

SuperSlicing Frame Restoration for Anisotropic ssTEM and Video Data

Dmitry Laptev

D LAPTEV@INF.ETHZ.CH

Joachim M. Buhmann

J BUHMANN@INF.ETHZ.CH

Department of Computer Science

ETH Zurich

8092 Zurich, Switzerland

Abstract

In biological imaging the data is often represented by a sequence of anisotropic frames — the resolution in one dimension is significantly lower than in the other dimensions. E.g. in electron microscopy it arises from the thickness of a scanned section. This leads to blurred images and raises problems in tasks like neuronal image segmentation. We present the details and additional evaluation of an approach originally introduced in Laptev et al. (2014) called SUPERSLICING to decompose the observed frame into a sequence of plausible *hidden sub-frames*. Based on sub-frame decomposition by SUPERSLICING we propose a novel automated method to perform neuronal structure segmentation. We test our approach on a popular connectomics benchmark, where SUPERSLICING preserves topological structures significantly better than other algorithms. We also generalize the approach for video anisotropy that comes from the long exposure time and show that our method outperforms baseline methods on a reconstruction of low frame rate videos of natural scenes.

Keywords: anisotropic data, super resolution, connectomics, segmentation, registration

1. Introduction

Digital imaging defines a quantization of the visual appearance of the world. The intensity of a pixel is the *cumulative* energy that has reached the physical sensor. In consequence, the details of a scene that are smaller than the spatial resolution of the sensor are getting *averaged away* (Fig. 1). Visually, averaging overcomes the problem of aliasing, but causes spatial blur and such data is called *anisotropic*.

Serial section transmission electron microscopy (ssTEM) Cardona and et al. (2010) of brain tissue is an important example. This method is the only available technique that guarantees sufficient resolution for reconstructing neuronal structures on the synapse level and, thereby, supports the scientific goals of connectomics Seung (2012) to understand brain functions. This technique renders the volume in a highly *anisotropic* way — the resolution across vertical dimension of the stack (thickness) is much lower than that of the horizontal dimensions.

The same phenomenon can be found in a low frame rate video recording. In case of *anisotropic video*¹ one can interpret the captured frame as an average of *hidden sub-frames*

1. Video is called *anisotropic* or *full-exposure* if exposure time equals to the time between two frames

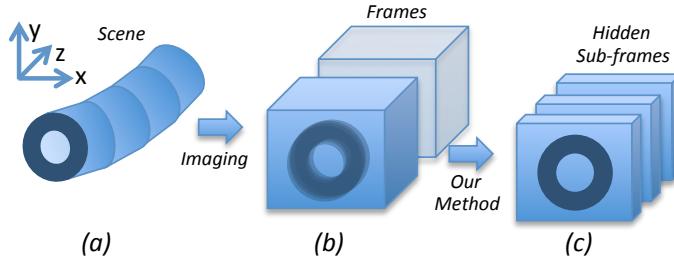


Figure 1: A schematic illustration of our approach: a) neuronal structure in brain tissue sample; b) the tissue sample is cut and captured with ssTEM, producing anisotropic frames with blur; c) the proposed method SUPERSLICING reconstructs *hidden sub-frames* with sharp details.

captured with shorter exposure time. The goal is then is to increase temporal resolution: estimate a high frame rate video from low frame rate.

We propose a method called SUPERSLICING (Super resolution frame Slicing). It reconstructs isotropic *hidden subframes* from a sequence of anisotropic frames, thereby increasing the depth or temporal resolution. This reconstruction states an inherently ill-posed problem as there exists an infinite number of possible sub-frames that can produce the same observed frame. We propose a regularisation that uses the information from the neighboring frames to resolve these ambiguities. The problem is formulated as energy minimization which appears to be convex and therefore guarantees the global optimum. The objective function is guided by two principal considerations: i) the physical constraints of the imaging process; ii) the structures in sub-frames should follow the correspondence between structures in the neighboring frames. To formalize the latter SUPERSLICING uses optical flow to find the correspondences between neighboring frames and interpolates them into sub-frames.

SUPERSLICING enables us to propose a novel automated method to perform neuronal structure segmentation (section 4). It recovers the crisp image of these structures and facilitates recognition of neural structures. The experiments on Drosophila first instar larva ventral nerve cord (VNC) dataset [Cardona and et al. \(2010\)](#) demonstrate significant improvement over the baselines.

2. Related Work

The first group of related techniques for frame enhancement interpolates between two neighboring frames. The simplest approach is a linear frame interpolation, which, although simple and fast, produces blurry results even when the initial frames are sharp. A more advanced technique [Baker et al. \(2011\)](#) is based on optical flow estimation and frame warping. However, in anisotropic data, frames are often reconstructed as blurred as initial frames because it takes into account no constraints on how imaging is performed. In contrast, SUPERSLICING reconstructs the changes *within* the frame, therefore recovering crisp details in each sub-frame. We use both of these approaches as baselines in our experiments.

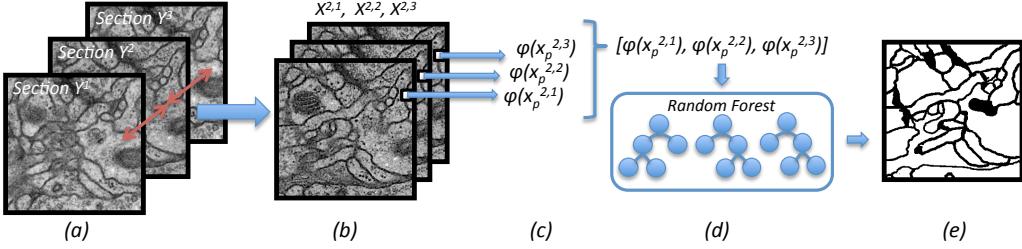


Figure 2: An illustration of the SUPERSLICING pipeline for neuronal structures segmentation. Based on the non-linear correspondences between neighboring frames Y^1 , Y^2 and Y^3 (a) the algorithm evaluates hidden sub-frames $X^{2,1}$, $X^{2,2}$, $X^{2,3}$ (b). Then, feature vectors in sub-frame pixels are evaluated: $\varphi(x_p^{n,1}), \dots, \varphi(x_p^{n,L})$ (c). After that the method concatenates them and passes the concatenated feature vector to a RF classifier (d) that returns the final segmentation (e).

Another approach [Hu and et al. \(2012\)](#) to solving the problem of spatial enhancement relies on using multiple ssTEM projections. Unlike these methods, we are considering a more general case and use only one sequence of frames from one ssTEM stack. And the third type of approaches [Shimano et al. \(2010\)](#) is based on exploring the recurrence of small self-similar patches in space and time. However, these methods assume that similar patches appear repeatedly within the frame sequence which is almost never the case for neuronal structures. In contrast to these methods we do not rely on high recurrence of self-similar patches and therefore, we solve a more general problem.

Neuronal structure segmentation and recognition has two general approaches. The first approach [Kaynig et al. \(2010\)](#) focuses on the detection of neuron membranes in each section independently based only on local information around every pixel. The second approach [Laptev et al. \(2012\)](#) incorporates context from different sections to resolve ambiguities that cannot be resolved within one section. The biggest challenge for the segmentation algorithm is posed by the blurry membranes (see Fig.5), that are often the result of anisotropy. We propose a novel method that first recovers the sharp sub-frames of a slice using SUPERSLICING and then uses them to perform segmentation. As the recovered sub-frames contain finer details the segmentation algorithm is able to identify the neuronal structures with higher accuracy than methods without SUPERSLICING.

3. Proposed Method

Let Y^n be the observed sequence of frames, $n \in [1, \dots, N]$, y_p^n – pixel p of the frame Y^n , $i(y_p^n)$ – the intensity of pixel y_p^n . Let $\epsilon(x_p^n)$ be a set of neighbors of pixel x_p^n . We want to reconstruct L hidden sub-frames $X^{n,l}$, $l \in [1, \dots, L]$ of the observed frames Y^n .

3.1. Optimization task

We define optimization problem 1 to approximate *hidden sub-frames* as an energy minimization problem for given correspondences Ω . The energy 1 consists of three terms. The

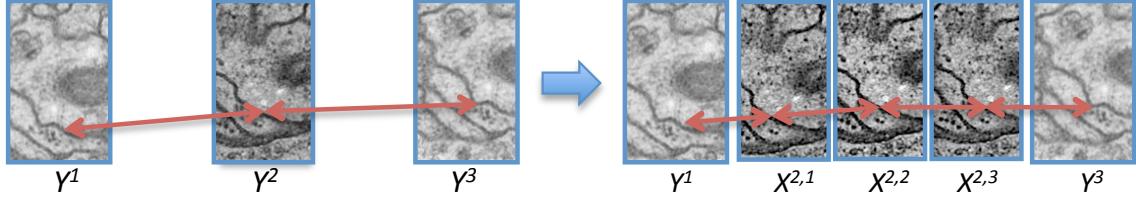


Figure 3: An illustration of correspondence interpolation. Left: arrows show correspondences between original frames; right: arrows shows interpolated correspondences between sub-frames. The second term of the energy function encourages the corresponding pixels to have low difference in intensities.

first term, the data term, represents the physical constraints that the observed frame should be equal to the average of the *hidden sub-frames*: $i(y_p^n) = \frac{1}{L} \sum_{l=1}^L i(x_p^{n,l}), \forall y_p^n \in Y^n$.

The second term promotes smoothness by favoring an alignment of pixel's intensities in the sub-frames along the structure's progression between the frames. The algorithm proceeds by finding correspondences between the anisotropic frames using optical flow and then interpolates them into the sub-frames using bilinear interpolation (see section 3.2).

The third term encourages the resulting sub-frames to be smooth to avoid visual artefacts. This goal is achieved by minimizing the difference of intensities between the neighboring pixels.

$$\begin{aligned}
E(X^{n,1}, \dots, X^{n,L}) = & \sum_{y \in Y^n} \left(i(y) - \frac{1}{L} \sum_{l=1}^L i(x_p^{n,l}) \right)^2 + \lambda \sum_{(\hat{x}_p^{n,l}, \hat{x}_q^{n,l+1}) \in \Omega} \left(\sum_{x \in \epsilon(\hat{x}_p^{n,l})} w(x, \hat{x}_p^{n,l}) i(x) \right. \\
& \left. - \sum_{x \in \epsilon(\hat{x}_q^{n,l+1})} w(x, \hat{x}_q^{n,l+1}) i(x) \right)^2 + \gamma \sum_{\substack{x_p^{n,l}; x_q^{n,l} \in \epsilon(x_p^{n,l}) \\ l=1, \dots, L}} \left(i(x_p^{n,l}) - i(x_q^{n,l}) \right)^2
\end{aligned} \tag{1}$$

Here λ and γ are Lagrange parameters that control the degree of regularization versus data fidelity. This is a quadratic functional with respect to $i(x_q^{n,l})$ and therefore we can achieve *global optimum* with any convex optimization technique (we used interior point method in our experiments).

3.2. Corresponding pixels

How can we find the set Ω of corresponding pixels? A central idea of this paper is to utilize the context of neighboring frames for reconstructing sub-frames. We first find the correspondences between the pixels in neighboring frames and only after these constraints have been identified, we interpolate these correspondences through sub-frames.

Assume that we observe the sequence of three images: $Y^1, Y^2 \equiv Y, Y^3$. For every pixel y_p^2 of Y^2 we find the corresponding pixel y_p^k from image Y^k , $k \in \{1, 3\}$ by finding the set

$\Omega_Y^k = \{(y_p^2, y_q^k) | \forall y_p^2 \in Y^2\}$ minimizing optical flow energy:

$$E_{fl}(\Omega_Y^k) = \sum_{y_p \in Y} \left(i(y_p) - i(y_q^k) \right)^2 + \alpha \sum_{y_p \in Y^2} \rho(y_p, y_q^k)^2$$

Here α is a model parameter, $\rho(y_p, y_q)$ is euclidean distance between the pixels y_p and y_q in pixel grid. Optical flow results in good correspondences, even though it allows only integer displacements, because the membrane displacements are smooth and need to be estimated only up to the thickness of a membrane, which is on average 3 to 7 pixels.

As soon as we have corresponding sets Ω_Y^1 and Ω_Y^3 , we can draw a curve φ through y_p^1 to y_q^2 and y_t^3 for every two correspondings (y_p^1, y_q^2) and (y_q^2, y_t^3) . Then we interpolate the pixels curve φ crosses in hidden sub-slices: $\hat{x}_{\varphi(1)}^1, \dots, \hat{x}_{\varphi(L)}^L$ (see Fig. 3). Then $\Omega_\varphi = \{(\hat{x}_{\varphi(l)}^l, \hat{x}_{\varphi(l+1)}^{l+1}) | l \in [1, \dots, L-1]\}$. The final set Ω is a union of all sets Ω_φ .

If pixel $\hat{x}_p^{n,l}$ does not fit to the pixel grid, we employ the bilinear interpolation technique and rewrite it as a weighted sum of direct neighbors in a grid $\hat{x}_p^{n,l} = \sum_{x \in \epsilon(\hat{x}_p^{n,l})} w(x, \hat{x}_p^{n,l}) x$, $w(.) \geq 0$, $\sum_{x \in \epsilon(\hat{x}_p^{n,l})} w(x, \hat{x}_p^{n,l}) = 1$. Here $w(x_1, x_2)$ is a bilinear weight that is closer to 1 if the distance between x_1 and x_2 is small and closer to 0 otherwise. We then write the second set of constraints enforcing that corresponding pixels of sub-frames assume the same intensity:

$$\sum_{x \in \epsilon(\hat{x}_p^{n,l})} w(x, \hat{x}_p^{n,l}) i(x) = \sum_{x \in \epsilon(\hat{x}_q^{n,l+1})} w(x, \hat{x}_q^{n,l+1}) i(x),$$

$\forall (\hat{x}_p^{n,l}, \hat{x}_q^{n,l+1}) \in \Omega$, where Ω is a set of all pairs of corresponding pixels.

4. Neuronal Segmentation

We propose a method that first reconstructs hidden sub-frames and uses features that are evaluated in pixels of recovered sub-frames for classification. Our workflow is illustrated in Figure 2. For a given section Y^n we first recover sub-frames $X^{n,1}, \dots, X^{n,L}$ with SUPERSLICING. Then, for every pixel $x_p^{n,l}$, $l \in [1, \dots, L]$ we calculate features $\varphi(x_p^{n,l})$, concatenate the feature vectors and use this extended feature vector as input to a Random Forest (RF) classifier [Breiman \(2001\)](#).

We select the method parameters γ and λ as well as optical flow parameter α with cross validation. We use RF with 255 trees and perform training on 10% of all the pixels. As features we use per pixel SIFT histograms [Lowe \(1999\)](#) and line filter transforms [Sandberg and Brega \(2007\)](#) with different parameters.

5. Experiments

To evaluate our approach we perform experiments on several different tasks and datasets. For all of the following experiments we select the method parameters γ and λ as well as optical flow parameter α with 5-fold cross validation and with respect to the corresponding metric used.

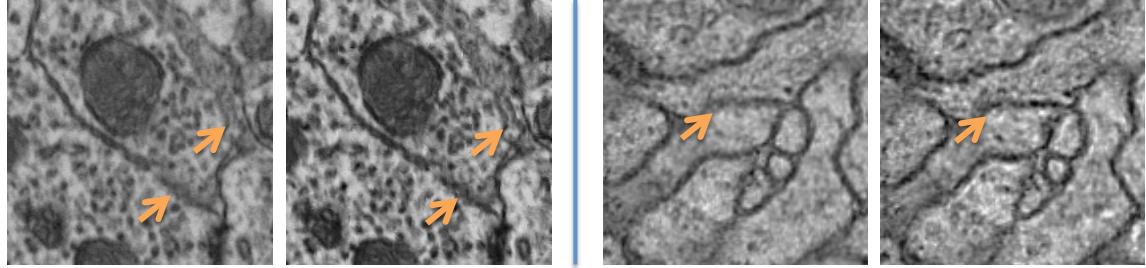


Figure 4: Two fragments of neuronal tissue captured with ssTEM: original sections (left) and one of sub-frames (right). Arrows point out membranes that were blurred out in the original images and appear more visible after sub-frame decomposition.

5.1. ssTEM imaging and Neuronal Reconstruction

We use publicly available segmentation challenge dataset [Cardona and et al. \(2010\)](#). Figures 4 and 5 qualitatively shows the results of our algorithm for hidden frame recovery. Membranes recovered in the sub-frames using SUPERSLICING are much sharper than the ones produced by the baseline methods.

To quantitatively test the approach for neuronal membrane segmentation presented in section 4, we compare segmentation results with two more methods: RF segmentation based on only features evaluated in one layer [Kaynig et al. \(2010\)](#), and RF segmentation based on context from neighboring sections [Laptev et al. \(2012\)](#). For fair comparison we implement the same set of features for all three methods and use the same RF structure with no post-processing to measure the impact of SUPERSLICING.

As we care about neurons topology, but not pixel-wise reconstruction, we also compare the results in terms of warping error [Jain and et al. \(2010\)](#). The warping error measures the topological error between proposed labeling \hat{X} and a reference labeling X^* . It is evaluated as squared Euclidean distance between X^* and the "best warping" F of \hat{X} onto X^* such that the warping F is from the class Λ that preserve topological structure: $\min_{F \in \Lambda} \sum_p \delta(F(\hat{X})_p, X_p^*)$. For further information about the warping error the interested reader is referred to [Jain and et al. \(2010\)](#). The results are summarized in table 1. The results on sub-frame stack produced by SUPERSLICING are 17% better than one sections segmentation and 11% better then the results based on three neighboring sections.

5.2. Natural videos

Rotating Fan We test the proposed algorithm on a rotating fan video from [Shahar et al. \(2011\)](#) to evaluate our method qualitatively². As the rotation speed is higher than the shutter speed the frame renders blurred fan blades. Based on three neighboring frames and

2. We do not compare with [Shahar et al. \(2011\)](#) directly, as their method operates under different assumptions and, moreover, they provide no quantitative results.

Method	Warping error
<i>One-section segmentation</i> Kaynig et al. (2010)	$2.876 * 10^{-3}$
<i>Three consecutive sections</i> Laptev et al. (2012)	$2.693 * 10^{-3}$
SUPERSLICING segmentation	$2.384 * 10^{-3}$

Table 1: Warping error on a testing set for one-section segmentation, segmentation based on three consecutive sections and for SUPERSLICING. Our method outperforms the baseline methods by 17% and 11%, respectively.

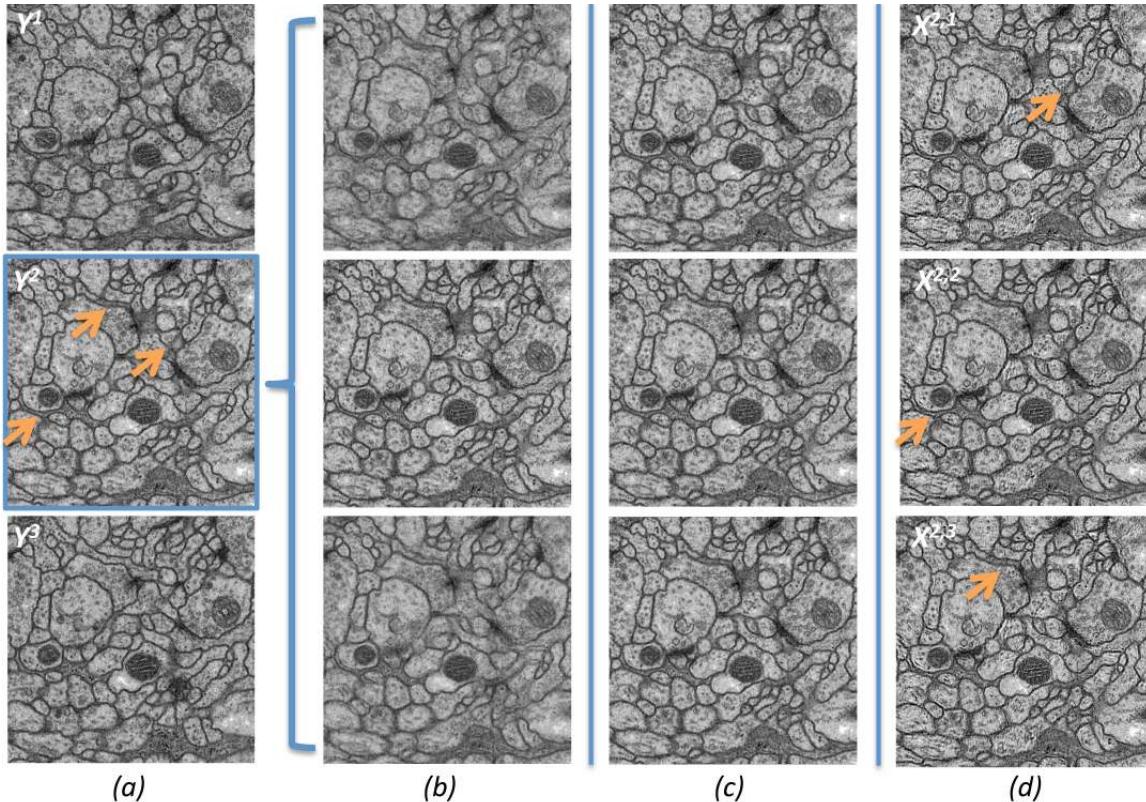


Figure 5: A qualitative comparison of our method with the baselines. Column (a) shows original anisotropic sections. Three following column shows $L = 3$ interpolated frames estimated with: linear interpolation (b), optical flow warping (c), SUPER-SLICING (d). Arrows point out blurred membranes that are better visible after sub-frame reconstruction.

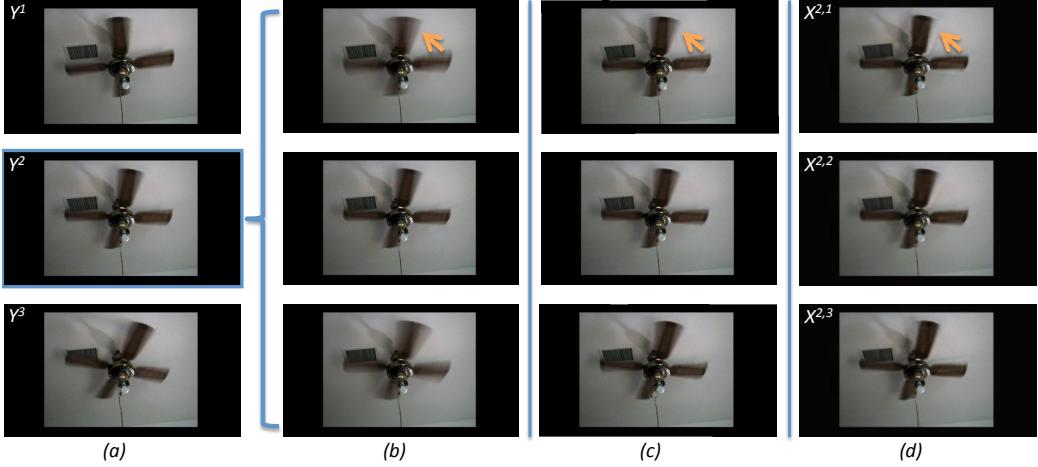


Figure 6: A comparison of SUPERSLICING with the results of alternative methods. Column (a) shows original frames Y^1 , Y^2 and Y^3 . Each following column shows three interpolated frames estimated with: linear interpolation (b), optical flow warping (c), SUPERSLICING (d). Arrows point out that SUPERSLICING results in less blurred fan blades.

no prior information we estimate $L = 3$ hidden sub-frames with linear interpolation, optical flow interpolation and the proposed method. Figure 6 shows the results of comparison. As can be seen linear interpolation blurs sub-frames even more. Optical flow interpolation shows the rotation of the fan, but as the initial frames are blurred, the resulting warping is blurred as well. SUPERSLICING shows superior results: it reconstructs the original shape of the blades and renders sharp sub-frames.

KTH dataset We perform synthetic experiments on the KTH action database [Schuldt et al. \(2004\)](#) to quantify the quality of SUPERSLICING’s reconstruction. This database consists of videos recorded at 24 frames per second. We first downsample the frame rate to 8 frames per second while taking an average of three neighboring frames (low frame rate videos). Then we reconstruct sub-frames with four different methods: frame repetition, linear interpolation, optical flow warping and SUPERSLICING. Figure 7 shows qualitative results for the number of hidden sub-frames equal $L = 2$ or 3 . Boxplots in Fig. 8 visualises the comparison of peak signal to noise ratio (PSNR) evaluated on 25 frames of video for $L = 3$ and $L = 2$ respectively. SUPERSLICING outperforms baseline methods for almost all frames and the average quantitative results appear to be significantly superior: 23% better for frame repetition and 10% for both linear interpolation and optical flow warping.

6. Conclusion

This paper addresses the problem of anisotropic data restoration in ssTEM microscopy. Our main contribution is a method called SUPERSLICING that decomposes an observed anisotropic frame into a sequence of *hidden isotropic sub-frames*. The proposed method

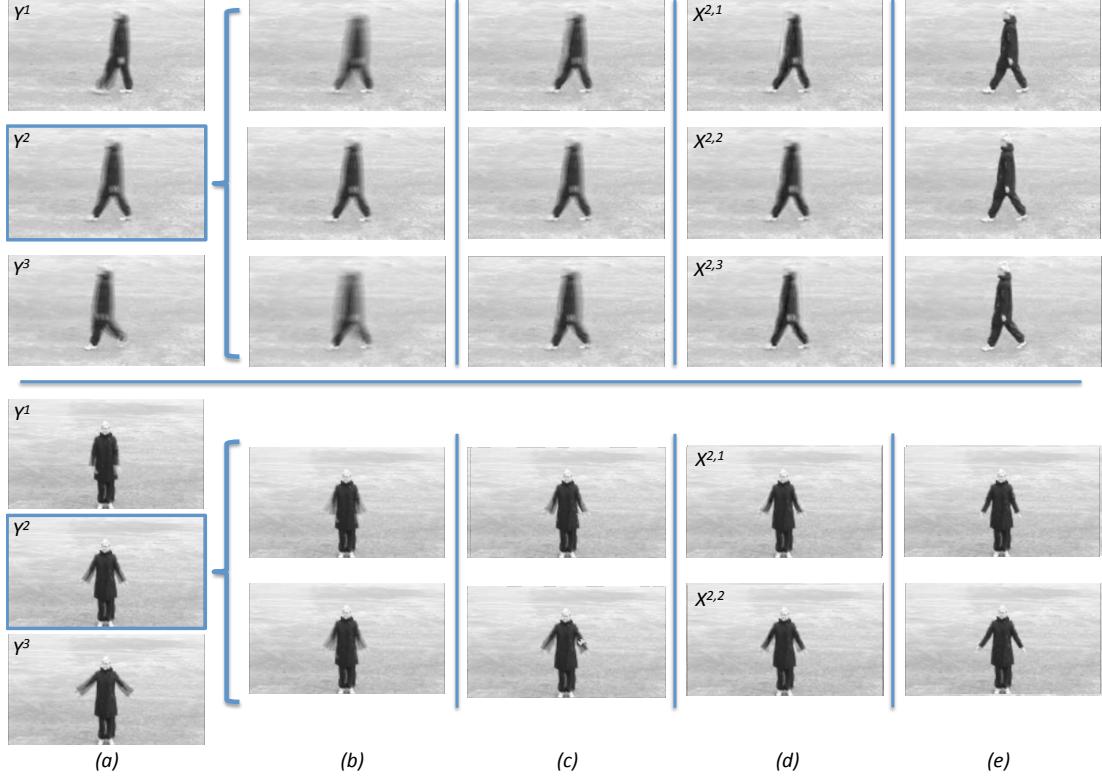


Figure 7: A comparison of our reconstruction results with the results of different methods and with ground truth. Top: walking person video reconstruction with $L = 3$ hidden sub-frames. Bottom: hand waving person video reconstruction with $L = 2$ hidden sub-frames. Column (a) shows original frames Y^1 , Y^2 and Y^3 from low frame rate video. Three following column shows $L = 3$ interpolated frames estimated with: linear interpolation (b), optical flow warping (c), SUPERSLICING (d). Column (e) shows ground truth from high frame rate video. Our results are less blurred and they are qualitatively closer to the ground truth than the results of the baseline methods.

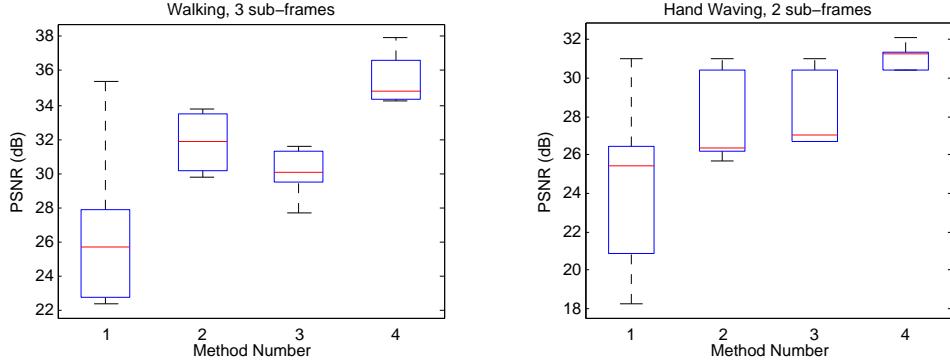


Figure 8: An illustration of quantitative results on KTH videos for different methods. Left plot: walking person video with $L = 3$ hidden sub-frames, right plot: hand waving person video with $L = 2$ hidden sub-frames. Each boxplot shows statistics for PSNR (in dB) evaluated for: frame repetition (1), linear interpolation (2), optical flow warping (3) and our method (4).

requires only two neighboring frames to perform the decomposition and it does not assume any special properties of the data.

SUPERSLICING incorporates two types of constraints. One of them represents physical properties of the involved imaging technique and the other constraint encourages the pixels that lie along the progression of objects between the frames to be of the same intensity. In order to find corresponding pixels we first find optical flow between observed frames and interpolate the flow into the sub-frames.

Based on SUPERSLICING we develop an algorithm for an automatic membrane segmentation in ssTEM sections. We show how to increase the performance of the segmentation algorithm by decomposing an observed anisotropic frame into isotropic sub-frames. We demonstrate the quality of the method on publicly available dataset where it performs, in term of warping error, 17% and 11% better than the baselines.

We also provide both qualitative and quantitative results for videos from the KTH action video dataset. We artificially synthesize blurred low frame rate video and decompose it into sub-frames. We evaluate PSNR and compare the results with three different baseline methods. Our results are on average 10% better than state-of-the-art.

Acknowledgments

This work was partially supported by the SNF grant Sinergia CRSII3_130470/1.

References

- S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92: 1–31, 2011.
- L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- A. Cardona and S. Saalfeld et al. An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy. *PLoS Biol*, 10, 2010.
- T. Hu and J. Nunez-Iglesias et al. Super-resolution using sparse representations over learned dictionaries: Reconstruction of brain structure using electron microscopy. *CoRR*, abs/1210.0564, 2012.
- V. Jain and B. Bollmann et al. Boundary learning by optimization with topological constraints. In *CVPR*, pages 2488–2495, 2010.
- V. Kaynig, T. J. Fuchs, and J. M. Buhmann. Geometrical consistent 3d tracing of neuronal processes in sstem data. In *MICCAI 2010*, pages 209–216. Springer Berlin / Heidelberg, 2010.
- D. Laptev, A. Vezhnevets, S. Dwivedi, and J. M. Buhmann. Anisotropic sstem image segmentation using dense correspondence across sections. In *MICCAI*, pages 323–330, 2012.
- D. Laptev, A. Vezhnevets, and J. M. Buhmann. Superslicing frame restoration for anisotropic sstem. In *IEEE 11th International Symposium on Biomedical Imaging ISBI 2014*. IEEE Xplore, 2014.
- D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, pages 1150–. IEEE, 1999.
- K. Sandberg and M. Brega. Segmentation of thin structures in electron micrographs using orientation fields. *Journal of Structural Biology*, 157(2):403–415, 2007.
- Christian Schuldt, Ivan Laptev, and Barbara Caputo. Recognizing human actions: A local svm approach. In *ICPR*, 2004.
- S. Seung. Connectome: How the brain’s wiring makes us who we are. *Houghton Mifflin Harcourt*, 2012.
- Oded Shahar, Alon Faktor, and Michal Irani. Space-time super-resolution from a single video. In *CVPR*, pages 3353–3360, 2011.
- M. Shimano, T. Okabe, I. Sato, and Y. Sato. Video temporal super-resolution based on self-similarity. In *ACCV*, pages 93–106, 2010.