495, pp. 177–188, 2011.

methods as well as total variation minimisation as convex objectives to process field data. They also showed that it is possible to recover the silhouettes of foreground objects by learning a low-dimensional compressed representation of the background image without learning the background itself to sense the innovations or the foreground objects. Mei et al [8] formulated the tracking problem similar to [6]. In order to find the tracking target at a new frame, each target candidate is sparsely represented in the space spanned by target templates and trivial templates. The sparse representation is obtained by solving an $L1$-regularised least squares problem to find good target templates. Then the candidate with the smallest projection error is taken as the tracking target. Subsequent tracking is continued using a Bayesian state inference framework in which a particle filter is used for propagating sample distributions over time.

Unlike above works, many data acquisition/processing applications do not require obtaining a precise reconstruction, but rather are only interested in making some kind of evaluations on the objective function. Particularly, human motion tracking essentially attempts to find the optimal value of the observation likelihood function. Therefore, we propose a new framework, called Compressive Annealed Particle Filter, for such a situation that bypasses the reconstruction and performs evaluations solely on compressive measurements. It has been proven [1] that the random projections can approximately preserve an isometry and pairwise distance, when the number of the linear measurements is large enough (still much smaller than the original dimension of the signal). Moreover, noticing the annealing schedule is a coarse-to-fine process, we introduce the staged wavelet decomposition with respect to each anneal layer so that the increasing anneal variable is absorbed into the wavelet decomposition. As a result, the number of compressive measurements is progressively increased to gain computational efficiency.

The rest of the paper is organised as follows. Section 2 describes the human body template. In Section 3, we provide a brief overview of the theoretical foundation of Compressive Sensing, followed by Compressive Annealed Particle Filter in Section 4 and the results of experiments with the HumanEvaII dataset in Section 5. Finally, Section 6 concludes with a brief discussion of our results and directions for future work.

2 Human Body Template

The textured body template in our work uses a standard articulated-joint parametrisation to describe the human pose, further leading to an effective representation of the human motion over time. Our articulated skeleton consists of 10 segments and is parameterised by 25 degrees of freedom (DOF) in Figure 1. It is registered to a properly scaled template skin mesh by Skeletal Subspace Deformation (SSD) [9]. Then, shape details and texture are recovered by an interactive volumetric reconstruction and the texture registration procedure. At last, the template model is imported to commercial software to be finalised according to the real subject. The example of the final template model is illustrated in Figure 1.

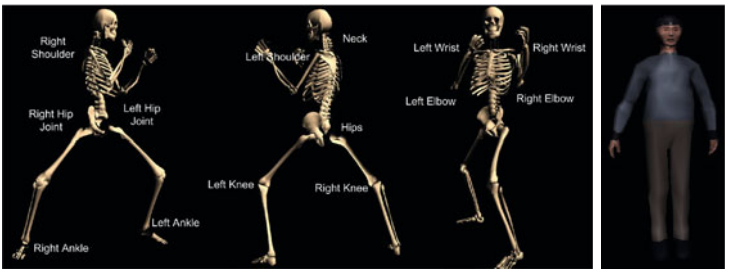

Fig. 1. From left to right: the articulated skeleton parameterised by 25 DOF and the textured template model after manual refinements used in this work

3 Compressive Sensing

The novel theory of Compressive Sensing (CS) [1,2,3,4] provides a fundamentally new approach to data acquisition that provides a better sampling and compression when the underlying signal is known to be sparse or compressible, yielding a sub-Nyquist sampling criterion.

3.1 Signal Sparse Representation

We consider that a signal $\mathbf{f} \in \mathbb{R}^N$ is sparse in some orthonormal basis $\mathbf{\Psi} \in \mathbb{R}^{N \times N}$ and can be represented as $\mathbf{f} = \mathbf{\Psi f'}$. If there are only a few significant entries in $\mathbf{f'}$, and insignificant entries can be discarded without much loss, then $\mathbf{f'}$ can be well approximated by $\mathbf{f'}_K$ that is constructed by keeping the K largest entries of $\mathbf{f'}$ unchanged and setting all remaining $N-K$ entries to zero. Then $\mathbf{f}_K = \mathbf{\Psi f'}_K$ is so called K-sparse representation. Since $\mathbf{\Psi}$ is an orthonormal matrix, hence $\|\mathbf{f} - \mathbf{f}_K\|_2 = \|\mathbf{f'} - \mathbf{f'}_K\|_2$. If $\mathbf{f'}$ is sparse or compressible in the sense that the sorted magnitudes of its components x_i decay quickly, then the relative error $\frac{\|\mathbf{f}-\mathbf{f}_K\|_2}{\|\mathbf{f}\|_2}$ is also small. Therefore, the perceptual loss of $\mathbf{f}_K$ with respect to $\mathbf{f}$ is hardly noticeable.

3.2 L1 Minimisation Recovery

Compressive sensing nevertheless surprisingly predicts that reconstruction from vastly undersampled non-adaptive measurements is possible-even by using efficient recovery algorithms. Let us consider M ($M << N$) non-adaptive linear measurements $\mathbf{z}$ (so called *Compressive Measurement*) of a signal $\mathbf{f}$ using $\mathbf{z} = \mathbf{\Phi f}$, where $\mathbf{\Phi} \in R^{M \times N}$ denotes the measurement matrix. Since $M << N$, the recovery of $\mathbf{f}$ from $\mathbf{z}$ is underdetermined. If, however, the additional assumption is imposed that the vector $\mathbf{f}$ has sparse representation, then the recovery can be realised by searching for the sparsest vector $\mathbf{f'}^*$ that is consistent with the measurement vector $\mathbf{z} = \mathbf{\Phi\Psi f'}$. The finest recovery $\mathbf{f}^* = \mathbf{\Psi f'}^*$ is achieved when the sparsest vector $\mathbf{f'}^*$ is found. This leads to solving a L_0-minimisation problem.

Unfortunately, the combinatorial L_0-minimisation problem is NP hard in general [10]. In [2] Candes et al have shown that the L_1 norm yields the equivalent solution to the L_0 norm, resulting in solving an easier linear program, for which efficient solution methods already exist. When the measurement process involves a small stochastic error term $\|\eta\|_2 \leq \epsilon$, $\mathbf{z} = \mathbf{\Phi\Psi f'} + \eta$, the L_1-minimisation approach considers the solution of:

$$\min \|\mathbf{f'}\|_1 \quad \text{subject to} \quad \|\mathbf{\Phi\Psi f'} - \mathbf{z}\|_2 \leq \epsilon \tag{1}$$

This is an instance of second order cone programming [3] which has a unique convex solution.

The exact recovery from non-adaptive linear measure is not universal but conditional. The primary result [4] of CS states, if $\mathbf{\Phi}$ is incoherent with $\mathbf{\Psi}$ so that the coherence $\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sqrt{N} \max_{l,k\in[1,N]} |\langle \phi_l, \psi_k \rangle|$[1] is close to 1 and $M \geqslant C\mu^2(\mathbf{\Phi}, \mathbf{\Psi})K \log N/\sigma$ for some positive constant C and small values of σ, then $\mathbf{f'}$ in $\mathbf{z} = \mathbf{\Phi f} = \mathbf{\Phi\Psi f'}$ can be exactly recovered with overwhelming probability $1-\sigma$. Moreover, it turns out that a randomly generated matrix $\mathbf{\Phi}$ from an isotropic sub-Gaussian distribution (e.g. from i.i.d. Gaussian or Bernoulli/ Rademacher 1 vectors) is incoherent with high probability to an arbitrarily fixed basis $\mathbf{\Psi}$.

4 Compressive Annealed Particle Filtering

The proposed approach resides on the APF framework that is first introduced in human tracking by Deutscher et al. [11]. APF incorporates simulated annealing [12] for minimising an energy function $E(\mathbf{y}_t, \mathbf{x}_t)$ or, equivalently, maximising the observation likelihood $p(\mathbf{y}_t|\mathbf{x}_t)$ that measures how well a particle (an estimate pose configuration) $\mathbf{x}_t$ fits the observation $\mathbf{y}_t$ at time t. The observation likelihood is essential for APF in order to approximate the posteriori distribution, and it is often formulated in a modified form of the Boltzmann distribution:

$$p(\mathbf{y}_t|\mathbf{x}_t) = \exp\{-\lambda E(\mathbf{y}_t, \mathbf{x}_t)\} \tag{2}$$

where the annealing variable λ is ,$1/(k_B T_t)$, an inverse of the product of the Boltzmann constant k_B and the temperature T_t at time t. The optimisation of APF is iteratively done according to a predefined L-phase schedule $\{\lambda = \lambda_1, ..., \lambda_L\}$, where $\lambda_1 < \lambda_2 < ... < \lambda_L$, known as the annealing schedule. At time t, considering a single phase l, initial particles are outcomes from the previous phase $l-1$ or drawn from the temporal model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$. Then, all particles are weighted by their observation likelihood $p(\mathbf{y}_t|\mathbf{x}_t)$ and resampled probabilistically to select good particles which are highly likely to near the global optimum. Finally, particles are perturbed by a Gaussian noise with a diagonal covariance matrix ${P_l}^2$.

[1] ϕ_l is a row of $\mathbf{\Phi}$. ψ_k is a column of $\mathbf{\Psi}$. To simplify the notation, ϕ_l can be concatenated as the basis with N elements so that $\langle \phi_l, \psi_k \rangle$ is always computable.

[2] The perturbation covariance matrix P_l is used to adjust the search range of particles.

Considering the pose space model in a dynamic structure that consists of a sequence of estimate poses $\mathbf{x}_t$ at successive time $t = 1, 2, ...$, and each pose is associated with an image observation $\mathbf{y}_t^{obs}$ or a compressive measurement $\mathbf{z}_t^d$. At time t, the compressive measurement can be defined by:

$$\begin{aligned}\mathbf{z}_t^d &= \mathbf{\Phi}\mathbf{\Psi}\mathbf{y}_t^d \\ &= \mathbf{\Phi}\mathbf{\Psi}(\mathbf{y}_t^{obs} - \mathbf{y}_t^{bg}) \\ &= \mathbf{z}_t^{obs} - \mathbf{z}_t^{bg}\end{aligned} \tag{3}$$

where, $\mathbf{\Psi}$ denotes wavelet basis. In particular, $\mathbf{y}_t^d$ is the difference image generated by subtracting the background image $\mathbf{y}_t^{bg}$ from the original observation image $\mathbf{y}_t^{obs}$. It is known that the images acquired from the natural scene have highly sparse representation in the wavelet domain. The difference image calculated by subtracting the static background from the observation image has more pixel values close to zero, hence, the difference image $\mathbf{\Psi}\mathbf{y}_t^d$ is also highly sparse and compressible in general.

On the other hand, given the estimate state $\mathbf{x}_t$, the estimate compressive measurement $\hat{\mathbf{z}}_t^d$ of the difference image can be calculated by subtracting the background image $\mathbf{y}_t^{bg}$ from the synthetic foreground image $s^{fg}(\mathbf{x}_t)$, which is generated by projecting the human model with the pose $\mathbf{x}_t$ and camera parameters onto the image plane. This difference image is also compressible in the wavelet domain so that it can be defined by:

$$\begin{aligned}\hat{y}_{t,i}^d &= sil_i(\mathbf{x}_t) * (s_i^{fg}(\mathbf{x}_t) - y_{t,i}^{bg}) \quad i = 1, ..., N \\ \hat{\mathbf{z}}_t^d &= \mathbf{\Phi}\mathbf{\Psi}\hat{\mathbf{y}}_t^d\end{aligned} \tag{4}$$

where, $sil(\mathbf{x}_t)$ is a synthetic silhouette mask generated by the estimate state $\mathbf{x}_t$ which has 0s on all background entries and 1s on all the foreground entries. This mask operation is used to make the synthetic difference image is comparable to the original difference image.

4.1 Restricted Isometry Property and Pairwise Distance Preservation

Another important result of CS is the Restricted Isometry Property (RIP) [1] which characterises the stability of nearly orthonormal measurement matrices. A matrix $\mathbf{\Phi}$ satisfies RIP of order K if there exists an isometry constant $\sigma_K \in (0, 1)$ as the smallest number, such that $(1 - \sigma_K)\|\mathbf{f}'\|_2^2 \leq \|\mathbf{\Phi}\mathbf{f}'\|_2^2 \leq (1 + \sigma_K)\|\mathbf{f}'\|_2^2$ holds for all $\mathbf{f}' \in \Sigma_K = \{\mathbf{f}' \in \mathbb{R}^N : \|\mathbf{f}'\|_0 \leq K\}$. In other words, $\mathbf{\Phi}$ is an approximate isometry for signals restricted to be K-sparse and approximately preserves the Euclidean length, interior angles and inner products between the K-sparse signals. This reveals the reason why CS recovery is possible because $\mathbf{\Phi}$ embeds the sparse signal set Σ_K in $\mathbb{R}^M$ while no two sparse signals in $\mathbb{R}^N$ are mapped to the same point in $\mathbb{R}^M$.

If $\mathbf{\Phi}$ has i.i.d. Gaussian entries and $M \geq 2K$, then there always exists $\sigma_{2K} \in (0, 1)$ such that all pair-wise distances between K-sparse signals are well preserved [13]:

$$(1-\sigma_{2K}) \leq \frac{\|\mathbf{\Phi}\mathbf{f}_i' - \mathbf{\Phi}\mathbf{f}_j'\|_2^2}{\|\mathbf{f}_i' - \mathbf{f}_j'\|_2^2} \leq (1+\sigma_{2K}). \tag{5}$$

Meanwhile, Baraniuk and Wakin [14] present a Johnson-Lindenstrauss (JL) lemma [15] formulation with the stable embedding of a finite point cloud under a random orthogonal projection, which has a tighter lower bound for M.

Lemma 1. *[14] Let $\mathbb{Q}$ be a finite collection of points in $\mathbb{R}^N$. Fix $0 < \sigma < 1$ and $\beta > 0$. Let $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ be a random orthogonal matrix and*

$$M \geq \left(\frac{4+2\beta}{\sigma^2/2 + \sigma^3/3}\right) \ln(\#\mathbb{Q})$$

If $M \leq N$, then, with probability exceeding $1-(\#\mathbb{Q})^{-\beta}$, the following statement holds: For every $\mathbf{f}_i', \mathbf{f}_j' \in \mathbb{Q}$ and $i \neq j$

$$(1-\sigma)\sqrt{\frac{M}{N}} \leq \frac{\|\mathbf{\Phi}\mathbf{f}_i' - \mathbf{\Phi}\mathbf{f}_j'\|_2}{\|\mathbf{f}_i' - \mathbf{f}_j'\|_2} \leq (1+\sigma)\sqrt{\frac{M}{N}}$$

where a random orthogonal matrix can be constructed by performing the Householder transformation [16] on M random length-N vectors having i.i.d. Gaussian entries, assuming the vectors are linearly independent.

4.2 Multilevel Wavelet Likelihood Evaluation on Compressive Measurements

The above Equation (5), Lemma (1) and orthonormality of $\mathbf{\Psi}$ guarantee the pairwise distance to be approximately preserved provided that M is sufficient large. Therefore the CS recovery is not necessary to evaluate the observation likelihood. Instead, the observation likelihood can be directly calculated via the distance of compressive measurements in Equation (3) and (4).

$$p(\mathbf{y}_t|\mathbf{x}_t) = \exp\{-\lambda\|\mathbf{z}_t^d - \hat{\mathbf{z}}_t^d\|_2\} \tag{6}$$

Notice $\lambda > 0$, the above equation can be transformed as:

$$\begin{aligned} p(\mathbf{y}_t|\mathbf{x}_t) &= \exp\{-\|\lambda\mathbf{z}_t^d - \lambda\hat{\mathbf{z}}_t^d\|_2\} \\ &= \exp\{-\|\mathbf{\Phi}\lambda(\mathbf{\Psi}\mathbf{y}_t^d - \mathbf{\Psi}\hat{\mathbf{y}}_t^d)\|_2\} \end{aligned} \tag{7}$$

In the equation (7), $\mathbf{\Psi}\mathbf{y}_t^d$ and $\mathbf{\Psi}\hat{\mathbf{y}}_t^d$ are wavelet coefficients. According to multilevel wavelet decomposition, we construct two wavelet coefficient sequences of $\mathbf{C} = \{\mathbf{c}_i | i = 1, 2...\}$ and $\hat{\mathbf{C}} = \{\hat{\mathbf{c}}_i | i = 1, 2...\}$ for $\mathbf{\Psi}\mathbf{y}_t^d$ and $\mathbf{\Psi}\hat{\mathbf{y}}_t^d$. Furthermore, $\mathbf{c}_i \subset \mathbf{c}_{i+1}$ the current level wavelet coefficient are always a subset of its super level wavelet coefficient. Hence, $\|\mathbf{c}_i\|_1 < \|\mathbf{c}_{i+1}\|_1$ and $\mathbf{C}$ is considered a monotonically increasing sequence in terms of the magnitude (the same can be applied to $\hat{\mathbf{C}}$). For instance, a four-level wavelet coefficient sequence is

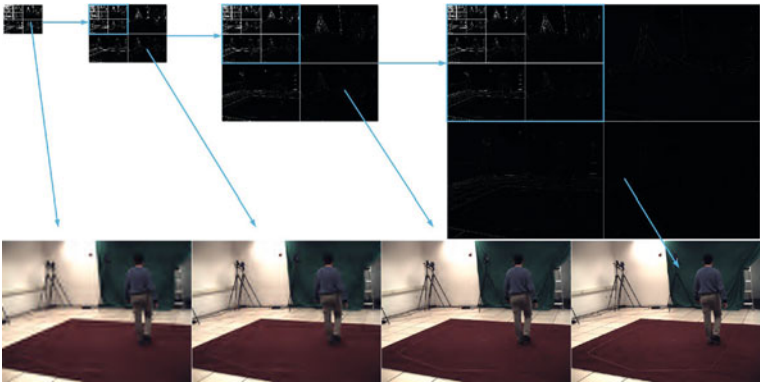

Fig. 2. The number of wavelet coefficients is progressively elevated as the wavelet decomposition process so that details are gradually enhanced through the anneal schedule. From left to right, we show 4 levels wavelet decomposition coefficients at the top of the figure. 1) using only the $K_4 = 2805$ largest coefficients (about 18.39% over all the level 4 coefficients) at the level 4, 2) $K_3 = 4345$ (7.18%) at the level 3, 3) $K_2 = 12086$ (5.01%) at the level 2 and 4) $K_1 = 30000$ (3.11%) at the level 1. The observation images at the bottom are reconstructed by using corresponding K_g sparse wavelet coefficients.

shown in the top of Figure 2. Obviously, $\mathbf{C}^{\Delta} = \mathbf{C} - \hat{\mathbf{C}}$ has the same monotonically increasing property $\|\mathbf{c}_i^{\Delta}\|_1 < \|\mathbf{c}_{i+1}^{\Delta}\|_1$. If defining a series of variables $\lambda_i = \|\mathbf{c}_{i+1}^{\Delta}\|_1 / \|\mathbf{c}_1^{\Delta}\|_1 \quad i = 1, 2, ...$, where $\lambda_i < \lambda_{i+1}$, alternatively, this monotonically increasing sequence $\mathbf{C}^{\Delta}$ can be described by $\mathbf{C}^{\Delta} = \{\mathbf{c}_1^{\Delta}, \lambda_1 \mathbf{c}_1^{\Delta}, \lambda_2 \mathbf{c}_1^{\Delta}, ...\}$. In other words, we always can construct a monotonically increasing wavelet coefficient sequence $\mathbf{C}^{\Delta}$ that has an equivalent counterpart series of λ. The precise value of λ for each anneal layer is not very critical, since λ is only used to roughly control the optimisation convergence rate. Therefore, we design directly evaluating the coarse-to-fine wavelet coefficients in difference levels to simulate increasing λ_l at each layer l. Then, an alternative of Equation (7) is given by:

$$p(\mathbf{y}_t|\mathbf{x}_t) = \exp\{-\|\mathbf{\Phi}(l)(\mathbf{\Psi}(l, \mathbf{y}_t^d) - \mathbf{\Psi}(l, \hat{\mathbf{y}}_t^d))\|_2\} \tag{8}$$

where, $\mathbf{\Psi}(l, \mathbf{y}_t^d)$ is wavelet coefficients of $\mathbf{y}_t^d$ at the l layer associated to the level g decomposition, and it has N_l wavelet coefficients. With l is increasing, g is decreasing and the more details encoded in wavelet coefficients $\mathbf{\Psi}(l, \mathbf{y}_t^d)$ are used. For instance, as shown in Figure 2. $\mathbf{\Phi}(l)$ is a $M_l \times N_l$ sub-matrix of $\mathbf{\Phi}$. $M_l = 2K_g$ is determined according to the sparsity K_g of the g level wavelet coefficients.

5 Experiments

Experiments are conducted on the benchmark dataset HumanEvaII [17] that contains two 1260-frame image sequences from 4 colour calibrated cameras synchronised with Mocap data at 60Hz. Those tracking subjects perform three different actions including walking, jogging and balancing. To generate compressive

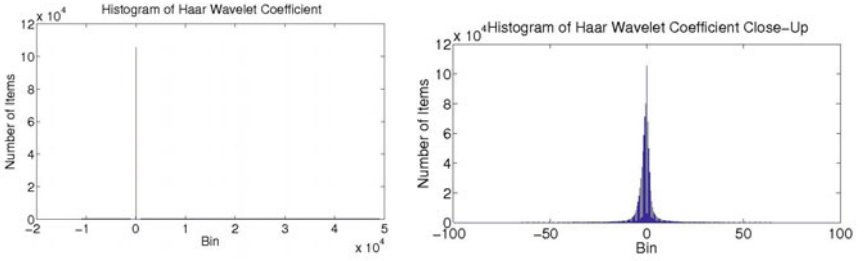

Fig. 3. Wavelet Coefficient Histogram and Wavelet Coefficient Histogram (close-up view) showing that 95% coefficients have very small values close to zero

measurements, we apply the 8-level haar wavelet 2D decomposition [18] to all observation images. The wavelet coefficients appear highly sparse, most of which are close to zero as illustrated in Figure 3. For instance, using solely the 30000 largest wavelet coefficients we are able to reconstruct the 964320 colour components of 656×490 RGB image with hardly noticeable perceptual loss. For the multilevel evaluation (Equation 8), the four sparsity levels $K_1 = 30000$, $K_2 = 12086$, $K_3 = 4345$ and $K_4 = 2805$ are evenly allocated in the 10 anneal layers[3]. The $M_l = 2K_g$ rows of $\mathbf{\Phi}$ are drawn i.i.d. from the normal distribution $N(0, 1/M_l)$ to approximately preserve the isometry as shown in Equation (5). On the other hand, the single level evaluation Equation (6) is used with a tight lower bound for M shown in Lemma (1). We presume there are one observation image and maximum 2000[4] synthetic images generated in the evaluation for each view and each frame. Then, for the 1260-frame sequence, there are total 2521260 unique compressive measurements required for tracking. Let $\sigma = 0.1$, $\beta = 1$ and $\#\mathbb{Q} = 2521260$, so $M = \left(\frac{4+2\beta}{\sigma^2/2+\sigma^3/3}\right) \ln(\#\mathbb{Q}) = 16583$. Moreover, the M rows of the $\mathbf{\Phi}$ are constructed by drawing i.i.d. entries from the normal distribution $N(0, 1/M)$ and performing the Householder transformation to orthogonalise $\mathbf{\Phi}$. Therefore, with high probability $1 - 1/2521260$, $\mathbf{\Phi}$ approximately preserves the pairwise distance. we also verified the performance of the number of compressive measurements in cases of $M = 10000$ and $M = 5000$.

As illustrated in the experimental results of HumanEvaII Subject 2 (the top of Figure 4), the evaluation using original images as the evaluation input obtains $54.5837 \pm 4.7516mm$[5]. The multilevel evaluation achieves the stable results $56.9442 \pm 4.4581mm$ which is comparable with the results using original images. When using the single level evaluation with $M = 16583$ compressive measurements, the tracking performance appears poorer than the multilevel evaluation but still maintains within $65.7548 \pm 5.4351mm$. When the number of compressive measurements are further reduced to $M = 10000$ and $M = 5000$, the

[3] Using $M_1 = 2 \times 2805, M_2 = 2 \times 2805, M_3 = 2 \times 2805, M_4 = 2 \times 4345, M_5 = 2 \times 4345, M_6 = 2 \times 4345, M_7 = 2 \times 12086, M_8 = 2 \times 12086, M_9 = 2 \times 30000, M_{10} = 2 \times 30000$.

[4] Given 10 layers and 200 particles as the maximum.

[5] The results are statistically presented by mean± standard deviation in Millimetres.

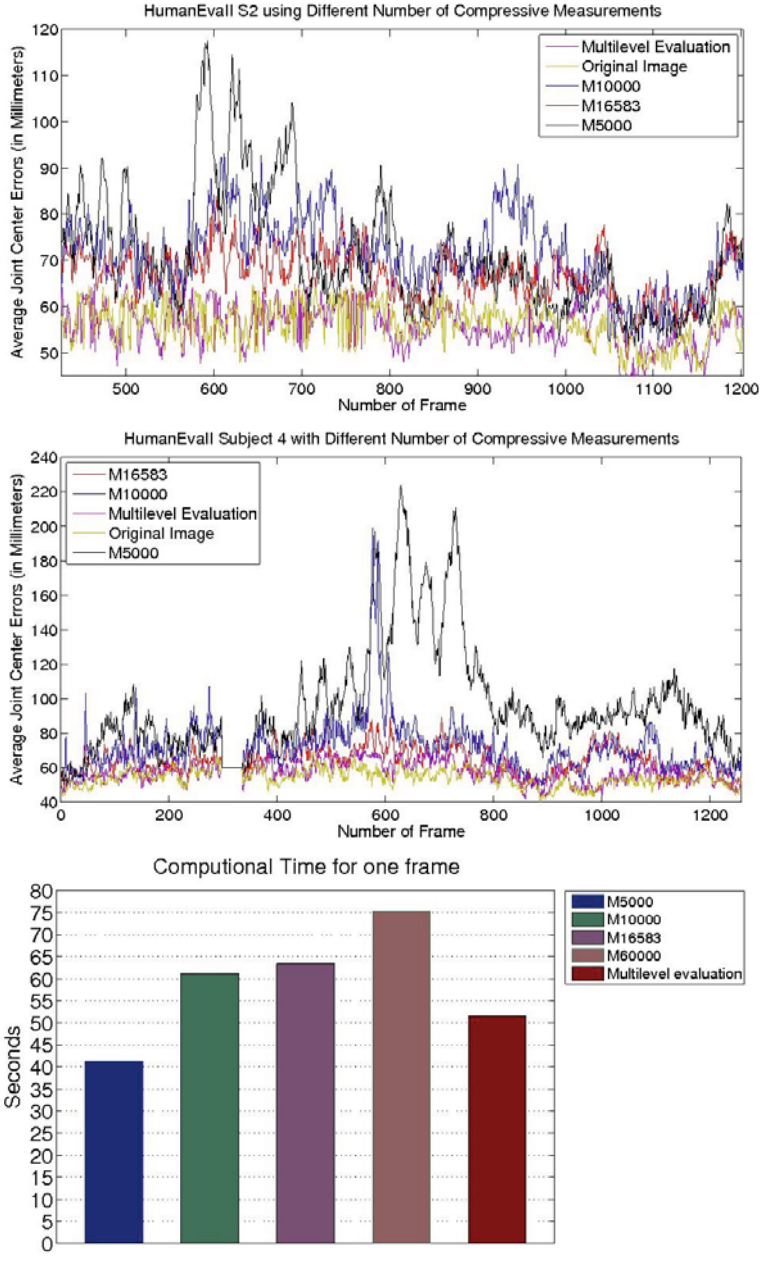

Fig. 4. From top to bottom, 1) tracking results of HumanEvaII Subject 2, 2) tracking results of HumanEvaII Subject 4 (the ground truth data is corrupted at 298-335 frames) and 3) computational time for one frame using the different number of compressive measurements

Fig. 5. HumanEvaII visual tracking results of Subject 4 and 2 are shown at the top four rows and the bottom four rows, respectively. The transparent visual model is overlapped with the tracking subject.

performance is degraded dramatically and we merely obtain $70.4249 \pm 7.5613mm$ and $68.2124 \pm 11.6153mm$, respectively. The middle of Figure 4 shows the experimental results of HumanEvaII Subject 4. The evaluation using original images achieves $54.2207 \pm 4.9250mm$ which is slightly better than $57.1705 \pm 6.0227mm$ achieved by the multilevel evaluation. Using $M = 16583$ compressive measurements experiences slightly more fluctuations comparing with the results of Subject 2. When the number of compressive measurements is decreased to $M = 10000$ and $M = 5000$, there are significant mistrackings and drifts with larger errors $71.6053 \pm 15.4005mm$ and $96.3663 \pm 32.8075mm$. More visual tracking results are shown in Figure 5.

The computational performance is also evaluated via the computational time for one frame using the different number of the compressive measurements shown in the bottom of Figure 4. As expected, the computational times from 40 to 75 seconds roughly correspond to increasing the number of the compressive

measurements M. On the other hand, the multilevel evaluation is able to reach the level of computational speed similar to merely using $M = 10000$ compressive measurements. Overall, the utilisation of progressive coarse-to-fine multilevel evaluation allows our approach to achieve the computational efficiency as only using $M = 10000$ compressive measurements and maintain the comparable tracking accuracy as using the original images.

6 Conclusion and Future Work

This paper has presented a compressive sensing framework for human tracking. It is realised by introducing a compressive observation model into the annealed particle filter. As the restricted isometry property ensures the preservation of the pairwise distance, compressive measurements with relative lower dimensions can be directly employed in observation evaluations without reconstructing the original image. Furthermore, noticing that there is a similar progressive process between the annealing schedule and the wavelet decomposition, we propose a novel multilevel wavelet likelihood evaluation in the coarse-to-fine fashion in which a fewer wavelet coefficients are used at the beginning, and then elevated gradually. This saves computational time and hence boosts the speed of evaluations. Finally, the robustness and efficiency of our approach are verified via the benchmark dataset HumanEvaII.

In compressive sensing recovery, many signal processing problems do not require full signal recovery and rather prefer to work on the compressive domain to benefit from dimensionality reduction. Indeed, RIP which approximately preserves an isometry allows to conduct evaluations and analysis on compressive measurements. However, the computational complexity of generating the sparse basis representation (in our case the wavelet decomposition) and compressive measuring still remains very high. In future work, we therefore would like to explore more about how to design more efficient the sparse basis representation and compressive measuring to handle the problem.

Acknowledgement. Authors would like to thank the support from National ICT Australia, and Leonid Sigal from Brown University provides the HumanEva dataset available.

References

1. Candes, E.J., Tao, T.: Decoding by linear programming. IEEE Transactions on Information Theory 51, 4203–4215 (2005)
2. Candès, E.J., Romberg, J.K., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory 52, 489–509 (2006)
3. Candès, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics 59, 1207–1223 (2006)

4. Candes, E.J., Romberg, J.: Sparsity and incoherence in compressive sampling. Inverse Problems 23, 969–985 (2007)
5. Duarte, M.F., Davenport, M.A., Takhar, D., Laska, J.N., Sun, T., Kelly, K.F., Baraniuk, R.G.: Single-pixel imaging via compressive sampling. IEEE Signal Processing Magazine 25, 83–91 (2008)
6. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 210–227 (2009)
7. Cevher, V., Sankaranarayanan, A., Duarte, M., Reddy, D., Baraniuk, R., Chellappa, R.: Compressive sensing for background subtraction. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 155–168. Springer, Heidelberg (2008)
8. Mei, X., Ling, H.: Robust visual tracking using l1 minimization. In: ICCV 2009, pp. 1436–1443 (2009)
9. Magnenat-Thalmann, N., Laperrière, R., Thalmann, D.: Joint-dependent local deformations for hand animation and object grasping. In: Proceedings on Graphics Interface 1988, Canadian Information Processing Society, pp. 26–33 (1988)
10. Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM Journal on Computing 24, 227–234 (1995)
11. Deutscher, J., Blake, A., Reid, I.: Articulated body motion capture by annealed particle filtering. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 126–133 (2000)
12. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science 220 (4598), 671–680 (1983)
13. Baron, D., Duarte, M.F., Wakin, M.B., Sarvotham, S., Baraniuk, R.G.: Distributed compressive sensing. The Computing Research Repository abs/0901.3403 (2009)
14. Baraniuk, R.G., Wakin, M.B.: Random projections of smooth manifolds. Foundations of Computational Mathematics 9, 51–77 (2009)
15. Johnson, W., Lindenstrauss, J.: Extensions of Lipschitz mappings into a Hilbert space. In: Conference in modern analysis and probability (New Haven, Conn., 1982). Contemporary Mathematics, vol. 26, pp. 189–206. American Mathematical Society, Providence (1984)
16. Householder, A.S.: Unitary triangularization of a nonsymmetric matrix. Journal of the ACM 5, 339–342 (1958)
17. Sigal, L., Black, M.J.: Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion. Technical report, Brown University, Department of Computer Science (2006)
18. Daubechies, I.: Ten Lectures on Wavelets. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia (1992)