Pattern Recognition Letters 43 (2014) 17–24

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Synapse classification and localization in Electron Micrographs

Vignesh Jagadeesh [a,*], James Anderson [b], Bryan Jones [b], Robert Marc [b], Steven Fisher [c], B.S. Manjunath [a,1]

[a] Department of Electrical and Computer Engineering, University of California, Santa Barbara 93106, United States
[b] Moran Eye Center, University of Utah, Salt Lake City, UT 84132, United States
[c] Department of Molecular, Cellular and Developmental Biology, University of California, Santa Barbara 93106, United States

ARTICLE INFO

Article history:
Available online 18 July 2013

Communicated by H. Bunke

Keywords:
Connectomics
Detection
Feature descriptors
Electron Micrographs

ABSTRACT

Classification and detection of biological structures in Electron Micrographs (EM) is a relatively new large scale image analysis problem. The primary challenges are in modeling diverse visual characteristics and development of scalable techniques. In this paper we propose novel methods for synapse detection and localization, an important problem in connectomics. We first propose an attribute based descriptor for characterizing synaptic junctions. These descriptors are task specific, low dimensional and can be scaled across large image sizes. Subsequently, techniques for fast localization of these junctions are proposed. Experimental results on images acquired from a mammalian retinal tissue compare favorably with state of the art descriptors used for object detection.

1. Introduction

Visual classification of structures of interest has a wide variety of applications in natural images, video sequences, aerial and biological images. At one end of the spectrum, detection and classification of objects in natural images has received a significant research interest in recent times with competitions like PASCAL and ImageNet. Algorithms such as the DPM and Sparselet have been shown to perform extremely well on such challenges. At the other end of the spectrum are emerging applications in bio-microscopic imagery, where automated image analysis is crucial due to high throughput image acquisition. Constructing an overarching classification/detection model that can work across any bio-microscopic imagery is challenging due to inherent variability in imaging protocols. For instance, a tissue imaged using different imaging conditions, such as the light, confocal or electron microscopy, can lead to visually very different images. Knowledge of associated meta-data such as molecule specific bio-markers used for imaging are critical for further processing and interpretation of such images. As a result, an algorithm developed for one modality is difficult to adopt to another modality, necessitating the development of application specific classification/detection algorithms. The scope and applications to problems in bio-microscopic imagery are fairly diverse, with many applications still relatively unexplored. We focus on one such application, namely structural connectomics.

Connectomics: Connectomics is a sub-field of neuroscience aiming to understand neuronal circuitry in the animal brain. Synapses, or edges in the neuronal circuit graph can be resolved only at nanometer (10^{-9} m) resolutions. Such resolutions require the acquisition of massive amounts of data, typically ranging into several terabytes. Due to recent developments in high throughput microscopy, such datasets can be acquired in a fully automated fashion without any human intervention. The bottleneck manifests in analyzing these large image mosaics, which could take human annotators several man years. In attempting to develop fully automated image analyzers, two main issues arise. Firstly, the low level visual features that would work best are unknown beforehand, and considerable effort is required to uncover features that work reliably. Secondly, the feature extractors and classifiers must be scalable to the size of datasets considered.

In this paper, we focus our attention on the problem of visually interpreting Electron Micrographs (EM). An example of Electron Micrograph imagery is shown in Fig. 1. These high resolution EM images tend to be highly textured and require expert interpretation in identifying cellular and sub-cellular structures of interest.

Scale of Data Considered: The dataset of interest in this paper, also referred to as the RC1 connectome is acquired from a rabbit's retinal tissue. It is physically 33 μm thick and has a diameter of 25 mm. The imaging is performed at an x–y resolution of 2.18 nm, and a z-resolution is 70 nm. As a result, the data is highly anisotropic, meaning that the sampling across z-direction is much coarser than sampling on the x–y direction. In other words, thin sections of the tissue are successively imaged with a z-spacing of about 70 nm between adjacent slices. A total of 341 z-slices are acquired using high throughput microscopy, leading to the creation

* Corresponding author. Tel.: +1 (805) 893 2526; fax: +1 (805) 893 3262.
E-mail address: vignesh@ece.ucsb.edu (V. Jagadeesh).
[1] Tel.: +1 (805) 893 2526; fax: +1 (805) 893 3262.

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.06.001

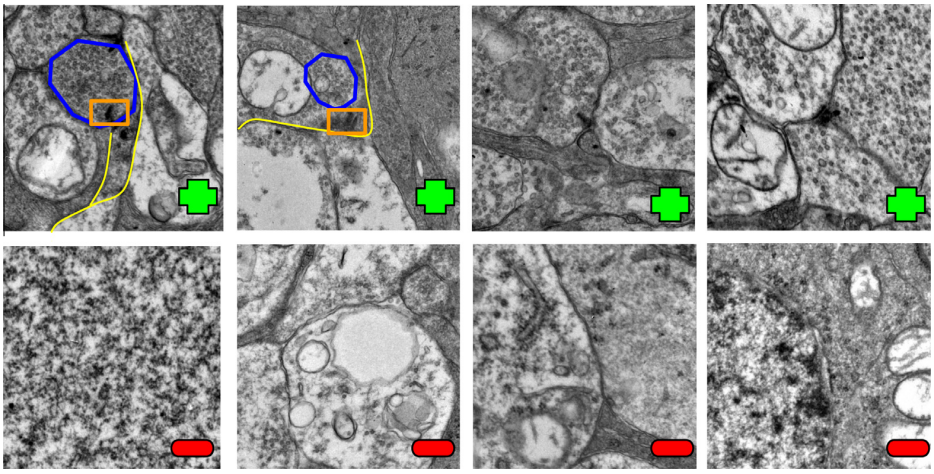


Fig. 1. The first row illustrates examples of synaptic junctions characterized by vesicles (blue contours), cell membrane (yellow contour) and ribbons (orange box). The second row of images illustrate negative examples which do not contain synaptic junctions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of a 3D image stack where the z dimension ranges from 1 to 341. Further, the raw data is stored as a multi resolution volume, comprising a total of six pyramid levels. The total storage requirements for the multi resolution volume alone is about 15 terabytes. Hence, even storing, accessing and handling the entire dataset is a major challenge to begin with. The scalable Viking viewer (Anderson et al., 2011) elegantly solves this problem by providing an interface to interact, annotate and study the data.

Each z-slice of the connectome volume comprises around 250,000 tiles of dimension 256×256. The entire connectome volume would then comprise a total of 90 million such tiles of dimension 256×256. A single slice from the connectome, after mosaicking has 125000×125000 pixels. Processing such large datasets require distributed computing infrastructures, and the availability of algorithms that can be parallelized and scaled across a large number of computing nodes. We focus our attention on the latter issue of developing scalable algorithms that solve an important problem of synapse localization in Electron Micrographs, see Fig. 2.

Significance of Synapses: Synapses are structures in the brain that help neurons in communicating with one another using chemicals known as neurotransmitters. Vesicles are the carriers of neurotransmitters, that are transmitted from one cell to another. Since vesicles are spherical in 3D, they have a circular shape when projected onto a 2D plane and imaged. The junction where communication between neurons happen is the cell membrane, also referred to as clefts. Vesicles and clefts co-occur in any type of chemical synaptic junction. Further, in some classes of synapses one can observe electron dense black regions referred to as ribbons near the cleft. Such synapses are known as ribbon synapses. The co-occurrence of three structures, namely vesicles, clefts and ribbons are often used by biologists to detect the presence/absence of a synaptic junctions. We also refer to the three structures (vesicles, clefts, ribbons) as semantic attributes that are the building blocks in constituting a synapse.

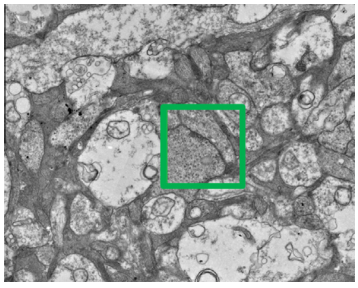


Fig. 2. Illustration of the Detection task. The aim here is to isolate synapses (green boxes) from the rest of structures in the image. The image comprises only a single channel, and has structures surrounding the synapse with similar visual properties, making the detection problem very challenging. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The primary objective of this work is to build a robust and lightweight feature descriptor for identifying synaptic junctions in large EM mosaics.

Strong Biological Priors: Object detection refers to the problem of identifying the spatial location of an object of interest in images. State of the art methods can detect faces and people with impressive accuracy. However, detection of generic object categories is still an open area of research which is being addressed in computer vision competitions such as PASCAL VOC. In most of the detection systems that work reliably, some form of gradient features based on histogram of oriented gradients are trained with linear SVMs for detection. Extensions based on deformable part models where latent part configurations are inferred during training are widely used. The deformable part model does have its share of disadvantages such as reduced accuracy on object classes without articulate parts, and considerably higher training time involved in learning.

Images for generic object detection can come in any scale/rotation/shear/clutter, while the target class (say a person) remains consistent across all images, thus necessitating rich feature sets and object localizers. In contrast, biological image datasets have highly constrained imaging protocols which are known in advance, as well as biological priors on shape and size of objects of interest. As a result, we propose to focus on exploiting strong constraints on imaging and biological knowledge to construct efficient and simple detectors. We focus on the problem of synapse detection for illustrating the usefulness of exploiting strong prior knowledge available in bioimaging scenarios.

Paper Organization: The rest of the paper is organized as follows. Section 2 presents the design of attribute based feature descriptors for synaptic patches, and their classification based on fusion of features. Section 3 discusses extension of classification to a localization framework. Section 4 presents comprehensive experimental

validation on individual and fused attributes, followed by conclusions and a discussion of future work in Section 5.

2. Synapse classification

We first consider the problem of classifying synaptic regions from non-synaptic regions. Example image patches in which synapses are present/absent are shown in Fig. 3. Given the typical wide intra-class variability of the images, it is often a challenging classification problem even for a trained human expert. This problem is in stark contrast to typical detection of common objects such as faces, humans, cars, and buildings, where the objects are immediately discernible and well defined. As noted earlier, a key observation is that the synaptic junctions are characterized by the co-occurrence of image primitives corresponding to vesicles, clefts and ribbons. The same intuition is used by the human experts while annotating such regions, and our feature fusion method described below builds upon this observation.

The work by Kreshuk et al. (2011) is closest in spirit to the proposed approach, with some important differences. Firstly, the technique in Kreshuk et al. (2011) was pixel based and does not model contextual cues. In contrast, the proposed approach models contextual (co-occurrence) cues, and classifies regions instead of pixels. Intuitively, it would be easier to provide patch exemplars to train a system (proposed), in contrast to precise user strokes on synapses. More importantly, the proposed approach extends the idea of pixel level features to regions by modeling the distribution of filter responses as a spatial pyramid. Further there have been work on detecting structures like mitochondria (Lucchi et al., 2010) or vesicles (Dıaz et al., 2010) from EM images. While these structures form closed contours and can be associated with some notion of shape, synaptic junctions considered in this work are regions comprising a mixture of structures (clefts,vesicles, ribbons) devoid of an intuitive shape representation. The works of Kaynig et al. (2010) and Jagadeesh et al. (2011) are other sources of reference for EM image analysis.

The proposed method is illustrated in Fig. 3. Initially, the problem of finding representations for vesicles, ribbons and clefts is addressed. Subsequently, a technique for fusing the above cues is discussed.

2.1. Vesicular features: spatial pyramids of matched filtering

Vesicles are spherical structures in three dimensions, and appear as circular structures when sliced across the z dimension. As illustrated by the blue contour in Fig. 3, vesicles are small circular blobs that are usually clustered in space. Hence, a detector of vesicles must be shape aware (circularity) and quantify spatial clustering behavior of detected vesicles. We propose using a filter matched to the size of vesicles observed during an offline training phase. Fig. 4(a) illustrates the result of convolving a vesicular patch with the matched filter. The vesicular blobs have a stronger filter response (in red) since they are matched to the filter, in comparison to the background. Performing a peak detection on the filtered response (black spots overlaid on Fig. 4(a)) illustrate detection of the centers of circles in the x–y plane. The analytical form of the LoG kernel is given by, $G(x,y) = -\frac{1}{\pi\sigma^4}\left[1-\frac{x^2+y^2}{2\sigma^2}\right]\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$. However, vesicles do not occur in isolation but with a lot of other subcellular structures which may be distracting in the filter responses. As a result, the local distribution of detected centers are computed to quantify the spatial clustering of vesicles. In order to make the representation robust, a spatial pyramid approach (Fig. 4(b)) is proposed. The basic idea is to capture the spatial distribution of feature points in addition to the magnitude of responses themselves. This is accomplished by successively subdividing an image into non overlapping sub-regions and accumulating strengths of filter responses that serve as feature vectors for classification. The subtle difference with existing spatial pyramid matching techniques is the absence of a pre-computed codebook (Lazebnik et al., 2006). In contrast, this work preserves the intuition that distribution of feature points be characterized along with strength of filter responses without an explicit codebook.

The vesicle descriptor construction is now explained in further detail. Assuming the image to be denoted by $I \in [0,1,\ldots,255]^{M\times N}$, the matched filter response for vesicles denoted by F_v for every pixel p is obtained as:

$$F_v(p) = I(p) * G(p) \tag{1}$$

Since one is interested in the peaks of filter responses to locate circular structures, we now define a binary vesicle interest map B_v that estimates locations of vesicle centroids, with respect to 8-connected neighborhoods $\mathcal{N}_p$ of a pixel p:

$$B_v = \mathcal{I}_{F_v(p)>F_v(q)}, \quad \forall q \in \mathcal{N}_p \tag{2}$$

where $\mathcal{I}_\psi = 1$, if $\psi = 1$ and is 0 otherwise. As described above, vesicles are spatially clustered in electron micrograph stacks, and any patch sampled should have vesicles that are uniformly distributed in space. In order to computationally capture this notion, we utilize spatial pyramids. For a spatial pyramid decomposition, the image is partitioned into multiple non-overlapping regions. Let us denote the binary mask corresponding to such a partition j in pyramid level i to be $Q_i^j \in \mathbb{B}^{M\times N}$, we obtain the descriptor defined over the set of pixels $\mathcal{P}$ constituting the image to be,

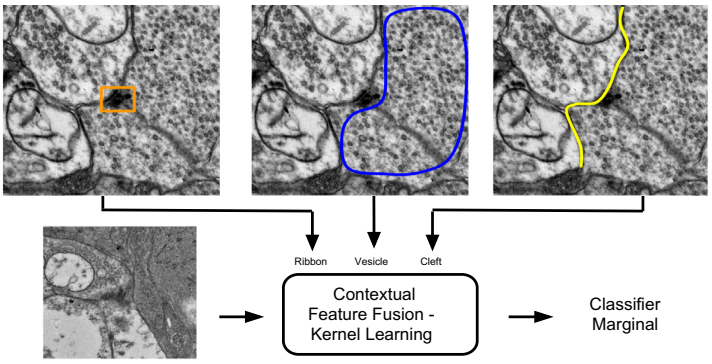
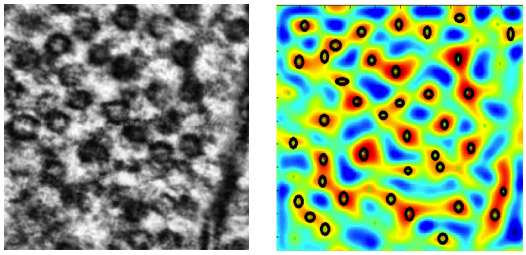
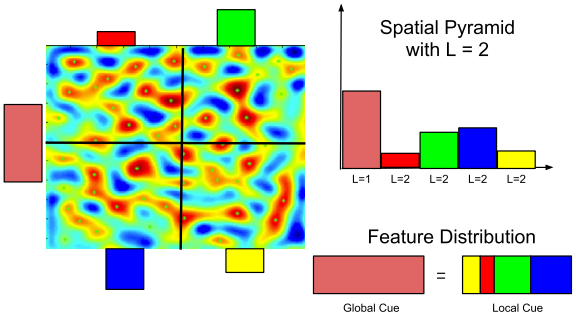


Fig. 3. An illustration of the multi attribute fusion procedure. The co-occurence of multiple attributes, namely ribbons (orange bounding box), vesicles (blue contour), and clefts (yellow contour) are described using task specific features. Subsequently, these features are fused using kernel learning to yield confidence measures for the existence of a synaptic junction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Convolution with a Matched Filter and Peak Finding for Vesicle Description



(b) Construction of Pyramidal Matched Filter responses.

Fig. 4. Matched filtering for vesicle description and pyramidal feature construction.

$$R_v(2^i + j) = \sum_{p \in \mathcal{P}} \left(F_v \odot B_v \odot Q_i^j \right)(p) \tag{3}$$

where $\odot$ refers to an element by element matrix multiplication. In our experiments, $i = 1$, leading to a 5-dimensional vesicle descriptor.

2.2. Cleft features: steered second order Gaussian derivatives

Clefts, as illustrated by the yellow open contour in Fig. 3 is the cell membrane separating two neuronal structures. It is well known that synapses occur only at locations where clefts are present. They have a characteristic ridge profile (see Fig. 5) and we propose using a second order Gaussian derivative filter to detect ridges. Since the clefts could occur at any orientation, a steered Gaussian derivative filter bank is utilized to detect despite orientation variability. In other words, the filter bank is constructed by rotating the Gaussian derivative kernel by uniformly sampling 0–180 degrees (see Fig. 6), with the set of angles contained in Θ. Denoting the image by I, second order Gaussian derivative filter at an angle θ by u''_θ, and the filter response at pixel p by $H_\theta(p) = I(p) * u''_\theta$, we compute response at every pixel to be $F_c(p) = \max_\theta H_\theta(p),\ \theta \in \Theta$. In other words, the input image is convolved with rotated second order derivative Gaussian filters, and the maximum response at each pixel is selected. The resulting response is binarized to obtain the cleft interest map over the image:

$$B_c(p) = \mathcal{I}_{I(p) > T_I} \tag{4}$$

where T_I is a threshold estimated form data using Otsu's method (1975). It is important to note that the goal of this procedure is not to perform accurate cleft segmentation. Instead, the idea is to accumulate evidence of contiguous ridge strength at the vicinity of vesicles. As illustrated in Fig. 5(a) the green contours capture a majority of clefts (there could be missing detections where signal strength is weak) without capturing any background clutter. A spatial pyramid histogram is constructed to represent clefts in a manner similar to the spatial pyramid description of vesicles. The individual features of the spatial pyramid representation are formed by accumulating the strengths of filter responses in the region of interest. Denoting the binary mask corresponding to such a partition j in pyramid level i to be $Q_i^j \in B^{M \times N}$, we obtain the descriptor defined over the set of pixels $\mathcal{P}$ constituting the image to be,

$$R_c(2^i + j) = \sum_{p \in \mathcal{P}} \left(F_c \odot B_c \odot Q_i^j \right)(p) \tag{5}$$

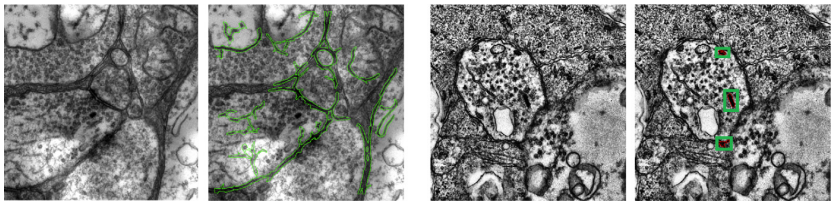
In our experiments, we set $i = 1$, obtaining a cleft feature descriptor that has 5 dimensions. The problem of validating clefts is tricky since clefts-like features are present in abundance throughout the connectome. Hence they are not very discriminative, but work in conjunction with vesicular features to reduce false positives, as will become evident later.

2.3. Ribbon features: region stability and shape

Ribbons are electron dense regions (see Fig. 5) that occur in some special kinds of synapses called ribbon synapses. They have a characteristic black appearance and close to an elliptical shape. In order to detect these structures the input image is adaptively equalized using Contrast Limited Adaptive Histogram Equalization (CLAHE) (Zuiderveld, 1994) to get an image denoted as I_{clahe}. We subsequently define a ribbon interest map by,

$$B_r(p) = \mathcal{I}_{I_{clahe}(p) > T_{I_{clahe}}} \tag{6}$$

where $T_{I_{clahe}}$ is a threshold estimated from the data. Ribbon detections are generated by using connected components analysis (CCA) on the ribbon interest map B_r. Further, let $CCA_i(B_r)$ denote a function that returns pixels corresponding to the ith largest



(a) Image comprising clefts (left), Cleft Detections (right)

(b) Image comprising ribbons (left), Ribbon Detections (right)

Fig. 5. Illustration of cleft and ribbon detection.

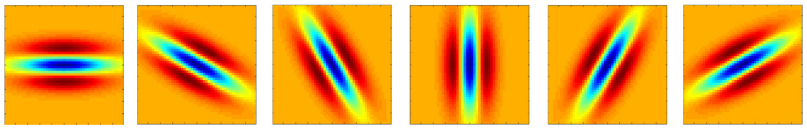


Fig. 6. Steered second order derivative of Gaussian filters at six orientations from 0–180 degrees.

connected component in B_r. The features for ribbon presence in a patch are size of the three ribbon candidates normalized by the image size,

$$RR_r(i) = \frac{|CCA_i(B_r)|}{|I|}, \quad 1 \leqslant i \leqslant K \tag{7}$$

where $|\cdot|$ denotes the cardinality of a set. In our experiments, $K = 3$ leading to a 3-dimensional descriptor for ribbons. Ribbons are known to be near elliptical in shape, a factor that can be used in refining ribbon detection. Denoting the covariance matrix of the binarized region corresponding to the ribbon, $C = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix}$ denote by σ_{c1}, σ_{c2} the eigenvalues corresponding to the covariance matrix. The ratio $s_{ribbon} = \sigma_{c1}/\sigma_{c2}$ is utilized for ascertaining the ellipticity of the shape, and the validity of a ribbon detection.

2.4. Fusion classifier for detection

The three different features correspond to vesicles (R_v, 5-dimensional), cleft (R_c, 5-dimensional) and ribbon (R_r, 3-dimensional) features. Let the above features corresponding to the three attributes be denoted by, f_j, $1 \leqslant j \leqslant 3$.

We now propose using two simple and intuitive fusion schemes. The aim is to learn a single discriminant function (Varma and Babu, 2009), $E(f^R) = \sum_{i=1}^{M} y_i \alpha_i K(f^R, f^i)$, where K is a kernel function measuring dissimilarity between the M support vectors f^i (formed by concatenating f_{js} for data point i) with weights α_i, and the input feature f^R. Specifically, we utilize the radial basis function for measuring dissimilarity in data points.

Early Fusion/Naive Learning works on feature vector concatenation. For every image all attribute features are extracted, and vectorized to a long column vector, which is the fused representation. The features are then mean normalized and scaled. Subsequently, a support vector machine (Chang and Lin, 2011) is trained over the input feature vectors to classify synapses in test data. Similarly, an online boosting algorithm is also utilized over the input features for synapse classification.

Late Fusion/Multi Kernel Learner works on the premise that not all feature dimensions are equally discriminative. In other words, a feature subset selector works along with the base SVM to weight discriminative feature channels more (Varma and Babu, 2009). The kernel function is where the actual fusion takes place, and is defined by a linear combination of a set of base kernels, $K(f^R, f^i) = \sum_{js} d_{js} K(f^R_{js}, f^i_{js})$, where j and s iterate over the different features, along with their scales respectively. K is the kernel function between the estimated support vector f^i and the input feature f^R. The variable K is a positive definite kernel and i indexes a set of M candidate support vectors chosen by the SVM. The above function can be interpreted as a kernel that can be further decomposed to a set of base kernels with weights d_{js}. Minimizing the MKL cost function yields a set of classifier weights across the different kernels, that is used during test time. The reader is referred to Varma and Babu (2009) for a more detailed treatment of MKL. The weights are learnt with respect to the different feature spaces indexed by j and the scale s at which features are extracted.

Algorithm 1. Procedure for Fast Interest Point Localization for Synapses

Require: I (the image), th (threshold), CC (connected components)
Ensure: $0 \leqslant th \leqslant 1$
  Normalize the Image Intensity $I = \frac{I - I_{mean}}{I_{max}}$
  Threshold image $I_T = T(I, th)$, where the threshold $th = 0.2$
  if TRAIN **then**
    Extract connected components $\mathcal{K}_{train} = CC(I_T)$
    Training: Learn $Pr(\mathcal{K}) = \mathcal{N}(\mu_{size}, \sigma_{size}) \leftarrow \mathcal{K}$ from the set of connected components
  else
    Extract connected components $\mathcal{K}_{test} = CC(I_T)$
    Evaluate $\mathcal{K}_{test}$ on the pre-learnt $Pr(\mathcal{K})$ to accept or reject a hypothesis
  end if

3. Synapse localization

The previous section discussed modeling the visual characteristics of a synaptic patch. We now turn our attention to localizing these regions effectively in very large mosaics.

For this purpose, we design an interest point detector similar in spirit to a Maximally Stable Extremal Region (MSER) of Matas et al.

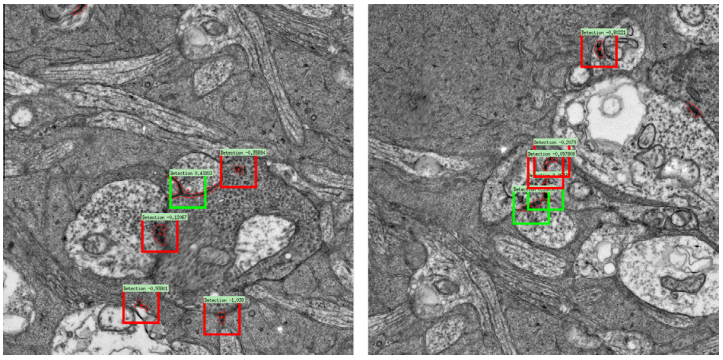


Fig. 7. Examples of synapse localization/detection on 2500 × 2500 image mosaics. The bounding boxes indicate locations that the algorithm places a high likelihood for the existence of a synaptic junction. Further, green boxes are locations where the algorithm is most confident, followed by the regions highlighted by red boxes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(2002). Our main observation is that of exploiting a region around the cleft that has an invariant black ridge profile. We propose the procedure in Algorithm 1 for efficient localization. This procedure is used to generate candidate locations (interest points) where the synapse attribute classifiers explained in the previous section are applied on.

Fig. 7 illustrates an example of synapse detection on a large 2500 × 2500 image mosaic. Detection using traditional scanning window techniques would take considerable time (order of minutes in our implementation) for detection. In contrast, our proposed technique can detect/classify synaptic junctions in a few seconds. In Fig. 7 the green bounding indicate regions where the algorithm is most confident of its detections, followed by the red bounding boxes. In the dataset considered, it is infeasible to evaluate accurate precision values because much of the dataset does not contain any annotations. As a result, any ground truth sampled from existing annotations would be partial, since many true positives are yet to be annotated. The PASCAL Average Precision score (Everingham et al., 2010), a standard procedure to validate detection does not apply here. However, the proposed detector achieves a recall of 0.85 on a validation set. This means that 85% of synapses that were annotated are successfully identified.

Fig. 8 illustrates a scenario where the proposed technique could be improved upon. The image shows multiple bounding boxes that are in the immediate vicinity of each other. A robust non-maxima suppression model would aid in achieving more accurate detections, and is a subject of future work.

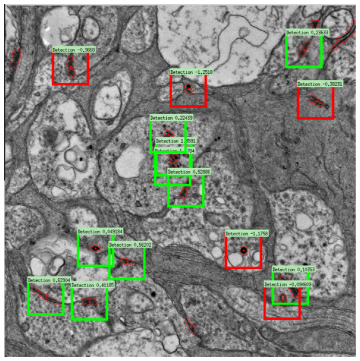


Fig. 8. Examples of synapse localization/detection on 2500 × 2500 image mosaics. A scenario where the proposed technique could be improved, with stronger non-maxima suppression models.

4. Experimental validation

Experimental results are first reported on the efficacy of each feature separately in a task specific manner, see Fig. 9(a)–(c). Subsequently experimental results are reported on the effectiveness of feature fusion, and the synapse detection task on a large dataset. Since the datasets considered have equal distribution of positive and negative samples, the detections are validated by their accuracy.

Experiments reported in Fig. 9 are performed on data from the RC1 retinal connectome. In order to understand each attribute's performance, they are initially validated separately, each on a dataset of 200 images (100 train/100 test) to identify relative attribute strengths. Subsequently, the fused description of three attributes is validated on the synapse classification task.

The results obtained using the 5-dimensional vesicle feature descriptor trained on an SVM with stratified ten fold cross validation is given in Fig. 9(a). Similar results for the individual cleft and ribbon descriptors are given in Fig. 9(b) and (c) respectively.

The previous sets of validation are on controlled data where the properties of individual attributes could be studied in greater detail. To test the efficacy of fusion, a dataset of 200 images with equal split of positive and negative samples are considered, with 50 percent of data used for training and the rest for testing. The prediction accuracy is reported in Fig. 9(d). Initially the task specific features alone are utilized for classification, followed by the fusion of all three structural features. As can be observed, the fusion yields much better results than any of individual features. Surprisingly, the naive learner gives almost the same performance as the multiple kernel learner as evident from Fig. 9(d).

Validating Synapse Classification: The next set of experiments deal with the task of synapse classification. In this task, a database of 2000 images with equal distribution of positive and negative samples (1024 × 1024) is used for validation. The experiments compare performance of texture descriptors based on Gabor Wavelets (Manjunath and Ma, 1996) (GW), Texton (VZ) filter bank energies (Varma and Zisserman, 2003), Local Binary Patterns (LBP) (Ojala et al., 2002), Spatial Envelope (GIST) (Oliva and Torralba, 2001) with the proposed approach. The proposed approach attains the highest performance by a fair margin. In summary, the proposed descriptor is simple, task specific and low dimensional, in comparison to texture descriptors that are fairly generic and of high dimensionality. Further the proposed method has better classification rate, and has semantic attributes associated with it, and is much faster to compute due to lesser convolution operations. The proposed approach can individually characterize vesicular density, cleft strength and ribbon sizes by their corresponding

Vesicle Detection	Positive Samples	Negative Samples
Positive Samples	**0.92**	0.08
Negative Samples	0.11	**0.89**

(a) Cross Validation on Control Set yields 90.5%

Cleft Detection	Positive Samples	Negative Samples
Positive Samples	**0.74**	0.26
Negative Samples	0.23	**0.77**

(b) Cross Validation on Control Set yields 76%

Ribbon Detection	Positive Sample	Negative Sample
Positive Samples	**0.90**	0.10
Negative Samples	0.12	**0.88**

(c) Cross Validation on Control Set yields 90%

Syn. Detect	Fused Feat.	Vesicle Feat.	Ribbon Feat.	Cleft Feat.
Naive	**82.29**	73.96	**71.88**	58.33
MKL	81.25	**75.00**	66.67	**60.42**
Boosting	73.96	57.29	70.83	52.08

(d) Performance of Fused Features along with Individual Structure Features

Syn. Detect	Gabor Wavelets	Texton Energy	LBP	GIST	Proposed
Det. Acc.	75.25	76.42	69.77	75.28	**81.25**

(e) Comparison to state of the art descriptors on a database of 2000 1024×1024 images

Fig. 9. Experimental validation of fusion and large scale experiments.

attributes, an aspect that generic features cannot provide. Note that a direct comparison with (Kreshuk et al., 2011) is not possible since the data used are inherently very different.Their most informative features resembling cleft features only yield 60% accuracy in the detection task.

Since the image sizes are large, the time taken for feature computation on a single node is substantial. For the work on classification, features are precomputed on a distributed computing environment with a 32 node cluster. The total time take on the cluster for computing the six sets of features was close to 12 h.

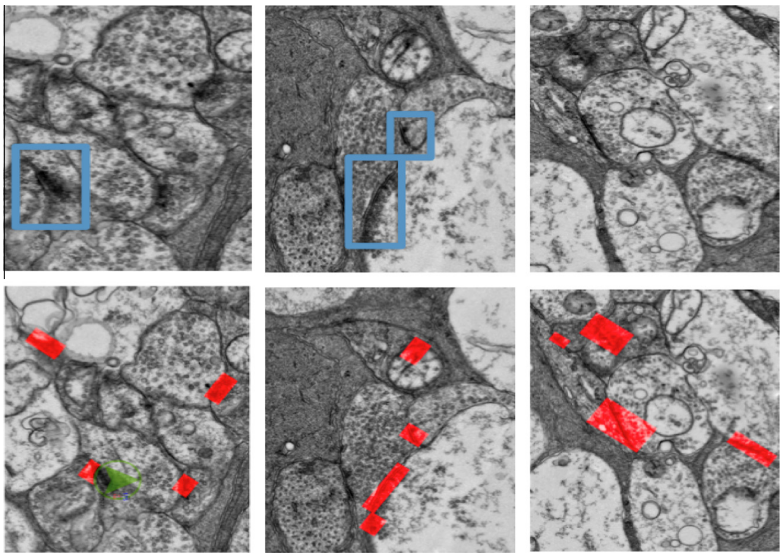


Fig. 10. The blue boxes are regions where the biologists have marked the presence of a synaptic junction, and the red transparencies are detections by the algorithm. As can be observed, not all locations where the algorithm has generated detections without corresponding human ground truth can be marked as false positives. As a result, it is necessary that a human in the loop validate the detection results. Strategies for efficiently utilizing an annotator's time through active learning is part of future work. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.1. Large scale localization experiments

As mentioned before, the dataset considered comprises large mosaics and it is of interest to investigate scalability aspects of the proposed technique. In order to test the same, we select a single z-slice of the connectome volume that comprises 250,000 tiles of size 256×256. We parallelize the algorithm on a 28 node cluster which takes about 100 min to process the entire slice that occupies about a hundred gigabytes of memory.

Fig. 10 is an example of large scale detection applied to the connectome data. The blue bounding boxes indicate the ground truth marked by biologists, while the red transparencies are results generated by the proposed algorithm. As can be observed, all locations where the algorithms has generated results without corresponding human annotations cannot be flagged as false positives. This is because the original data has not been fully annotated. The best course of action in these cases would be active learning, where a user in the loop is consistently queried with example detections to validate the algorithm.

5. Conclusions

This paper explored the problem of synapse detection in Electron Micrographs by addressing the problem of attribute based synapse description and fast synapse localization. Firstly, a novel attribute based synaptic junction descriptor that models the visual characteristics of clefts, ribbons and vesicles was presented. Subsequently, techniques for fast synapse localization was proposed, with validation on large image mosaics. Future work includes the use of active learning based techniques for interactive querying and corrections from the end user.

Acknowledgment

This work was funded by Grant NSF-OIA 0941717.

References

Anderson, J., Mohammed, S., Grimm, B., Jones, B., Koshevoy, P., Tasdizen, T., Whitaker, R., Marc, R., 2011. The viking viewer for connectomics: scalable multi-user annotation and summarization of large volume data sets. Journal of Microscopy 241, 13–28.

Chang, C.-C., Lin, C.-J., 2011. Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2, 27.

Daz, E., Ayala, G., Daz, M.E., Gong, L.-W., Toomre, D., 2010. Automatic detection of large dense-core vesicles in secretory cells and statistical analysis of their intracellular distribution. IEEE/ACM Transactions on Computational Biology and Bioinformatics 7, 2–11.

Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The pascal visual object classes (VOC) challenge. International Journal of Computer Vision 88, 303–338.

Jagadeesh, V., Vu, N., Manjunath, B., 2011. Multiple structure tracing in 3d electron micrographs. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011. Springer, pp. 613–620.

Kaynig, V., Fuchs, T., Buhmann, J.M., 2010. Neuron geometry extraction by perceptual grouping in system images. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 2902–2909.

Kreshuk, A., Straehle, C.N., Sommer, C., Koethe, U., Cantoni, M., Knott, G., Hamprecht, F.A., 2011. Automated detection and segmentation of synaptic contacts in nearly isotropic serial electron microscopy images. PLOS One 6, e24899.

Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2. IEEE, pp. 2169–2178.

Lucchi, A., Smith, K., Achanta, R., Lepetit, V., Fua, P., 2010. A fully automated approach to segmentation of irregularly shaped cellular structures in em images. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010. Springer, pp. 463–471.

Manjunath, B.S., Ma, W.-Y., 1996. Texture features for browsing and retrieval of image data. IEEE Transactions on Pattern Analysis and Machine Intelligence 18, 837–842.

Matas, J., Chum, O., Urban, M., Pajdla, T., 2002. Robust wide baseline stereo from maximally stable extremal regions. In: British Machine Vision Conference, vol. 1, pp. 384–393.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 971–987.

Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. International Journal of Computer Vision 42, 145–175.

Otsu, N., 1975. A threshold selection method from gray-level histograms. Automatica 11, 23–27.

Varma, M., Babu, B.R., 2009. More generality in efficient multiple kernel learning. In: Proceedings of the 26th Annual International Conference on Machine Learning. ACM, pp. 1065–1072.

Varma, M., Zisserman, A., 2003. Texture classification: are filter banks necessary? IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, Proceedings, vol. 2. IEEE, p. II-691.

Zuiderveld, K., 1994. Contrast limited adaptive histogram equalization. In: Graphics Gems IV. Academic Press Professional, Inc., pp. 474–485.