

3-D Density Kernel Estimation for Counting in Microscopy Image Volumes using 3-D Image Filters and Random Decision Trees

Dominic Waithe³(✉), Martin Hailstone⁴, Mukesh Kumar Lalwani², Richard Parton⁴, Lu Yang⁴, Roger Patient², Christian Egeling^{1,3}, Ilan Davis⁴

1 MRC Human Immunology Unit, 2 MRC Molecular Haematology Unit and 3 Wolfson Imaging Centre, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom.

`dominic.waithe@imm.ox.ac.uk`

4 Department of Biochemistry, University of Oxford, Oxford, United Kingdom.
The final publication is available at link.springer.com ECCV 2016 Paper ID 5

Abstract. We describe a means through which cells can be accurately counted in 3-D microscopy image data, using only weakly annotated images as input training material. We update an existing 2-D density kernel estimation approach into 3-D and we introduce novel 3-D features which encapsulate the 3-D neighbourhood surrounding each voxel. The proposed 3-D density kernel estimation (DKE-3-D) method, which utilises an ensemble of random decision trees, is computationally efficient and achieves state-of-the-art performance. DKE-3-D avoids the problem of discrete object identification and segmentation, common to many existing 3-D counting techniques, and we show that it outperforms other methods when quantification of densely packed and heterogeneous objects is desired. In this article we successfully apply the technique to two simulated and to two experimentally derived datasets and show that DKE-3-D has great potential in the biomedical sciences and any field where volumetric datasets are used.

Keywords: Density Kernel Estimation, 3-D, Random Decision Trees, Microscopy, Counting

1 Introduction

For over 30 years 3-D fluorescence light-microscopy has been used to visualise and investigate many aspects of cell biology and physiology. In recent years however, microscopy has seen unprecedented advances in resolution and speed allowing unobtrusive visualisation of live specimens at very high speeds. These microscopes which include highly sensitive confocal microscope, structured illumination, 3D-STED and light-sheet microscopes are capable of generating giga-bytes of information from individual experiments [1,2,3,4,5,6]. These systems create significant challenges in terms of image processing and analysis which remains a bottleneck for the effective use of data generated by these systems [7]. As a

consequence there is a real demand for combining intelligent analysis solutions which are powerful and computationally efficient. For this study images were acquired using a confocal microscope, but the technique could be equally applied to images taken on other microscopes.

Within the field of 3-D microscopy the most common approach to quantification of structures or objects within a specimen rely on segmentation of the global intensity histogram and often require a number of processing and filtering steps to isolate the cell-type or structure of interest [8,9]. These solutions require a skilled image analyst who is familiar with image processing techniques to process each image. From the medical imaging community discriminative and generative machine learning approaches have also been applied to segment organs in 3-D medical imaging of the body [10,11,12]. These approaches are advanced, yet do not tend to tackle the broad needs of the light microscopy imaging community which require algorithms that can be tuned to a number of similar yet discrete applications by relatively unskilled users.

The method proposed here is a discriminative approach utilising regression random decision trees to learn the association between features generated using basic image filters and a structured representation of each object within the training dataset. This approach follows on from the 2-D density approaches which were first applied to density estimation in biological images and also for estimating the number of individuals in a crowd [13,14,15,16,17,18,19]. The core strength of the density estimation approach is that it does not try to identify objects in their entirety but instead learns to associate pixel features with a particular density value. This approach means that if objects differ in morphology or appearance or are densely packed, the technique can still provide an accurate prediction of the number of objects present. The density estimation approach is perfect for application in microscopy due to the dense and heterogeneous nature of the objects being counted. The current state of the art for density estimation in 2-D is achieved through using Fully Convolutional Regression Networks (FCRN) which out-perform regression decision trees with sufficient training [19]. In this study we decided to use regression random decision trees however, due to their high speed and their top performance, especially with relatively small amounts of training. Processing time when analysing 3-D data is a key consideration and also, in biomedical applications, dataset sizes can be small and so optimum performance with a small amount of training is key advantage.

In previous density kernel estimation approaches, images features have been calculated from pixels using dense SIFT or an array of basic image filters which describe the pixel and its local environment in 2-D data [13,14]. We found that these filters could be applied to 3-D data by processing each slice of the image volume independently and through ensuring that the input density kernel used for the training was 3-D and not 2-D. We denote this adapted 2-D density kernel estimation algorithm as DKE-2-D. A more powerful approach however, that we developed during this study, was to use image filters that function in 3-D and aggregate information from the local 3-D environment into the feature description of each voxel, a technique we call DKE-3-D. For this method we have

developed an array of image filters which can process and aggregate the local 3-D voxel environment to provide a very rich description of each voxel within the 3-D image volume. Through applying these filters with our density kernel estimation method we have achieved a high level of accuracy. Our main contributions in this study are therefore, that we outline the theoretical basis of density kernel estimation in 3-D, we develop specialised 3-D filters for feature description, which out-perform comparable 2-D filters, and we validate our technique against competitive methods on four distinct 3-D volumetric datasets.

2 Method

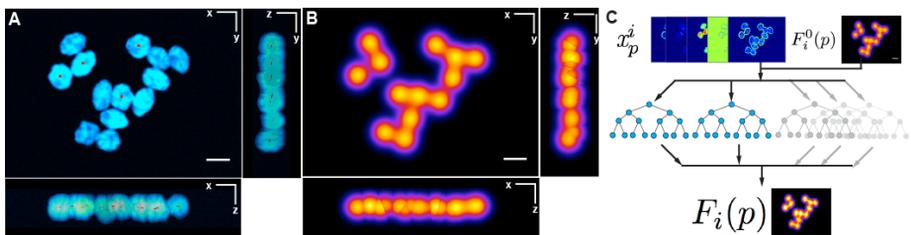


Fig. 1. Training scheme in 3-D. (A) Input image volume with dot annotation (red, exaggerated) in xy dimension and also orthogonal max projections (zy, right)(xz, bottom). (B) Ground-truth density function in xy dimension and also orthogonal max projections (zy, right)(xz, bottom). Scale bar is 40 pixels. (C) Schematic for data input and output with the ensemble of decision trees. Input features are calculated from input image volumes and associated with ground-truth density volume. Sampled voxel features and associated ground-truth voxels are then concatenated into a long vector (number of voxels, number of features) and used to train the ensemble of decision trees. Each decision tree receives a boot-strap sample of input voxels. At inference time, for a given voxel, the output density is calculated using the average output of each tree for that voxel to produce the predicted density image.

The training data is provided by the user in the form of N image volumes ($I_{i=1}, I_2, I_3, \dots, I_N$) with N corresponding annotations ($A_{i=1}, A_2, A_3, \dots, A_N$). A_i is a sparse matrix (\mathbf{R}^3) where the location of each cell in the corresponding image is marked with a single point (dot), by the user (Fig. 1). Each point in the annotation is stored in a vector of 3-D centroid positions $P_i = \{P_1, P_2, \dots, P_{C(i)}\}$ and C_i is the total cell count for a particular image. From each annotation volume in the training set we produce a ground-truth density representation which for each pixel (p) is defined as the sum of all the Gaussian kernels centred on the dot annotations:

$$\forall p \in I_i, F_i^0(p) = \sum_{\mu \in \mathbf{P}_i} \mathcal{N}(p; \mu, \sigma^2) \quad (1)$$

and $\sigma^2 = \text{diag}(\sigma_x^2, \sigma_y^2, \sigma_z^2)$. The kernel is anisotropic as the acquisition sampling resolution is often different in the axial z-dimension compared to the x-y lateral dimension. Figure 1 shows an example input image with dot annotations superimposed (Fig. 1A) and also the ground-truth density image calculated from these annotations (Fig. 1B). For this application we chose the sigma to represent approximately 1/3 the dimensions of the cells being counted which allows the whole kernel to be contained within an average cell, due to the so-called three-sigma rule of thumb [20]. For example if on average the cells being counted have dimension which is 30 x 30 x 10 pixels it would have a kernel of sigma 10 x 10 x 3 applied to it. Typically, the ground-truth density map is produced by convolving A_i with a 3-D Gaussian kernel of size σ . The integral of the ground-truth density map and the sum of the annotation volume can differ for a particular image when one or more of the cells are overlapping the periphery of the scene. It is a desirable consequence that the integral density be slightly reduced with respect to the annotations in this case, as we only want to consider the parts of cells within the scene. For each pixel in the input image, during training and testing, a corresponding feature vector is calculated, $x_i^p \in \mathbf{R}^{21}$ for DKE-2-D and $x_i^p \in \mathbf{R}^{25}$ for DKE-3-D. Each descriptor of the feature vector is created through processing of the input image or volume with one of a bank of filters which included: Gaussian, magnitude of Gaussian, Laplacian of Gaussian, eigenvalues of curvature, as in [14]. These filters are applied at multiple scales (sigma = 0.8, 1.6, 3.2 and 6.4) to aggregate data from the surrounding voxels into the feature descriptor at that pixel or voxel. This scale range was sufficient for the cases used in this study and were not changed. If the objects being counted are large with respect to the pixel resolution and smooth in appearance one should consider downsampling the images to improve the computational efficiency of the system. For the DKE-2-D algorithm, 2-D features were calculated. For the DKE-3-D algorithm, 3-D features were calculated, whereby the filters were calculated in 3-D across each volume and responses encoded as a feature vector for each pixel. The same filters were used as in the 2-D case but using 3-D implementations of each. The sigma range used was also the same as in the 2-D case using isotropic kernels throughout.

2.1 Decision tree framework

To solve the 3-D counting problem a machine learning density kernel estimation (DKE) approach was employed using an ensemble of decision trees [14]. The role of the DKE approach is to learn the non-linear mapping $\mathcal{F} : x_p^j \rightarrow F_i(p)$ which maps the input pixel features to the annotation derived ground-truth densities. Once optimised, this model would then allow us to predict the density values associated with a given pixel and allow estimation of the cell count for an entire image volume through summation of the individual densities. We chose

an ensemble of regression decision trees as our non-linear model [21]. Regression trees are simple binary trees which are built according to a classical top-down procedure[22]. In general, regression trees are fast to train, as split parameters are chosen randomly, and also they can handle large amounts of data as their complexity increases linearly with the number of data samples. Both of these factors are important for 3-D datasets where a typical image volume can contain more than 7 million pixels. From the top of the tree to the bottom, the data is split recursively into more and more specialised subsets based on the features and labels of the training data. Before generation of the first decision tree the training data from each image volume is concatenated into a single 2-D array (S) with dimension (num. of pixels x num. of features). The total volume of data is then reduced through random sampling by (1/200), as this was found to improve the speed of the fitting without deleterious affects on the accuracy of the algorithm due to pixel redundancy in each image. The recursive function which splits the data at each node, does so through selecting K candidate descriptors at random and with replacement from the available feature list. For each candidate, a threshold value is generated at random within the permissible min/max range for that feature $[a_{min}^s, a_{max}^s]$ producing a list of candidate splits s_1, \dots, s_K . Some decision tree types will enforce that multiple thresholds be considered for each feature selected, but in this case we kept the value as one, to keep the fitting efficient. A cost function (Score(s_i , S)) is calculated for each candidate split based on the subset of data at the node. The score used in this application was based on reducing the output variance in the labels associated with the left and right side of the proposed split:

$$- \sum_{p \in S_R} (F(p) - \bar{F}_R) - \sum_{p \in S_L} (F(p) - \bar{F}_L) \quad (2)$$

where S_L represents all the pixels which had values less than the proposed split point and S_R represents the complementary set within the superset S . \bar{F}^L and \bar{F}^R represent the mean value for all the pixel labels in the left and right side of the split respectively. From the K candidates the highest scoring feature and threshold which splits the accompanying density labels is chosen. The recursive splitting is repeated on the left and right node subsets until one of the stop conditions is fulfilled. Termination in this case was fulfilled when a decision tree reached a depth of 20, or there were only 20 samples left in a split. Each terminal node is then assigned the average density of the labels which descended into it. During inference, unlabelled image pixel features descend the trees based on the now fixed test functions at each split. The pixels will inherit the density value associated with the node in which their descent terminates. The generation of just one decision tree will tend to result in overfitting of data and so typically with random decision tree methods an ensemble of trees is generated from bootstrap samples of the input data. In this case 30 trees were generated during training. The average density value for a pixel, from all the generated trees, is then used to generate the output density label for that location.

2.2 Alternative Methods

As a means of comparison for our adapted approach (DKE-2-D) proposed approach (DKE-3-D) we compare our techniques to two current methodologies for 3-D counting of cells in 3-D volumetric data as well as to a competitive state-of-the-art technique known as FARSIGHT. The first comparative method is a generic segmentation strategy based on analysis of the image volume intensity histogram. For this, a Matlab script was developed which first thresholds the images automatically using the Otsu algorithm and then applies a 3-D watershedding algorithm to the image to split any clumped cells. The number of cells was then counted through Matlab’s in-built region-props functions [23] [24]. A second approach, which is a simple regression analysis was also developed using Matlab’s in-built poly function. Using 10 randomly selected training images, the relationship between the independent variable (the cell number in each image) and the dependent variable (the integral intensity) was learnt through linear regression and then inference was performed by querying unseen image intensities using this learnt distribution. Our third method of comparison represents the current state-of-the-art for counting cells in 3-D microscopy data which is a technique employing a morphological multi-model approach to segmentation and cell classification and has been shown to be highly accurate (FARSIGHT) [25,26]. The FARSIGHT algorithm required no additional parameters to perform segmentation and required no training.

3 Results

3.1 The Datasets

To establish the efficacy of the proposed technique we created/acquired four datasets and assessed the performance of the algorithm on each. The datasets and source code for this project are available through the website: http://github.com/dwaithe/Density_Kernel_Estimation_3D.

The first dataset (dataset 1) comprised of simulated 3-D image volumes which contain relatively few cells, with between 1 and 33 cells in each image. The simulation closely resembles cultures of HL60 cells with their nuclei stained with DAPI, which is a fluorescent DNA binding agent commonly applied in the life sciences for cell counting [27]. The assumption is that each cell only contains one nucleus, which is usually the case, and so can be used to directly estimate the number of cells present. The dataset contains 30 image volumes in total with dimension 404x283x65 and also included corresponding dot annotations. The texture and shape of the nuclei staining varies across each example and varies from cell-to-cell. The dimension of the input kernel applied at each dot in the image was set to $\sigma_x = \sigma_y = 12, \sigma_z = 13$.

The second dataset (dataset 2) has been designed to closely simulate colon tissue sampled from human patients suffering from adenocarcinoma, a type of cancer [27]. This dataset comprises 30 images volumes of dimension 325x258x65 and includes ground-truth data which exactly represents cell numbers present

in each scene. This dataset is challenging due to the density of cells within the scene and their close packing. Images were reduced in scale before processing as this was found to improve the computational speed without reducing the quality of the prediction. No other preprocessing was performed. The input kernel dimension applied at each dot in the image was set to $\sigma_x = \sigma_y = 6, \sigma_z = 10$.

The third test dataset (dataset 3) is experimentally derived and consists of 3-D confocal micrographs acquired from *Drosophila melanogaster* fly brains cultured for several days, fixed and then stained with DAPI nuclear dye, which stained the nuclei of the cells within the brain. 26 images were acquired on a Olympus Fluoview FV1000 microscope for this application with dimension varying between 512x512x15 and 512x512x33. The ground-truth for this data was created through manual identification of cell centres using commercially available Imaris software, due to its excellent 3-D visualisation capabilities, although any other software could have been used. The dimension of the kernel used to generate the ground-truth density image from the annotation image was set to $\sigma_x = \sigma_y = 15, \sigma_z = 4$.

The fourth dataset (dataset 4) represents deep 3-D volume imaging of the heart organ from intact Zebrafish embryos. In this study, a transgenic zebrafish line $Tg(myf7:EGFP; myf7:dsRednuc)$ was used which expresses red fluorescent protein (dsRed) in nuclei and enhanced green fluorescent protein (EGFP) in the cytoplasm of cardiomyocytes. The zebrafish embryos were stained for immunofluorescence so that the ventricular cardiomyocytes could be specifically counted. The red (dataset 4a) and green (dataset 4b) channels were imaged using a Zeiss confocal microscope and reconstructed independently to form two discrete image volumes varying in z-depth between specimen from 256x256x73 to 256x256x212. The red fluorescence staining is relatively punctate, representing the nuclei of the cells, whereas the cytoplasmic signal is much more diffuse and is continuous across the entire organ. For counting cells in the red nuclear fluorescence channel a sigma kernel of dimension $\sigma_x = \sigma_y = \sigma_z = 1.8$ was used. The ground-truth for this data was created through manual identification of cell centres using commercially available Imaris software. Cells could not be recognised from the green channel alone, due to the diffuse nature of the cytoplasmic staining, so for the basis of training on this dataset, the same cell positions and kernel sigma values from the red channel annotation were used.

To judge the performance of each of the algorithms, for datasets 1-3, a thorough cross-validation strategy was employed. For each comparison the algorithm was trained on 10 image volumes and then used to estimate the counts in 15 hold-out images, with the exception of the segmentation and FARSIGHT approaches which required no training. For the fourth dataset, which is a smaller dataset, training was performed on 8 image volumes and then the evaluation was performed on 2 hold-out images. Results for each assessment are described in terms of a percentage accuracy metric: $Acc.(%) = (1 - ((GT - PC)/GT)) * 100$, GT is the integral of the ground-truth density image and PC is the integral of the predicted density image. For each algorithm and dataset comparison, the training and test images were shuffled and the procedure repeated ten times

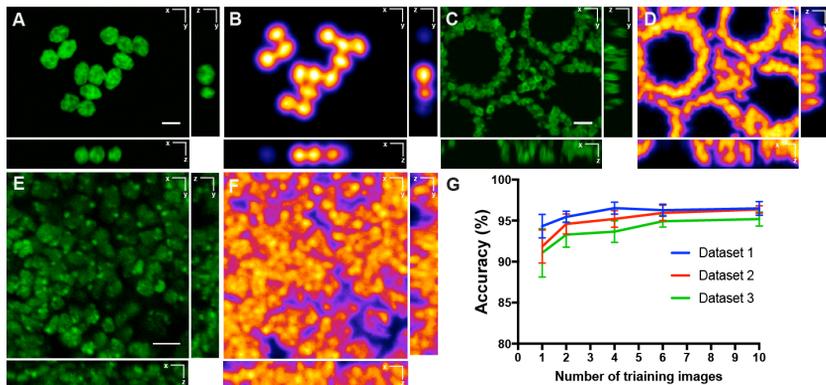


Fig. 2. DKE-3-D evaluation data. Example images (A) from dataset 1 with (B) density estimation, (C and D) dataset 2 and (E and F) dataset 3. Scale bar is 40 pixels in A and B and $6 \mu\text{m}$ in C. The depth of each stack is 65 pixels in A and B and $13.75 \mu\text{m}$ (25 pixels) in C. Orthogonal views show middle point of xy image in zy and xz plane. (D) The line plot shows the accuracy of the DKE-3-D algorithm when between 1 and 10 images are used in training. Error bars represent standard deviation of data.

with the average accuracy and its standard deviation being recorded for all the comparisons.

Table 1. Performance of algorithms. Performance is shown as percentage accuracy \pm standard deviation.

method	dataset 1	dataset 2	dataset 3	dataset 4a	dataset 4b
DKE-2-D	$93.9 \pm 1.3\%$	$93.1 \pm 1.3\%$	$95.4 \pm 0.8\%$	$73.3 \pm 24.9\%$	$88.5 \pm 8.1\%$
DKE-3-D	$96.5 \pm 0.8\%$	$96.4 \pm 0.5\%$	$95.2 \pm 0.9\%$	$85.6 \pm 10.8\%$	$84.0 \pm 16.1\%$
FARSIGHT	$96.1 \pm 2.0\%$	$81.2 \pm 3.0\%$	$< 0\%$	$82.8 \pm 3.5\%$	NA
Otsu Water.	$92.5 \pm 1.1\%$	46.2 ± 0.7	$< 0\%$	$65.0 \pm 3.4\%$	NA
Lin. Reg.	$94.3 \pm 1.1\%$	92.3 ± 1.2	$88.3 \pm 2.4\%$	$38.5 \pm 72.5\%$	$76.7 \pm 15.8\%$

The DKE-3-D approach proved to be a highly effective means of predicting the number of cells in each sample and out-performed the adapted DKE-2-D approach, the standard approaches (linear regression, Otsu watershed) and the FARSIGHT algorithm especially in dense scenes (dataset 2 and 3). Table 1 shows that the accuracy of the proposed algorithm (DKE-3-D) was better than the other approaches for dataset 1, although all the techniques performed well on this data. Dataset 1 represents a relatively sparse dataset with relatively few cells in each image and represents typical data for which a basic segmentation approach would normally be sufficient. Each technique achieved accuracy of over 92.5% with the DKE-3-D approach being slightly more accurate with an aver-

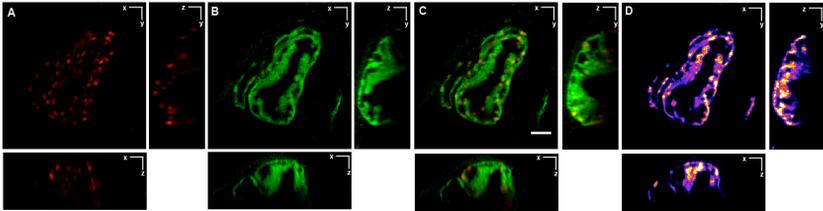


Fig. 3. Indirect density estimation. Example image from dataset 4 (A) in green channel (cytoplasmic stain) (B) in red channel (nuclear stain) (C) merged image of red and green channel. (D) Indirect predicted density output from model when using cytoplasmic stain for input features and ground-truth cell positions from red nuclear stain. Scale bar is $30 \mu\text{m}$, image volume is $117 \mu\text{m}$ (77 pixels) deep. Orthogonal views show middle point of xy image in zy and xz plane.

age accuracy of 96.5%. This shows that in relatively simple datasets, with few cells, it is probably sufficient to use a conventional approach like segmentation or linear regression, as top accuracy can be achieved. In datasets 2 the proposed DKE-3-D algorithm proved to be the most effective, out-performing all the other approaches including the DKE-2-D approach and achieved accuracy of 96.4% compared to 93.1% and 92.3% for the DKE-2-D and linear regression approaches respectively. The performance of the proposed algorithm is also evident in Dataset 3 where it out-performed the linear regression approach by over 7% and the segmentation approaches failed almost completely. Interestingly the DKE-2-D and DKE-3-D approaches performed equally well on this dataset probably because the data volumes were relatively shallow in this data ($\sigma_z=4$) and so aggregation of 3-D data was less impactful. The FARSIGHT algorithm performed well on the relatively sparse dataset 1 (96.1%) and respectably well on the denser dataset 2 (81.2%) but failed completely on dataset 3, showing that even cutting-edge segmentation strategies fail on data where cells are densely packed. Although slower, than both the linear regression and segmentation strategies the DKE-3-D is fast, training with 10 input images takes 5 s (dataset 1, 2.3 GHz Intel Core i7, 16 GB of RAM) and evaluation of a single image volume takes 10 s. The largest bottleneck for DKE-3-D is calculating the pixel features as this took around 40 s per image volume twice as long as it took for DKE-2-D (20 s), this could however be sped up considerably through parallelization. For the accuracy comparisons ten training images were used to train the DKE-3-D algorithm, but it is possible to train the algorithm with far fewer images and still achieve highly accurate results at test time. Figure 2D shows the average performance of the DKE-3-D algorithm when trained with between 1 and 10 images and evaluated on 15 holdout images as before. The DKE-3-D algorithm will after being trained with 2 images achieve an average 94.4% accuracy (Datasets 1-3) which is less than 2% less than the accuracy achieved with 10 training images 96.0%.

Dataset 4 represents challenging data for any of the tested counting systems (Figure 3). DKE-3-D performed best on this dataset however it is clear that the heterogeneity in the intensity of cells proved challenging for all the tested algorithms. Although DKE-3-D is able to handle a large diversity of object appearances, the accuracy will drop when there is too much diversity. This data was useful for showing the limits of the present approaches but also offered a unique opportunity, due to its two colour channels, to showcase a unique advantage of the DKE algorithm. Dataset 4a represents the red nuclear stain (Fig. 3A) which can be accurately annotated by a human, to label every cell in the heart tissue. When trained on this data the DKE-3-D algorithm achieved an accuracy of 85.6% which was better than any of the other tested techniques. Dataset 4b represents the cytoplasmic channel from the same image volumes as the red channel (Fig. 3B), a ground-truth cannot be made directly using the green channel due to the diffuse and overlapping nature of the stain. If however we use the ground-truth positions from Dataset 4a with the features calculated in Dataset 4b we can achieve a respectable 88.5 and 84.0% accuracy for DKE-2-D and DKE-3-D respectively (example output Fig. 4D). It is likely that the reason the DKE-2-D algorithm outperformed the DKE-3-D algorithm on this data because the diffuse nature of the cytoplasmic data means the 3-D features present in the DKE-3-D algorithm did not contribute much information. This indirect learning has great potential, whereby we can indirectly approximate cell counts based on stains which are not easily quantifiable by a human annotator. This means acquisition times and sample preparation times could be reduced as it would only be necessary to fully label and image samples for the training and then the technique could be applied to different potentially cheaper or more coarsely imaged representation. The only other technique, which can do this, is the simple linear regression approach but this technique achieved only 76.7% accuracy, due to most likely insufficient feature description, when compared to the DKE-3-D technique.

4 Conclusion and Limitations

The DKE method is a highly accurate and computationally efficient means through which 3-D objects with complex intensity distributions can be easily counted using only a dot annotation as training. This work shows that density kernel estimation is an invaluable approach for solving complex counting tasks in 3-D microscopy data. We make a significant step towards perfecting the technique through incorporating novel 3-D filters into the work-flow. The conventional DKE-2-D approach, adapted to 3-D, works relatively well on all data, but the proposed DKE-3-D algorithm with 3-D filters outperforms all other techniques, especially in dense environments where the boundaries of cells are not easily identified. This technique is suitable for counting any object which is roughly spheroidal or spherical in shape, but is limited, like other techniques, if the integral intensity between cells is wildly different. In summary, the DKE-3-D algorithm is a vital approach for tackling counting in 3-D volumes and due to

its unique density approach can out-perform conventional methodologies on a range of data.

5 Acknowledgements and Funding

We acknowledge the WIMM, The Dunn School of Pathology and the Biochemistry Department for infrastructure support. Authors are grateful to the staff of the Biomedical Services Unit at the John Radcliffe Hospital site for aquatic support. We thank the Wolfson Imaging Centre Oxford and to MICRON Oxford (<http://micronoxford.com>, supported by the Wellcome Trust Strategic Award 091911) for access to equipment and assistance with data acquisition and analysis. MKL and RP acknowledge funding from the BHF-Centre for Regenerative Medicine, Oxford-UK (grant ref RM/13/3/30159). The work was supported by the Wolfson Foundation, the Medical Research Council (MRC, grant number MC_UU_12010/unit programmes G0902418 and MC_UU_12025), MRC/BBSRC/EPSC (grant number MR/K01577X/1), and Wellcome Trust (grant ref 104924/14/Z/14). MH was supported through the ONBI DPhil programme in biomedical imaging technology development funded by the MRC and Engineering and Physical Sciences Research Council (EPSRC) (grant number EP/L016052/1). I.D. and R.M.P. were supported by a Wellcome Trust Senior Research Fellowship (081858) to I.D. LY was supported by a Clarendon Fund Scholarship in Humanities and by a Goodger fund Scholarship. DW was supported by funding from the MRC and EPSRC (grant number EP/L016052/1). None of the funding organisations have had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Streibl, N. Three-dimensional imaging by a microscope. *JOSA A* 2, 2 (1985), pp. 121-127.
2. Chen, B.-C., Legant, W. R., Wang, K., Shao, L., Milkie, D. E., Davidson, M. W., Jane-topoulos, C., Wu, X. S., Hammer, J. A., Liu, Z., et al. Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science* 346, 6208 (2014), pp. 1257998.
3. Reynaud, E. G., Krzic, U., Greger, K., and Stelzer, E. H. Light sheet-based fluorescence microscopy: More dimensions, more photons, and less photodamage. *HFSP journal* 2, 5 (2008), pp. 266-275.
4. Huang, B., Wang, W., Bates, M., and Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* 319, 5864 (2008), pp. 810-813.
5. Harke, B., Ullal, C. K., Keller, J., and Hell, S. W. Three-dimensional nanoscopy of colloidal crystals. *Nano letters* 8, 5 (2008), pp. 1309-1313.
6. Shao, L., Kner, P., Rego, E. H., and Gustafsson, M. G. Super-resolution 3d microscopy of live whole cells using structured illumination. *Nature methods* 8, 12 (2011), pp. 1044-1046.
7. Reynaud, E. G., Peychl, J., Huisken, J., and Tomancak, P. Guide to light-sheet microscopy for adventurous biologists. *Nature methods* 12, 1 (2015), pp. 30-34.

8. Long, F., Zhou, J., and Peng, H. Visualization and analysis of 3d microscopic images. *PLoS Comput Biol* 8, 6 (2012), e1002519-e1002519.
9. Peng, H., Bria, A., Zhou, Z., Iannello, G., and Long, F. Extensible visualization and analysis for multidimensional images using Vaa3D. *Nature protocols*, 9(1), (2014), pp. 193-208.
10. Cuingnet, R., Prevost, R., Lesage, D., Cohen, L. D., Mory, B., and Ardon, R. Automatic detection and segmentation of kidneys in 3d ct images using random forests. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*. Springer, (2012), pp. 66-74.
11. Lempitsky, V., Verhoek, M., Noble, J. A., and Blake, A. Random forest classification for automatic delineation of myocardium in real-time 3d echocardiography. In *Functional Imaging and Modeling of the Heart*. Springer, 2009, pp. 447-456.
12. Hu, S., Hoffman, E., Reinhardt, J. M., et al. Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images. *Medical Imaging, IEEE Transactions on* 20, 6 (2001), pp. 490-498.
13. Lempitsky, V., and Zisserman, A. Learning to count objects in images. In *Advances in Neural Information Processing Systems (2010)*, pp. 1324-1332.
14. Fiaschi, L., Nair, R., Koethe, U., Hamprecht, F., et al. Learning to count with regression forest and structured labels. In *Pattern Recognition (ICPR), 2012 21st International Conference on (2012)*, IEEE, pp. 2685-2688.
15. Arteta, C., Lempitsky, V., Noble, J. A., and Zisserman, A. Interactive object counting. In *Computer Vision ECCV 2014*. Springer, 2014, pp. 504-518.
16. Waithe, D., Rennert, P., Brostow, G., and Piper, M. D. Quantify: Robust trainable software for automated drosophila egg counting. *PLoS one* 10.5 (2015): e0127659.
17. Pham, V.-Q., Kozakaya, T., Yamaguchi, O., and Okada, R. Count forest: Co-voting uncertain number of targets using random forest for crowd density estimation. In *Proceedings of the IEEE International Conference on Computer Vision (2015)*, pp. 3253-3261.
18. Kainz, P. and Urschler, M. and Schultze, S. and Wohlhart, P. and Lepetit, V. You Should Use Regression to Detect Cells. *Medical Image Computing and Computer-Assisted Intervention—MICCAI (2015)*, pp. 276-283.
19. Xie, W. and Noble, J. A. and Zisserman, A. Microscopy Cell Counting with Fully Convolutional Regression Networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization MICCAI 1st Workshop on Deep Learning in Medical Image Analysis*, 2015.
20. Pukelsheim, F. The three sigma rule. *The American Statistician* 48, 2 (1994), pp. 88-91.
21. Breiman, L. Random forests. *Machine learning* 45, 1 (2001), pp. 5-32.
22. Geurts, P., Ernst, D., and Wehenkel, L. Extremely randomized trees. *Machine learning* 63, 1 (2006), pp. 3-42.
23. Otsu, N. A threshold selection method from gray-level histograms. *Automatica* 11, 285-296 (1975), pp. 23-27.
24. Meyer, F. Topographic distance and watershed lines. *Signal processing* 38, 1 (1994), pp. 113-125.
25. Schmitz, C. and Eastwood, B. S. and Tappan, S. J. and Glaser, J. R. and Peterson, D. A. and Hof, P. R. Current automated 3D cell detection methods are not a suitable replacement for manual stereologic cell counting. *Frontiers in neuroanatomy*. 8 (2014).
26. Lin, G. and Chawla, M. K. and Olson, K. and Barnes, C. A. and Guzowski, J. F. and Bjornsson, C. and Shain, W. and Roysam, B. A multi-model approach to

simultaneous segmentation and classification of heterogeneous populations of cell nuclei in 3D confocal microscope images. *Cytometry Part A* 71, 9 (2007), pp. 724-736.

27. Svoboda, D., Homola, O., and Stejskal, S. Generation of 3d digital phantoms of colon tissue. *Image Analysis and Recognition* (2011), pp. 31-39.