Signal Processing Challenges in Quantitative 3-D Cell Morphology: More than meets the eye

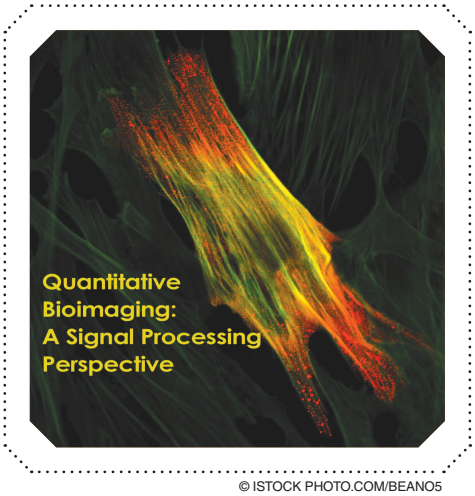
ARTICLE *in* IEEE SIGNAL PROCESSING MAGAZINE · JANUARY 2015

Impact Factor: 5.85 · DOI: 10.1109/MSP.2014.2359131

Alexandre Dufour, Tzu-Yu Liu, Christel Ducroz, Robin Tournemenne, Beryl Cummings, Roman Thibeaux, Nancy Guillen, Alfred Hero III, and Jean-Christophe Olivo-Marin

Signal Processing Challenges in Quantitative 3-D Cell Morphology

More than meets the eye



Modern developments in light microscopy have allowed the observation of cell deformation with remarkable spatiotemporal resolution and reproducibility. Analyzing such phenomena is of particular interest for the signal processing and computer vision communities due to the numerous computational challenges involved, from image acquisition all the way to shape analysis and pattern recognition and interpretation. This article aims at providing an up-to-date overview of the problems, solutions, and remaining challenges in deciphering the morphology of living cells via computerized approaches, with a particular focus on shape description frameworks and their exploitation using machine-learning techniques. As a concrete illustration, we use our recently acquired data on amoeboid cell deformation, motivated by its direct implication in immune responses, bacterial invasion, and cancer metastasis.

MOTIVATION AND CHALLENGES

Cell deformation and migration are dynamic processes regulated by a complex machinery with major implications on a number of key processes in biology including development, immune responses, and invasive processes [1]. A method of choice for studying this mechanism lies in light microscopy, whereby living cells evolving in their three-dimensional (3-D) environment (both in vitro and in vivo) can be imaged over prolonged periods of time with limited invasiveness, producing time-lapse sequences of volumetric 3-D images [2]. Due to the considerable complexity of cell deformation and migration, visual analysis of such processes is no longer limited just by user bias and fatigue but also fails to apprehend large-scale, population-wise patterns that may otherwise appear random or disorganized. Systematic quantitative analysis and understanding of cellular dynamics is becoming a major interest for the signal processing and computer vision communities, given the wide range of computational challenges to overcome. These challenges principally fall into one of the following five categories, covering many aspects of the experimental pipeline (cf. Figure 1).

IMAGE RECOVERY

Modern optical light microscopy techniques have substantially expanded the diversity and reliability of live cell imaging applications, constantly improving on speed, penetration depth, and spatial resolution, though usually at the expense of the signal-to-noise ratio. An important part of the literature therefore focuses on the development of deconvolution and denoising techniques adapted to the peculiarities of bioimaging data (e.g.,

Digital Object Identifier 10.1109/MSP.2014.2359131
Date of publication: 5 December 2014

1053-5888/15©2015IEEE

mixed Poisson–Gaussian noise and anisotropic lateral-to-axial resolution [3]–[5]), while recent advances in computational optics (notably compressive sensing and superresolution techniques) have given rise to new challenges in signal reconstruction and inverse problems [6]–[9].

SEGMENTATION AND TRACKING

The increasing diversity and complexity of environments in which motile cells can be observed has made their reliable detection and tracking most challenging, notably in crowded or cluttered environments [10]. Added to the sheer amount of routinely produced 3-D imaging data, the need to develop fast and semi- to fully automated approaches remains a long-standing challenge in the community [11], [12].

SHAPE REPRESENTATION

Although informative, raw shape and trajectory information are usually too large and complex to produce interpretable results, unless their dimensionality is sufficiently and adequately reduced. Unfortunately, the natural variability of shape configurations observable within so-called homogeneous cell populations pose significant challenges in defining descriptors that are both robust to noise and retain enough specificity across populations. While much effort has been conducted to develop such descriptors in two-dimensional (2-D) [13] or pseudo-3-D [14], 3-D-shape descriptors that permit robust morphological analysis and facilitate human interpretation are still under active investigation [15]–[21].

LEARNING AND INTERPRETATION

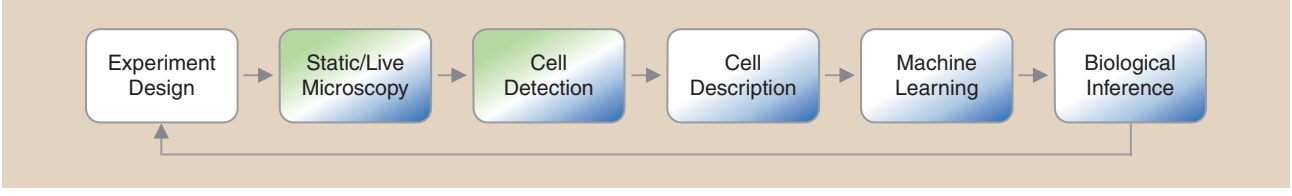
Linked to the issue of shape description is that of invariant analysis of cell populations across various experimental conditions. The difficulty here lies in two aspects: 1) developing pattern recognition and machine-learning approaches able to capture the differences between populations while remaining robust to intraclass variability [22], [23] and 2) highlighting such differences in a human-readable form and ultimately leading to the inference of standardized computational models with the aim of deriving novel biological hypotheses [24].

SYSTEMATIC QUANTITATIVE ANALYSIS AND UNDERSTANDING OF CELLULAR DYNAMICS IS BECOMING OF MAJOR INTEREST FOR THE SIGNAL PROCESSING AND COMPUTER VISION COMMUNITIES, GIVEN THE WIDE RANGE OF COMPUTATIONAL CHALLENGES TO OVERCOME.

AVAILABILITY AND REPRODUCIBILITY

The vast majority of developments in the community generally appears in the literature in the form of theoretical workflows that facilitate understanding and software (or hardware) implementation. Unfortunately, the increasing complexity of these protocols renders their implementation and validation very tedious for nonspecialists, hindering both their adoption and reproducibility. Ironically, while many scientific findings are based on computerized analysis, the associated computer codes are only rarely made public in contrast to reproducible research practices in other scientific domains [25]. Community efforts such as the Reproducible Research Initiative strive to make both code and data publicly available, although more support from publishers and/or research sponsors is required [26].

In the specific context of 3-D bioimaging, the image recovery and segmentation aspects have received extensive focus from the signal processing community over the last several decades, as illustrated by the recent introduction of challenges at the IEEE International Symposium on Biomedical Imaging, with special sessions on image deconvolution (http://bigwww.epfl.ch/deconvolution/challenge), particle tracking [27], and cell segmentation and tracking [12]. The shape representation and machine-learning aspects have been comparatively less thoroughly investigated, even though they provide essential keys to decipher the cell machinery. Here we review recent developments in the fields of cell-shape description and associated machine-learning approaches, highlighting the current state of the art and the challenges ahead toward a comprehensive understanding of cellular dynamics. We also review a number of open-source software solutions that permit reliable and reproducible quantification of cellular images. We shall illustrate this review using the example of amoeboid cell deformation, which is a mechanism of strong interest in the life science community due to its importance in immune response, infectious diseases, and cancer metastasis. Amoeboid motion is characterized by the emission of localized protrusions at the cell surface that permit environment scanning and motion initiation (cf. Figure 2) [28], [29] and poses significant challenges in terms of quantitative characterization and comparative phenotyping.



[FIG1] Typical experimental pipeline for cellular (notably phenotypic) studies. Boxes marked with green and blue labels indicate the availability of associated open hardware and software developments, respectively.

3-D-SHAPE DESCRIPTORS AND THE CELL

Proper description of the cell shape in a numerical form is inherently dependent on its underlying representation, i.e., how it was extracted from the image data [13]. While most segmentation methods usually produce voxel-based masks or distance maps as an indicator of the cell location and cell interior (analogous to the input data), outline-based approaches produce a description that is homotopic to the cell surface (e.g., using control points or parameterized curves [14], [30], [31]). From any of these representations, various low-level descriptors can be extracted (e.g., volume, surface, elongation, ellipticity, compacity, etc.), providing a coarse appreciation of the global shape conformation (most of these descriptors can be extracted using the approaches further reviewed below). However, when dealing with more complex shapes such as cells or organs, such low-level descriptors quickly become sensitive to increasing amounts of noise and usually fail to capture subtle shape variations occurring at different spatial scales, from large conformation changes (e.g., elongation, contraction) to smaller variations at the cell surface (e.g., protrusions) [32]. Higher-lever descriptors based on multiscale decomposition become particularly interesting in this context, as they decompose the shape in a coarse-to-fine manner, allowing one to restrict the analysis to the scale of interest and thereby increasing robustness to negligible or unlikely shape variations. These approaches typically fall into one of the three following categories, classified in increasing order of complexity.

LANDMARK-BASED APPROACHES

Landmark-based approaches are a popular choice for morphological studies [33], with numerous applications in medical imaging [34], [35], evolutionary biology [21], and face recognition [36]. The common denominator in these fields is the availability of a reliable low-dimensional model for the shapes of interest, allowing one to reduce the description of the shape (or its tolerated deformations) to a small number of control points (or parameters). This simplified representation in turn permits efficient registration, statistical analysis, and template modeling [37]. Applications in cellular morphology are, however, not as common, mostly due to the fact that deforming cells generally have many more degrees of freedom that cannot be accurately captured using such methods (counterexamples can be found in specific biological applications, e.g., [38]).

IRONICALLY, WHILE MANY SCIENTIFIC FINDINGS ARE BASED ON COMPUTERIZED ANALYSIS, THE ASSOCIATED COMPUTER CODES ARE ONLY RARELY MADE PUBLIC IN CONTRAST TO REPRODUCIBLE RESEARCH PRACTICES IN OTHER SCIENTIFIC DOMAINS.

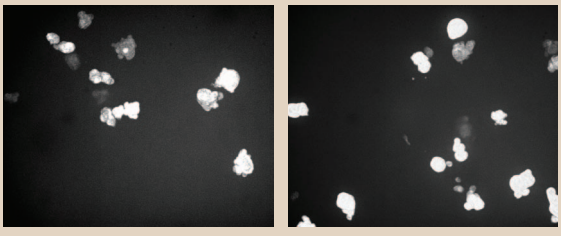
GRAPH-BASED REPRESENTATIONS

Graph-based representations fall in two subcategories, depending on whether they describe the interior or the outline of the shape of interest. In the former case, the cell body is converted (e.g., from a initial binary mask) into a hierarchical treelike graph connecting virtual landmarks inside the cell. Typical examples include morphological skeletons, medial axis transforms, or Voronoi tessellations [39], [40]. Once the graph is obtained, local shape features at the cell surface (the leaves of the tree) are semantically segregated from large shape conformation (closer to the root). As the graph generation process may be subject to noise, adequate graph pruning algorithms are required to differentiate structures at the cell edge (e.g., filopodia from erratic spikes), thereby permitting an unbiased analysis of the cell deformation over time [40]. However, such approaches remain limited to 2-D analysis, and their extension to 3-D is computationally challenging.

The latter category considers a surface-based graph representation of the shape of interest. While the topic of signal processing on arbitrary graphs is only in its early days [41], several methods have been developed for the specific case of closed surfaces (homeomorphic to the two-sphere), such as energy-minimizing graph matching (developed for protein surface alignment [42]) and graph-based spherical wavelets (applied to cell-shape analysis in [18]).

MOMENT-BASED APPROACHES

These approaches consider the shape of interest as an arbitrary spatial distribution function that is then mathematically represented as a sum of known polynomial functions, thereby permitting the extraction of geometrical moments with suitable invariants [43]. Such methods generalize traditional Fourier analysis to arbitrary distributions and therefore share the same descriptive properties: low-order moments describe the coarse conformation, while high-order moments retain information at higher frequency. For this reason, these approaches have been utilized in many areas of image processing, with popular choices of



[FIG2] Planar slices of two field of views representing (a) wild type and (b) chemically modified parasites. Distinguishing between these populations based on shape information is particularly challenging, even for the trained eye, and requires robust quantitative tools for shape description and machine learning.

bases including Lengendre, Zernike, Tchebichef polynomials (see [44] for a comparison), and splines [45].

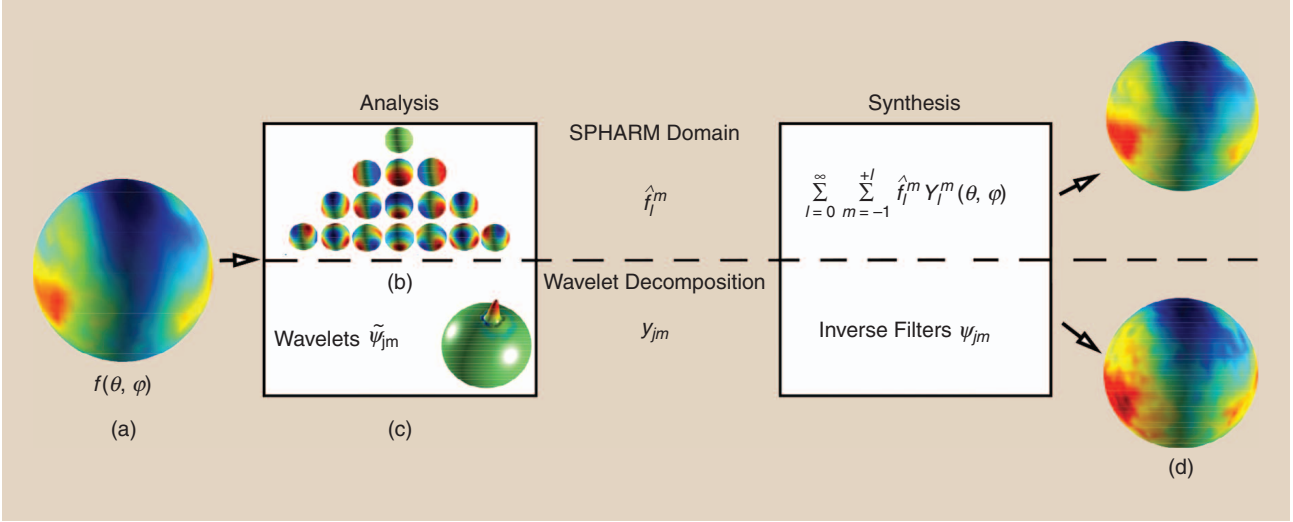
In the context of cell-shape description, such approaches are generally not applied on the raw image data. Instead, a binary mask or outline of the shape is first extracted and then projected onto an appropriate basis [14], [16], [19], [20], [46], [47]. These methods can further be decomposed into two categories. On the one hand, the 2-D cell outline is projected directly onto a chosen basis (e.g., Fourier [46] or splines [14]), and the process is repeated on each slice of the cell shape to obtain a 3-D set of descriptors [14]. On the other hand, the surface of interest is first mapped onto the sphere using appropriate spherical parameterization techniques [48] and then projected onto a reference function basis living on the sphere. Two popular and complementary candidates in this family are the spherical harmonics (SPHARM) [47] and spherical wavelets (SWAVE) [49], which significantly differ from the eponym graph-based approach, notably in the way they are constructed. In the former case, the spherical signal is projected onto a basis of Legendre polynomials, extending the classical Fourier analysis to signals on the two-sphere [cf. Figure 3(b)]. SPHARM therefore have global spatial support, and each coefficient describes the general conformation of the shape of interest at different spatial scales. Applications of SPHARM include molecular surface modeling [50], [51], medical-shape analysis [52], and cell-shape analysis [15], [16], [19], [53]. In the latter case, the function basis is formed of wavelets (hence its name), and are constructed by analogy to wavelets in the plane via appropriate spherical projections [17], [20], [54]. Here the local spatial support provided by SWAVE is of particular interest to localize specific features along the surface [cf. Figure 3(c)].

While the function bases utilized here are not specific to cellular shapes, other approaches have been proposed to increase their specificity by locally adapting the basis to the data set at hand (e.g., using the Laplace–Beltrami operator, as in [19] and [55]). Nevertheless, the use of standard function bases preserves two major advantages: 1) they are ubiquitousness in signal processing applications that propels the creation of ever-more efficient computational implementations and 2) they permit an unbiased description and comparison of shapes across multiple experimental conditions, and also serve as a basis to perform shape synthesis (cf. Figure 3) or build so-called generative models of the cell [24].

SHAPE EXTRACTION AND DESCRIPTION HAVE LED TO SIMPLIFIED REPRESENTATIONS OF THE RAW DATA FROM 3-D IMAGES DOWN TO A SMALLER SET OF DESCRIPTORS PER CELL.

RECOGNITION, CLASSIFICATION, AND INTERPRETATION

In terms of dimensionality, shape extraction and description have led to simplified representations of the raw data from 3-D images (on the order of $10^{8–10}$ voxels) down to a smaller set of descriptors (also called *features*) per cell (on the order of $10^{1–3}$). Unfortunately, these feature sets are rarely translatable to a concrete, biologist-friendly interpretation of the biological experiment, rather appearing as large arrays of poorly informative numbers. This motivates the following questions: Which features really matter? What is the influence of the experimental conditions on these features? How do they translate into biological terms? Machine-learning approaches are particularly well suited to answer these questions, and choosing the appropriate technique depends on the application and how



[FIG3] The comparison of spherical harmonics and spherical wavelets for signal processing on the sphere. (a) The signal of interest $f(\theta, \varphi)$ can be projected on a basis of functions with global spatial support such as (b) spherical harmonics, or a basis of functions with local support such as (c) spherical wavelets. In both cases, the spherical signal can be synthesized back (d) from the coefficients.

easily the initial question can be cast into a machine-learning framework [23].

Machine learning consists of building computer models that accurately describe a given population of individuals with the ultimate goal to either characterize subgroups of individuals with similar properties, or to predict the properties of a new unknown (or simulated) individual [62]. Machine-learning techniques are generally split into two categories: 1) supervised techniques require that a subset of individuals in each subpopulations be manually annotated to classify the rest of the data set, while 2) unsupervised techniques learn the model directly from the inherent structure of the feature set without user intervention. Both families have their set of advantages and drawbacks, which we discuss next in the context of cell-shape analysis (we refer the reader to [23] for an applications-oriented review, and [63] for a more theoretical introduction).

QUANTITATIVE MORPHOLOGY OF SINGLE CELLS IS ONLY THE VISIBLE PART OF THE DIGITAL BIOIMAGING "ICEBERG."

Supervised techniques require a subset of the data to be annotated, i.e., one or more examples describing all known subclasses of the population must be indicated beforehand. This so-called training set is then used to learn an optimal classifier. Popular approaches for cell classification from image data include k-nearest neighbors [64], decision trees [65], and support vector machines (SVMs) [57]. These approaches can be further combined, resulting in so-called ensemble learning techniques, e.g., boosting [22] and random forests [66]. Supervised techniques are widely popular due to their robustness to noise and apparent intuitiveness and versatility. Annotating a small finite number of examples is generally acceptable in many applications, while the classifier itself remains sufficiently generic to accept a wide range of applications. Unfortunately, some of these methods may suffer from overfitting when the dimension becomes much larger than the number of samples. Variable selection techniques provide a solution to this issue, while improving learning accuracy and often facilitating interpretation. Typical examples include forward selection, backward elimination and sparsity-constrained classifiers [67]. The major limitation of supervised learning is that it requires class labels to be available: nonannotated subpopulations will not be learned and, hence, not recognized, and by extension, novel unknown subpopulations (e.g., unpredicted cell phenotypes) cannot be discovered.

Unsupervised techniques do not require a training set and can be applied to unlabeled populations. They learn the inherent structure of the data set using a predefined metric (e.g., a similarity or distance measure between individuals), permitting homogeneous groups or dimensions to be distinguished. One usually distinguishes clustering techniques, which aim at extracting subpopulations sharing similar properties according to the considered metric, from dimensionality reduction techniques, which aim at selecting a subset of essential principal components that best represent a high-dimensional data set to facilitate user interpretation. Classical clustering techniques include *k*-means [68] and Gaussian mixture modeling (GMM) [69], while dimensionality reduction techniques include principal component analysis (PCA) [46] and independent component analysis (ICA) [70]. The major advantage of unsupervised learning is that the data labels need not be known in advance, alleviating the need for data annotation while allowing the discovery of unexpected subpopulations, giving them remarkable exploratory potential in biology [71]. Unfortunately, unsupervised techniques also have their drawbacks: they are more sensitive to noise, defining the appropriate metric for the data set at hand can be complex for high-dimensional data set, as is the interpretation of the results.

It is worth pointing out that most of these techniques can also be applied directly to the raw image data without necessarily needing a preliminary shape extraction and description step. Shape description becomes necessary as soon as both qualitative and quantitative characterization or modeling of the cell shape is required, notably when studying the effect of known experimental conditions on the cell phenotype.

AVAILABILITY AND REPRODUCIBILITY

Emerging interdisciplinary fields such as bioimage informatics foster interactions across an ever broader portfolio of scientific expertise (this article only mentions six of them: optics, signal processing, image segmentation, object tracking, shape description, and machine learning). Unfortunately, novel algorithmic developments in many of these fields are only rarely published in the form of ready-to-use software, while reimplementing the underlying method becomes increasingly challenging for the nonspecialist in

[TABLE 1] OPEN-SOURCE SOFTWARE SOLUTIONS WITH DEDICATED MODULES FOR CELL-SHAPE ANALYSIS.

NAME	REFERENCE	SUPPORTED LANGUAGES	2-D/3-D	SHAPE DESCRIPTORS	MACHINE LEARNING
CELLCLASSIFIER	[56]	MATLAB*	2-D	GEOMETRIC	SUPERVISED
CELLCOGNITION	[57]	PYTHON, C++	2-D	GEOMETRIC	SUPERVISED AND UNSUPERVISED
CELLORGANIZER	[24]	MATLAB*	2-D, 3-D	SPLINES	UNSUPERVISED (GENERATIVE)
CELLPROFILER	[22]	PYTHON, VISUAL PROGRAMMING	2-D	GEOMETRIC	SUPERVISED
EBIMAGE	[58]	R	2-D	GEOMETRIC	(VIA R)
ICY	[59]	JAVA, SCRIPTING, VISUAL PROGRAMMING	2-D, 3-D	GEOMETRIC, SPHARM	(VIA PLUGINS)
IMAGEJ/FIJI	[60]	JAVA, SCRIPTING, MACRO RECORDING	2-D, 3-D	GEOMETRIC	(VIA PLUGINS)
TANGO	[61]	IMAGEJ, R	3-D	GEOMETRIC	(VIA R)

*MATLAB is licensed by Mathworks.

SPHARM AS SHAPE DESCRIPTORS

SPHARM are defined as

$$Y_l^m(\theta,\varphi) = k_{l,m} P_l^m(\cos\theta) e^{im\varphi},$$

where θ and φ parameterize the spherical domain, l and m are respectively the degree and order of the harmonic, $k_{l,m}$ is the expansion coefficient, and P_l^m is the associated Legendre polynomial.

Spherical harmonic analysis is a natural extension of traditional Fourier analysis for signals defined on the unit sphere. Hence, any arbitrary function f defined on the sphere can be expanded using the SPHARM transform, given by

$$f(\theta,\varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C_l^m \cdot Y_l^m(\theta,\varphi),$$

where C_l^m are the generalized Fourier coefficients with respect to the SPHARM basis, or more simply, SPHARM coefficients.

To conduct a SPHARM expansion of the cell shape, its surface must be written as a spherical function, which is done via so-called spherical parameterization techniques [47]. While some surfaces may not be bijectively transposable to the sphere by a simple radial projection (also referred to as nonstar-shaped surfaces), a classical approach is to project each Cartesian component of the surface independently, yielding a vector of spherical functions $f = [f_x(\theta,\varphi)\, f_y(\theta,\varphi)\, f_z(\theta,\varphi)]$ [48]. Expanding f thus yields three sets of SPHARM coefficients $C_l^m = [(C_l^m)_x\, (C_l^m)_y\, (C_l^m)_z]$. Rotation invariant coefficients are subsequently obtained by considering their L2-norm:

$$\hat{C}_l^m = \| C_l^m \| = \sqrt{(C_l^m)_x^2 + (C_l^m)_y^2 + (C_l^m)_z^2}.$$

[TABLE 2] A COMPARISON OF SEVERAL SUPERVISED CLASSIFICATION APPROACHES FOR CELL SHAPE CLASSIFICATION.

CLASSIFICATION METHOD	SPHARM FEATURE SET	ERROR (%)
k-NEAREST NEIGHBORS [64]	STANDARD ($K = 1, Q = 1$)	43.96
	POPULATION	33.75
	TEMPORAL	26.25
	COMBINED	29.23
DECISION TREES [65]	STANDARD ($K = 1, Q = 1$)	45.52
	POPULATION	35.94
	TEMPORAL	31.67
	COMBINED	30.94
SVM [57]	STANDARD ($K = 1, Q = 1$)	42.69
	POPULATION	27.73
	TEMPORAL	18.08
	COMBINED	20.00
STRUCTURED SVM [74]	STANDARD ($K = 1, Q = 1$)	40.45
	POPULATION	31.36
	TEMPORAL	17.69
	COMBINED	19.23

SPHARM-BASED FEATURE DESIGN FOR CLASSIFICATION

Protocol

- SPHARM expansion of the cell surfaces (cf. "Spherical Harmonics as Shape Descriptors") is conducted with an empirical precision of $l = 5$, yielding an array of 21 rotationally invariant coefficients per cell (this value depends on the application, and defines the balance between shape and noise information).
- The data set is then divided into groups of $K \in \{1, 2, \ldots, 5\}$ randomly selected cells observed over $Q \in \{1, 5, 10, 15, 20, 25\}$ consecutive frames (the starting frame is random if Q is less than the entire length of the video). We avoid imbalance in the training samples by randomly subsampling the larger class so that the sizes of both classes are identical.
- Finally, the coefficients are indexed throughout the data set as $\hat{C}_l^m(i, \text{cell}_k, q)$, where $i \in \{1, 2, \ldots, n\}$ is the sample indicator, $k \in \{1, \ldots, K\}$ indicates the cell, and $q \in \{1, \ldots, Q\}$ indicates the time at which the frame is acquired. From this data set, one can design a structured combination of features. Here we illustrate two possible combinations:

Population features (μ) are obtained by averaging each coefficient over Q frames for each cell of a group, yielding $21K$ features:

$$\mu_l^m(i, \text{cell}_k) = \frac{1}{Q} \sum_{q=1}^{Q} \hat{C}_l^m(i, \text{cell}_k, q).$$

Temporal features ($\tilde{\mu}$) are obtained by averaging each coefficient over a group of K cells in each frame, yielding $21Q$ features:

$$\tilde{\mu}_l^m(i, q) = \frac{1}{K} \sum_{k=1}^{K} \hat{C}_l^m(i, \text{cell}_k, q).$$

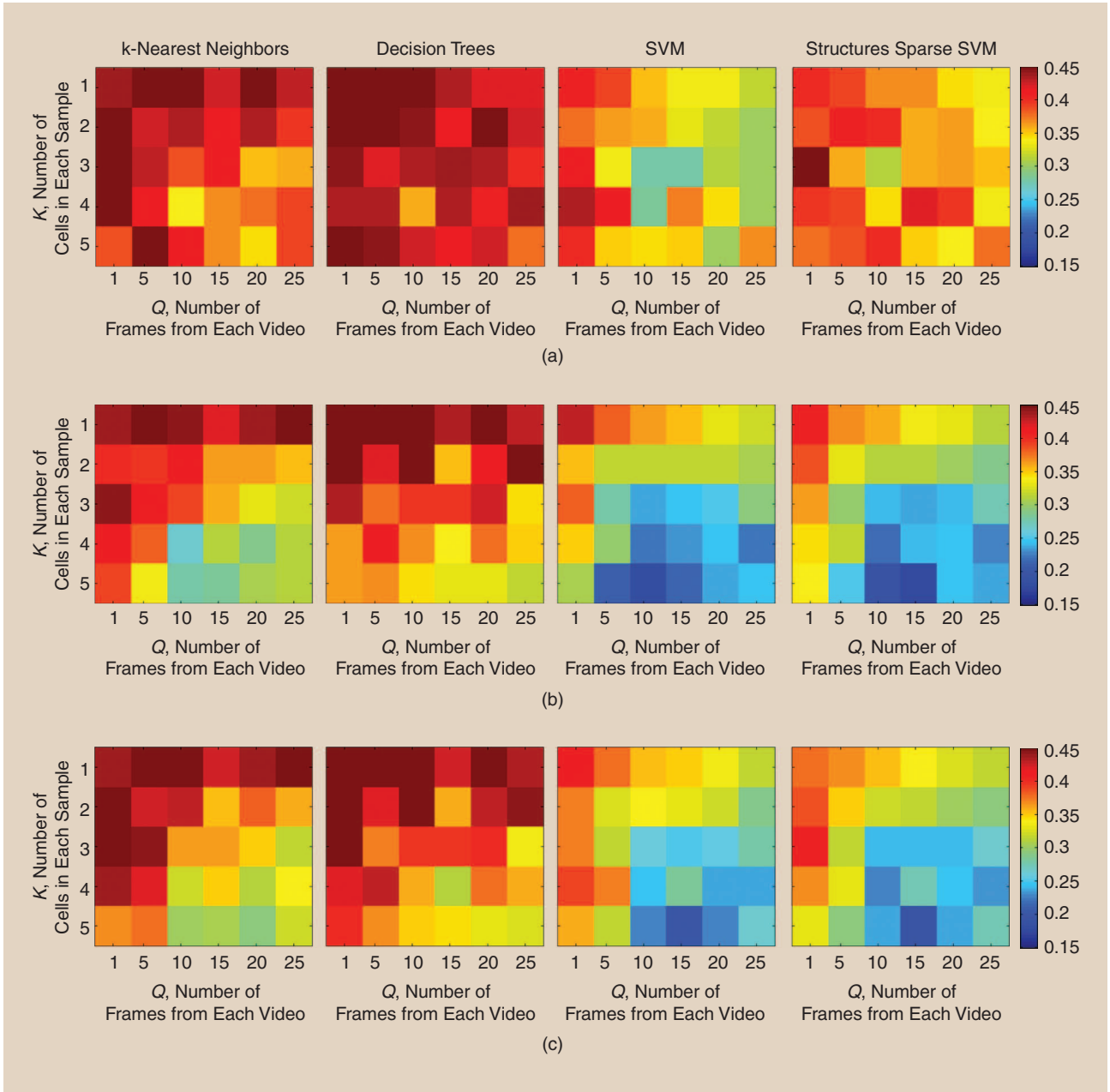
From these descriptors, a set of group structures can be derived:

$$I_{\mu l^*} = \{\mu_l^m(i, \text{cell}_k) : l = l^*, \forall i, k\}$$
$$I_{\tilde{\mu} l^*} = \{\tilde{\mu}_l^m(i, t) : l = l^*, \forall i, t\}.$$

addition to being time-consuming and error prone [25]. Fortunately, the bioimage informatics community has been a proactive driver of the Reproducible Research Initiative [26] with the appearance of multiple software solutions for quantification in bioimaging over the last decades (cf. [72] and Table 1). Some of these frameworks rely on a so-called plug-in architecture, allowing their enrichment via third-party contributions, removing much of the redundant work including data loading and visualization, and streamlining the publishing process. Each platform will tend to be specialized in a specific discipline (mostly driven by its core developers), and current efforts in the community are devoted to provide higher interoperability across software, so as to combine the best solutions available for to tackle a given problem. Table 1 provides a nonexhaustive list of such software platforms, focusing specifically on the solutions available for cell-shape analysis and

quantification. Such software can be classified in two categories: general-purpose image-based quantification software for which specific third-party modules for cellular analysis have been developed (this is notably the case of Fiji/ImageJ [60] and Icy [59]), and more streamlined software dedicated almost exclusively to cell-based analysis (such as CellClassifier [56], CellCognition [57], CellOrganizer [24], CellProfiler [22], EBImage [58], and Tango [61]). While some of these software packages limit their analysis to the 2-D case, the methods are in principle extensible to 3-D. This limitation is notably present in the field of high-throughput, high-content screening (HT-HCS), where 3-D imaging considerably increased acquisition and analysis times, alongside data management issues [22].

THE ROAD TO BUILDING A THE CELL IS PAVED WITH MAJOR DIFFICULTIES IN TERMS OF MATHEMATICAL MODELING AND SIGNAL PROCESSING.



[FIG4] The performance comparison of several classification methods on various combinations of SPHARM features. Each table cell indicates the classification error for a particular combination of the number of cells per group (K) and the number of time points per cell (Q), as defined in "SPHARM-Based Feature Design for Classification." (a) Population features (μ). (b) Temporal features ($\tilde{\mu}$). (c) Population and temporal features ($\mu, \tilde{\mu}$).

CLASSIFICATION METHODS

We consider an input feature set $x_i \in R^p$ and the class labels $y_i \in \{+1, -1\}$, where $i = 1, \ldots, N$. Under this notation, we can describe the classification methods used in this review as follows.

k-Nearest Neighbors

Let $N_k(x)$ be the neighborhood of x found by the k nearest samples, defined by some metric, e.g., Euclidean distance. The decision rule is defined by a majority vote on $\{y_i \,|\, x_i \in N_k(x)\}$.

Decision Trees

The decision tree is an greedy algorithm that adds splitting nodes to the tree by defining half planes $P_1 = \{x \,|\, x_j \leq s\}$ and $P_2 = \{x \,|\, x_j > s\}$, in which x_j is the splitting variable and s is the splitting point. At each candidate node, compute the impurity, e.g., the Gini index

$$I = p_{-1}(1 - p_{-1}) + p_{+1}(1 - p_{+1}),$$

in which p_k is the fraction of class k observed at that node. The splitting nodes are selected to improve the homogeneity sequentially, and the decision at each leaf node is by majority vote.

SVMs

Let $f(x) = w'x + b$ be a decision function that assigns observation x to $sign(f(x))$, then $f(x) = 0$ is a hyperplane in R^p. In the general case, where classes are not linearly separable, the SVM is written as

$$\min_f \sum_{i=1}^{n} V(f, x_i) + \lambda R(f)$$

with

$$V(f, x_i) = [1 - y_i f(x_i)]_+, R(f) = \frac{1}{2} \| w \|_2^2.$$

Structured SVMs

The structured sparse SVM is formulated as an SVM but with a different regularization:

$$V(f, x_i) = [1 - y_i f(x_i)]_+, R(f) = \sum_{g=1}^{G} \| w_{I_g} \|_2,$$

where I_g is the set indexing the variables that are in the gth group.

On the hardware side, recent efforts to make public blueprints and to increase the interaction between hardware and software has led to the development of so-called smart acquisition systems. Some of the key contributions on the hardware side include the Warwick Open Microscope System (http://wosmic.org) and the OpenSPIM project (http://openspim.org), while integrated hardware and software approaches have been developed in several projects including μManager (http://micro-manager.org) and MicroPilot [73].

CASE STUDY: AMOEBOID CELL DEFORMATION

In this section, we illustrate the use of shape description and classification in the context of studying morphology and motility of *Entamoeba histolytica,* a unicellular parasite responsible for the amoebiasis disease. Recent studies in vitro and ex vivo have suggested that parasites specifically modified to prevent their degrading of the extracellular matrix remained able to migrate at the same speed as unmodified parasites, possibly due to subtle shape changes that simple descriptors could not fully capture [32]. Illustrative slices of the 3-D data set are presented in Figure 2. Here cells are segmented using active contours [30], and described by SPHARM features (cf. [16] and "SPHARM as Shape Descriptors" and "SPHARM-Based Feature Design for Classification"). Although other combinations of methods and descriptors are possible, our aim here is to show how an adequate design of feature vectors may lead to discoveries of significant and interpretable differences between populations that otherwise seem visually identical.

Table 2 reports the comparative performance of several classification approaches (described in "Classification Methods") with various combinations of features. The performance is evaluated by leave-one-out cross-validation, while the entire leave-one-out process is repeated for ten trials to report the averaged performance for each method. Both the temporal features (averaged over groups of cells) and population features (averaged over a fixed number of frames) are tested against variable group sizes, as illustrated in the error rate heat-maps presented in Figure 4. As the group size increases, the classification performance systematically improves, suggesting that more robust classification can be achieved by group-based analysis. Table 2 further indicates that the best performance is obtained either using temporal features or a combination of temporal and population features. However, combining both feature sets does not necessarily yield the best performance, suggesting that temporal information is more discriminant. Finally, structured SVM imposes sparsity to select an optimal subset of features and has better classification performance. Also, by specifying the most predictive subset of features, the structured SVM has the advantage of providing feature interpretability. While each feature alone does not necessarily translate directly to biological knowledge, this example illustrates that a careful and systematic design of the features can highlight significant and unexpected discrepancies between cell populations, which in turn can potentially lead to new interpretations and hypotheses that enhance the design of the next experiment.

FUTURE PERSPECTIVES AND CHALLENGES

Numerous breakthroughs in imaging and computational techniques have had a considerable impact on the amount of quantitative data that describe the behavior of single cells evolving in their 3-D environment. After suitable standardization, such data become amenable to proper mathematical characterization, invariant description, and classification. Thus, these tools have exciting potential to reveal the complexity of biological mechanisms at all spatial scales: 1) at the single cell level, the mechanisms and signals responsible for the transition between modes of migration (notably in cancer development); 2) at the group level, the short- and long-distance signaling queues that induce cells to interact, differentiate,

or migrate to distant locations (e.g., for tissue repair and immune responses); and 3) at the tissue level, how information is propagated across a dense cell network, and how this information is locally interpreted to drive cell intercalation, differentiation, and renewal (e.g., during embryo- and morphogenesis). In this article, we have highlighted some of the latest developments in quantitative tools and associated software packages to study some of these processes and illustrated how group-based analysis of cell morphology provides a much more powerful and discriminant description of a cell population as compared to single-cell analysis, while temporal information carries a significant potential to improve the overall classification performance. Yet, quantitative morphology of single cells is arguably only the visible part of the digital bioimaging "iceberg."

THE NEXT MAJOR CHALLENGES IN THE BIOIMAGING AND BIOSIGNAL PROCESSING FIELD LIE IN STUDYING SPATIOTEMPORAL PROCESSES BEYOND SINGLE CELLS, FROM THE NANOSCOPIC TO THE MACROSCOPIC SCALE.

The next major challenges in the bioimaging and biosignal processing field lie in studying spatiotemporal processes beyond single cells, from the nanoscopic to the macroscopic scale. At the subcellular level, the mechanisms that underlie cell deformation and motility are still poorly understood, mostly due to the lack of proper visual insight into the various architectural components (down to individual proteins) forming the cell cytoskeleton. The road to building a computational model of the cell (not to mention the huge variety of cell types in plant or animal models) is paved with major difficulties in terms of mathematical modeling and signal processing and will require the development of novel biophysics-inspired algorithms to: understand how the cytoskeleton is formed, acts, and reacts to internal and environmental signals; and generates force and adhesion that ultimately lead to deformation, movement, and division. At the macroscopic level, studying biological processes at large spatial and temporal scales requires the integrating of single-cell analyses over millions of cells and hours of imaging data in multiple modalities and experimental conditions, raising major visualization and computational bottlenecks. One example of such a challenge is illustrated by the recent advances in selective plane illumination microscopy and its application to the reliable observation of embryonic development in numerous animal models, from the single cell up to tens of thousands of cells [10]. Such data sets have already initiated many developments in image denoising, cell segmentation, cell tracking, and data manipulation software, however, comprehensive modeling of the morphology and trajectory of these cells and their clustering into biologically relevant subpopulations remain open challenges.

AUTHORS

Alexandre Dufour (adufour@pasteur.fr) received the M.Sc. degree in artificial intelligence and pattern recognition from Paris Pierre and Marie Curie University, France, in 2004, and the Ph.D. degree in image processing from Paris Descartes University, France, in 2007. From 2005 to 2007, he was a junior scientist at the Institut Pasteur of Seoul, South Korea. Since 2008, he has been with the Institut Pasteur of Paris, France, where he focuses on three-dimensional deformable models and shape modeling approaches to study morphological and dynamic processes at the cellular and multicellular level. He is also actively involved in open-source software development for reproducible research in bioimage informatics. He is a member of the IEEE Signal Processing Society, the IEEE Bioimaging and Signal Processing Technical Committee, and served as area chair and associate editor in multiple conferences including the 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing; the 2014 IEEE International Conference on Image Processing; the European Signal Processing Conference; and the International Symposium on Biomedical Imaging.

Tzu-Yu Liu (tyliu@eecs.berkeley.edu) received the B.S. degree from the National Taiwan University in 2007 and the Ph.D. degree from the University of Michigan, Ann Arbor, in 2013, both in electrical engineering. Her research interests include statistical learning from high-dimensional and small sample size problems, optimization, structured variable selection, and their applications to biomedical data. She is a postdoctoral researcher at the University of California, Berkeley, where she is focusing on computational biology and machine learning.

Christel Ducroz (christel.ducroz89@gmail.com) received the M.Sc. degree in automatic control, signal, and image in 2011 from Ecole Centrale de Nantes, France. She is currently in her fifth year of medical school. From 2010 to 2012, she was a research assistant with the Quantitative Image Analysis Unit at Institut Pasteur, France, working on three-dimensional cell-shape characterization.

Robin Tournemenne (rtournemenne@gmail.com) received the M.Sc. degree in biomedical engineering in 2013 from RWTH in Aachen, Germany, and graduated from the engineering school Ecole Centrale de Nantes, France, the same year. He is currently pursuing his Ph.D. degree at the Institut de Recherche en Communications et Cybernétique de Nantes, France, where he focuses on physical modeling applied to sound synthesis.

Beryl Cummings (bcummings@g.harvard.edu) received the B.Sc. degree in biology and a certificate in genome sciences and policy from Duke University in Durham, North Carolina. She is currently pursuing her Ph.D. degree in the Harvard Biological and Biomedical Sciences Program, Genetics and Genomics Division in Boston, Massachusetts. Her research is focused on utilizing three-dimensional microscopy to elucidate the relationship between phenotypic plasticity in parasites and their disease pathogenesis.

Roman Thibeaux (roman.thibeaux@pasteur.fr) received the B.Sc. degree in molecular and cellular biology in 2006 and the M.Sc. degree in genetics in 2008 from Paris Denis Diderot

University, France. He then pursued his Ph.D. studies at the Institut Pasteur in Paris, France. Since 2011, he has been a postdoctoral research fellow pursuing his research on tissue invasion and places a particular emphasis on determining the role of host inflammatory response during this invasive process. His research interests include host–pathogen interactions, parasitology, virology, microbiology, extracellular matrix biology, tissue engineering, live imaging, and two-photon microscopy.

Nancy Guillen (nguillen@pasteur.fr) is Research Director 1 at the National Center for Scientific Research, France. She is head of the Cell Biology of Parasitism Unit-INSERM U786, Department of Cell Biology and Infection, Institut Pasteur, Paris, France. She is a member of the Scientific Council, Institut Pasteur, Animal Health Division, Institut National de Recherche Agronomique, and is scientific coordinator of the Pasteur-Weizmman Research Council. Her research interests include the cellular and molecular biology studies of the pathogenic process in *Entamoeba histolytica*, the agent of human amoebiasis, parasitology, cytoskeleton, myosin, cell motility, transcriptome, and physiopathology.

Alfred Hero III (hero@eecs.umich.edu) received the B.S. degree (summa cum laude) from Boston University in 1980 and the Ph.D. degree from Princeton University in 1984, both in electrical engineering. Since 1984, he has been with the University of Michigan, Ann Arbor, where he is the R. Jamison and Betty Williams Professor of Engineering. He received the 2011 University of Michigan Distinguished Faculty Achievement Award. He has been plenary and keynote speaker at major workshops and conferences. He has received numerous Best Paper Awards. He received an IEEE Signal Processing Society (SPS) Meritorious Service Award (1998), the IEEE Third Millenium Medal (2000), an IEEE SPS Distinguished Lecturership (2002), and an IEEE SPS Technical Achievement Award (2014). He was president of the IEEE SPS from 2006 to 2007. He was on the Board of Directors of the IEEE (2009–2011). His recent research interests are in statistical signal processing, machine learning, and the modeling and analysis of high-dimensional spatiotemporal data. He is a Fellow of the IEEE.

Jean-Christophe Olivo-Marin (jcolivo@pasteur.fr) received the Ph.D. and Habilitation à Diriger des Recherchesdegrees in optics and signal processing from the Institut d'Optique Théorique et Appliquée, University of Paris-Orsay, France. He is the head of the Bioimage Analysis Unit and the director of the Center for Innovation and Technological Research at Institut Pasteur, Paris. He has chaired the Cell Biology and Infection Department and was a cofounder of the Institut Pasteur Korea, Seoul, South Korea. His research interests are in image analysis of multidimensional microscopy images, computer vision, and motion analysis for cellular dynamics, and in multidisciplinary approaches for biological imaging. He is a Fellow of the IEEE, an IEEE Signal Processing Society Distinguished Lecturer, chair of the IEEE International Symposium on Biomedical Imaging Steering Committee, a senior area editor of *IEEE Signal Processing Letter*s, and a member of the editorial boards of *Medical Image Analysis* and *BMC Bioinformatic*s. He was the general chair of the 2008 IEEE International Symposium on Biomedical Imaging.

REFERENCES

[1] A. D. Doyle, R. J. Petrie, M. L. Kutys, and K. M. Yamada, "Dimensions in cell migration," *Curr. Opin. Cell Biol.*, vol. 25, no. 5, pp. 642–649, 2013.

[2] M. M. Frigault, J. Lacoste, J. L. Swift, and C. M. Brown, "Live-cell microscopy—Tips and tools," *J. Cell Sci.*, vol. 122, pp. 753–767, Mar. 2009.

[3] Y. Le Montagner, E. Angelini, and J.-C. Olivo-Marin, "An unbiased risk estimator for image denoising in the presence of mixed Poisson-Gaussian noise," *IEEE Trans. Image Processing*, vol. 23, no. 3, pp. 1255–1268, 2014.

[4] M. a. Bruce and M. J. Butte, "Real-time GPU-based 3D deconvolution," *Opt. Express*, vol. 21, pp. 4766–4773, Feb. 2013.

[5] M. Arigovindan, J. C. Fung, D. Elnatan, V. Mennella, Y.-H. M. Chan, M. Pollard, E. Branlund, J. W. Sedat, and D. a. Agard, "High-resolution restoration of 3D structures from widefield images with extreme low signal-to-noise-ratio," in *Proc. Natl. Acad. Sci. USA*, vol. 110, Oct. 2013, pp. 17344–17349.

[6] M. Marim, E. Angelini, and J.-C. Olivo-Marin, "A compressed sensing approach for biological microscopic image processing," in *Proc. IEEE Int. Symp. Biomedical Imaging*, 2009, no. 1, pp. 1374–1377.

[7] P. Ye, J. L. Paredes, G. R. Arce, Y. Wu, C. Chen, and D. W. Prather, "Compressive confocal microscopy," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Apr. 2009, pp. 429–432.

[8] L. Shao, P. Kner, E. H. Rego, and M. G. L. Gustafsson, "Super-resolution 3D microscopy of live whole cells using structured illumination," *Nat. Methods*, vol. 8, pp. 1044–1046, Dec. 2011.

[9] M. Ting, R. Raich, and A. O. Hero, "Sparse image reconstruction for molecular imaging," *IEEE Trans. Image Processing*, vol. 18, no. 6, pp. 1215–1227, 2009.

[10] R. Tomer, K. Khairy, F. Amat, and P. J. Keller, "Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy," *Nat. Methods*, vol. 9, pp. 755–63, July 2012.

[11] E. Meijering, "Cell segmentation: 50 years down the road," *IEEE Signal Processing Mag.*, vol. 29, no. 5, pp. 140–145, 2012.

[12] M. Maska, V. Ulman, D. Svoboda, P. Matula, P. Matula, C. Ederra, A. Urbiola, T. España, S. Venkatesan, D. M. W. Balak, P. Karas, T. Bolcková, M. Streitová, C. Carthel, S. Coraluppi, N. Harder, K. Rohr, K. E. G. Magnusson, J. Jaldén, H. M. Blau, O. Dzyubachyk, P. Krízek, G. M. Hagen, D. Pastor-Escuredo, D. Jimenez-Carretero, M. J. Ledesma-Carbayo, A. Muñoz Barrutia, E. Meijering, M. Kozubek, and C. Ortiz-de Solorzano, "A benchmark for comparison of cell tracking algorithms," *Bioinformatics*, vol. 30, no. 11, pp. 1609-1617, June 2014.

[13] Z. Pincus and J. a. Theriot, "Comparison of quantitative methods for cell-shape analysis," *J. Microsc.*, vol. 227, pp. 140–156, Aug. 2007.

[14] T. Peng and R. Murphy, "Image-derived, three-dimensional generative models of cellular organization," *Cytom. Part A*, vol. 79, no. 5, pp. 383–391, 2011.

[15] K. Khairy, J. Foo, and J. Howard, "Shapes of red blood cells: Comparison of 3D confocal images with the bilayer-couple model," *Cell. Mol. Bioeng.*, vol. 1, pp. 173–181, Sept. 2010.

[16] C. Ducroz, J.-C. Olivo-Marin, and A. Dufour, "Characterization of cell shape and deformation in 3D using spehrical harmonics," in *Proc. IEEE Int. Symp. Biomedical Imaging*, Barcelona, Spain, 2012, pp. 1–4.

[17] C.-j. Du, J. G. Ferguson, P. T. Hawkins, L. R. Stephens, and T. Bretschneider, "Local shape representation in 3D: From weighted spherical harmonics to spherical wavelets," in *Proc. British Machine Vision Conf.*, 2012, pp. 1–12.

[18] C. Ducroz, J.-C. Olivo-Marin, and A. Dufour, "Automatic detection of 3D cell protrusions using spherical wavelets," in *Proc. IEEE Int. Conf. Image Processing*, Melbourne, Australia, 2013, pp. 3499–3502.

[19] C.-J. Du, P. T. Hawkins, L. R. Stephens, and T. Bretschneider, "3D time series analysis of cell shape using Laplacian approaches," *BMC Bioinform.*, vol. 14, p. 296, Jan. 2013.

[20] R. Tournemenne, C. Ducroz, J.-C. Olivo-Marin, and A. Dufour, "3D shape analysis using overcomplete spherical wavelets: Application to bleb detection in cell biology," in *Proc. IEEE Int. Symp. Biomedical Imaging*, 2014, pp. 1–4.

[21] S. Torcida, S. I. Perez, and P. N. Gonzalez, "An integrated approach for landmark-based resistant shape analysis in 3D," *Evol. Biol.*, vol. 41, no. 2, pp. 351–366, 2014.

[22] T. R. Jones, A. E. Carpenter, M. R. Lamprecht, J. Moffat, S. J. Silver, J. K. Grenier, A. B. Castoreno, U. S. Eggert, D. E. Root, P. Golland, and D. M. Sabatini, "Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning," *Proc. Natl. Acad. Sci. USA*, vol. 106, pp. 1826–1831, Feb. 2009.

[23] C. Sommer and D. W. Gerlich, "Machine learning in cell biology—Teaching computers to recognize phenotypes," *J. Cell Sci.*, vol. 126, pp. 5529–5539, Dec. 2013.

[24] T. E. Buck, J. Li, G. K. Rohde, and R. F. Murphy, "Toward the virtual cell: Automated approaches to building models of subcellular organization 'learned' from microscopy images," *BioEssays: News Rev. Mol. Cell. Dev. Biol.*, vol. 34, pp. 791–799, Sept. 2012.

[25] D. C. Ince, L. Hatton, and J. Graham-Cumming, "The case for open computer programs," *Nature*, vol. 482, pp. 485–488, Feb. 2012.

[26] P. Vandewalle, J. Kovacevic, and M. Vetterli, "Reproducible research in signal processing," *IEEE Signal Processing Mag.*, vol. 26, pp. 37–47, May 2009.

[27] N. Chenouard, I. Smal, F. de Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, A. R. Cohen, W. J. Godinez, K. Rohr, Y. Kalaidzidis, L. Liang, J. Duncan, H. Shen, Y. Xu, K. E. G. Magnusson, J. Jaldén, H. M. Blau, P. Paul-Gilloteaux, P. Roudot, C. Kervrann, F. Waharte, J.-Y. Tinevez, S. L. Shorte, J. Willemse, K. Celler, G. P. van Wezel, H.-W. Dan, Y.-S. Tsai, C. O. de Solórzano, J.-C. Olivo-Marin, and E. Meijering, "Objective comparison of particle tracking methods," *Nat. Methods*, vol. 11, pp. 281–289, Mar. 2014.

[28] T. Lämmermann and M. Sixt, "Mechanical modes of 'amoeboid' cell migration," *Curr. Opin. Cell Biol.*, vol. 21, pp. 636–644, Oct. 2009.

[29] M. Bergert, S. D. Chandradoss, R. a. Desai, and E. Paluch, "Cell mechanics control rapid transitions between blebs and lamellipodia during migration," *Proc. Natl. Acad. Sci. USA*, vol. 109, pp. 14434–14439, Sept. 2012.

[30] A. Dufour, R. Thibeaux, E. Labruyère, N. Guillén, and J.-C. Olivo-Marin, "3D active meshes: Fast discrete deformable models for cell tracking in 3D time-lapse microscopy," *IEEE Trans. Image Processing*, vol. 20, pp. 1925–1937, July 2011.

[31] R. Delgado-Gonzalo, N. Chenouard, and M. Unser, "Spline-based deforming ellipsoids for interactive 3D bioimage segmentation," *IEEE Trans. Image Processing*, vol. 22, pp. 3926–3940, Oct. 2013.

[32] R. Thibeaux, A. Dufour, P. Roux, M. Bernier, A.-C. Baglin, P. Frileux, J.-C. Olivo-Marin, N. Guillén, and E. Labruyère, "Newly visualized fibrillar collagen scaffolds dictate Entamoeba histolytica invasion route in the human colon," *Cell. Microbiol.*, vol. 14, pp. 609–621, Jan. 2012.

[33] F. L. Bookstein, *Morphometric Tools for Landmark Data: Geometry and Biology*, vol. 10. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[34] J. Ashburner and K. J. Friston, "Voxel-based morphometry—The methods," *NeuroImage*, vol. 11, pp. 805–821, June 2000.

[35] S. M. Weinberg, N. C. Andreasen, and P. Nopoulos, "Three-dimensional morphometric analysis of brain shape in nonsyndromic orofacial clefting," *J. Anatomy*, vol. 214, pp. 926–936, June 2009.

[36] P. Perakis, G. Passalis, T. Theoharis, and I. A. Kakadiaris, "3D facial landmark detection under large yaw and expression variations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1552–1564, 2013.

[37] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. New York: Wiley, 1998.

[38] J. Neustupa and L. Hodac, "Changes in shape of the coenobial cells of an experimental strain of Pediastrum duplex var. duplex (Chlorophyta) reared at different pHs," *Preslia*, vol. 77, pp. 439–452, 2005.

[39] Y. Xiong, C. Kabacoff, J. Franca-Koh, P. N. Devreotes, D. N. Robinson, and P. a. Iglesias, "Automated characterization of cell shape changes during amoeboid motility by skeletonization," *BMC Syst. Biol.*, vol. 4, p. 33, Jan. 2010.

[40] D. Tsygankov, C. G. Bilancia, E. a. Vitriol, K. M. Hahn, M. Peifer, and T. C. Elston, "CellGeo: A computational platform for the analysis of shape changes in cells with complex geometries," *J. Cell Biol.*, vol. 204, pp. 443–460, Feb. 2014.

[41] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

[42] M. Lozano and F. Escolano, "Protein classification by matching and clustering surface graphs," *Pattern Recognit.*, vol. 39, no. 4, pp. 539–551, 2006.

[43] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inform. Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[44] P. Yap, R. Paramesran, and S. Ong, "Image analysis by Krawtchouk moments," *IEEE Trans. Image Processing*, vol. 12, pp. 1367–1377, Jan. 2003.

[45] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Mag.*, vol. 16, no. 6, pp. 22–38, 1999.

[46] L. Tweedy, B. Meier, J. Stephan, D. Heinrich, and R. G. Endres, "Distinct cell shapes determine accurate chemotaxis," *Sci. Rep.*, vol. 3, p. 2606, Jan. 2013.

[47] C. Brechbühler, G. Gerig, and O. Kübler, "Parametrization of closed surfaces for 3-D shape description," *Comput. Vis. Image Understand.*, vol. 61, no. 2, pp. 154–170, 1995.

[48] L. Shen and F. Makedon, "Spherical mapping for processing of 3D closed surfaces," *Image Vis. Comput.*, vol. 24, pp. 743–761, July 2006.

[49] J.-P. Antoine and P. Vandergheynst, "Wavelets on the 2-sphere: A group-theoretical approach," *Appl. Computat. Harm. Anal.*, vol. 7, pp. 262–291, Nov. 1999.

[50] D. W. Ritchie and G. J. L. Kemp, "Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces," *J. Computat. Chem.*, vol. 20, no. 4, pp. 383–395, 1999.

[51] R. J. Morris, R. J. Najmanovich, A. Kahraman, and J. M. Thornton, "Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons," *Bioinformatics (Oxford, England)*, vol. 21, pp. 2347–2355, May 2005.

[52] M. Styner, I. Oguz, S. Xu, C. Brechbuhler, D. Pantazis, J. Levitt, M. Shenton, and G. Gerig, "Framework for the statistical shape analysis of brain structures using SPHARM-PDM," *Insight J.*, vol. 1071, pp. 242–250, 2006.

[53] S. Singh, F. Janoos, T. Pecot, E. Caserta, K. Huang, J. Rittscher, G. Leone, and R. Machiraju, "Non-parametric analysis of cell phenotypes," in *Proc. MICCAI*, part II, pp. 343–351, 2011.

[54] B. T. T. Yeo, W. Ou, and P. Golland, "On the construction of invertible filter banks on the 2-sphere," *IEEE Trans. Image Processing*, vol. 17, pp. 283–300, Mar. 2008.

[55] M. Niethammer, M. Reuter, F. Wolter, S. Bouix, N. Peinecke, M.-S. Koo, and M. Shenton, "Global medical shape analysis using the Laplace-Beltrami spectrum," in *Proc. MICCAI*, 2007, vol. 10, pp. 850–857.

[56] P. Rämö, R. Sacher, B. Snijder, B. Begemann, and L. Pelkmans, "CellClassifier: Supervised learning of cellular phenotypes," *Bioinformatics*, vol. 25, pp. 3028–3030, Nov. 2009.

[57] M. Held, M. H. A. Schmitz, B. Fischer, T. Walter, B. Neumann, M. H. Olma, M. Peter, J. Ellenberg, and D. W. Gerlich, "CellCognition: Time-resolved phenotype annotation in high-throughput live cell imaging," *Nat. Methods*, vol. 7, no. 9, pp. 747–754, 2010.

[58] G. Pau, F. Fuchs, O. Sklyar, M. Boutros, and W. Huber, "EBImage—An R package for image processing with applications to cellular phenotypes," *Bioinformatics*, vol. 26, pp. 979–981, Apr. 2010.

[59] F. de Chaumont, S. Dallongeville, N. Chenouard, N. Hervé, S. Pop, T. Provoost, V. Meas-Yedid, P. Pankajakshan, T. Lecomte, Y. Le Montagner, T. Lagache, A. Dufour, and J.-C. Olivo-Marin, "Icy: An open bioimage informatics platform for extended reproducible research," *Nat. Methods*, vol. 9, no. 7, pp. 690–696, 2012.

[60] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: An open-source platform for biological-image analysis," *Nat. Methods*, vol. 9, pp. 676–682, July 2012.

[61] J. Ollion, J. Cochennec, F. Loll, C. Escudé, and T. Boudier, "TANGO: A generic tool for high-throughput 3D image analysis for studying nuclear organization," *Bioinformatics (Oxford, England)*, vol. 29, pp. 1840–1841, July 2013.

[62] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4. New York: Springer, 2006.

[63] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. New York: Springer, 2009.

[64] X. Chen, X. Zhou, and S. T. C. Wong, "Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy," *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 762–766, Apr. 2006.

[65] E. Segal, C. B. Sirlin, C. Ooi, A. S. Adler, J. Gollub, X. Chen, B. K. Chan, G. R. Matcuk, C. T. Barry, H. Y. Chang, and M. D. Kuo, "Decoding global gene expression programs in liver cancer by noninvasive imaging," *Nat. Biotechnol.*, vol. 25, pp. 675–680, June 2007.

[66] C. Sommer, C. Straehle, U. Kothe, and F. A. Hamprecht, "Ilastik: Interactive learning and segmentation toolkit," in *Proc. IEEE Int. Symp. Biomedical Imaging*, 2011, pp. 230–233.

[67] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[68] C. Ducroz, J.-C. Olivo-Marin, and A. Dufour, "Spherical harmonics based extraction and annotation of cell shape in 3D time-lapse microscopy sequences," in *Proc. IEEE Engineering in Medicine and Biology Conf.*, Boston, MA, 2011, pp. 6619–6622.

[69] Q. Zhong, A. G. Busetto, J. P. Fededa, J. M. Buhmann, and D. W. Gerlich, "Unsupervised modeling of cell morphology dynamics for time-lapse microscopy," *Nat. Methods*, vol. 9, pp. 711–713, July 2012.

[70] K. Masood, N. Rajpoot, K. Rajpoot, and H. Qureshi, "Hyperspectral colon tissue classification using morphological analysis," in *Proc. Int. Conf. Emerging Technologies*, 2006, pp. 735–741.

[71] B. Neumann, T. Walter, J.-K. Hériché, J. Bulkescher, H. Erfle, C. Conrad, P. Rogers, I. Poser, M. Held, U. Liebel, C. Cetin, F. Sieckmann, G. Pau, R. Kabbe, A. Wünsche, V. Satagopam, M. H. A. Schmitz, C. Chapuis, D. W. Gerlich, R. Schneider, R. Eils, W. Huber, J.-M. Peters, A. A. Hyman, R. Durbin, R. Pepperkok, and J. Ellenberg, "Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes," *Nature*, vol. 464, no. 7289, pp. 721–727, 2010.

[72] K. W. Eliceiri, M. R. Berthold, I. G. Goldberg, L. Ibáñez, B. S. Manjunath, M. E. Martone, R. F. Murphy, H. Peng, A. L. Plant, B. Roysam, N. Stuurman, N. Stuurmann, J. R. Swedlow, P. Tomancak, and A. E. Carpenter, "Biological imaging software tools," *Nat. Methods*, vol. 9, pp. 697–710, July 2012.

[73] C. Conrad, A. Wünsche, T. H. Tan, J. Bulkescher, F. Sieckmann, F. Verissimo, A. Edelstein, T. Walter, U. Liebel, R. Pepperkok, and J. Ellenberg, "Micropilot: Automation of fluorescence microscopy-based imaging for systems biology," *Nat. Methods*, vol. 8, pp. 246–249, Mar. 2011.

[74] T. Y. Liu, A. Wiesel, and A. O. Hero, "A sparse multiclass classifier for biomarker screening," in *Proc. IEEE Global Conf. Signal and Information Processing (GlobalSIP)*, 2013, pp. 77–83.

[SP]