ShapePheno: unsupervised extraction of shape phenotypes from biological image collections

Theofanis Karaletsos¹,*, Oliver Stegle¹,*, Christine Dreyer², John Winn³ and Karsten M. Borgwardt¹,4
¹ Machine Learning and Computational Biology Research Group, Max Planck Institute for Intelligent Systems and Max Planck Institute for Developmental Biology, ²Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany, ³Machine Learning and Perception Group, Microsoft Research Ltd, Cambridge CB3 0FB, UK and ⁴Zentrum für Bioinformatik, Eberhard Karls Universität, 72076 Tübingen, Germany

ABSTRACT

Motivation: Accurate large-scale phenotyping has recently gained considerable importance in biology. For example, in genome-wide association studies technological advances have rendered genotyping cheap, leaving phenotype acquisition as the major bottleneck. Automatic image analysis is one major strategy to phenotype individuals in large numbers. Current approaches for visual phenotyping focus predominantly on summarizing statistics and geometric measures, such as height and width of an individual, or color histograms and patterns. However, more subtle, but biologically informative phenotypes, such as the local deformation of the shape of an individual with respect to the population mean cannot be automatically extracted and quantified by current techniques.

Results: We propose a probabilistic machine learning model that allows for the extraction of deformation phenotypes from biological images, making them available as quantitative traits for downstream analysis. Our approach jointly models a collection of images using a learned common template that is mapped onto each image through a deformable smooth transformation. In a case study, we analyze the shape deformations of 388 guppy fish (Poecilia reticulata). We find that the flexible shape phenotypes our model extracts are complementary to basic geometric measures. Moreover, these quantitative traits assist the observations into distinct groups and can be mapped to polymorphic genetic loci of the sample set.

Availability: Code is available under: http://bioweb.me/GEBI

Contact: Theofanis.karaletsos@tuebingen.mpg.de; oliver.stegle@tuebingen.mpg.de

Supplementary information: Supplementary data are available at Bioinformatics online.

Received on September 22, 2011; revised on February 7, 2012; accepted on February 9, 2012

1 INTRODUCTION

With the advent of high-throughput genotyping techniques an unprecedented breadth of genotypic datasets can be generated, opening doors to large-scale association studies, promising sufficient power to understand the genetic underpinning of more subtle phenotypes that characterize the sample. As phenotyping often requires manual labor and expert knowledge, a major bottleneck now lies with the identification and quantification of informative traits. Currently, the quantification of phenotypic traits is predominantly done in a semi-manual fashion, rendering the task of analyzing large datasets expensive, time-consuming and error-prone. In order to address these shortcomings, the automated analysis of biological images has become a staple in modern biology.

High-throughput imaging techniques for various types of microscopy and other imaging modalities have become common in the experimental environment. Automated image analysis for bioimaging attempts to deal with the flood of data and subsumes a large variety of tasks and methods; for a comprehensive review, see Peng (2008) and Walter et al. (2010). Common tasks include the counting of cells in microscope images and differential analysis of distinct cell types (Buchs et al. 2014; Eau et al. 2014). Key challenges in bioimage informatics stem from the breadth and individuality of natural variation within these images and dealing with the inherent noise in biological imaging tasks. In order to deal with these factors, machine learning techniques have raised considerable attention and are used to tackle various complicated tasks in realistic settings (Ning et al. 2009; Shamir et al. 2011).

For example, in the analysis of appearance phenotypes machine vision has been used to quantify the extent of existence of predefined visual features or detect interesting appearance features that characterize the data (Whibley et al. 2009). Visual appearance features usually pertain to specific local properties of the depicted objects. However, more general visual phenotypes often are also biologically informative, such as the description of the shape of an object and the quantification of global (including size and height) as well as local (i.e. locally deformed parts of an image) shape variations. An example where such a method is useful is the characterization of the shapes of guppy fish, which so far can only be analyzed by labor-intensive manual geometric phenotype measurements on hundreds of fish, as performed in Tripathi et al. (2009).

Our goal is to automatically determine and quantify differences among observed shapes in biological images in order to interpret them as shape phenotypes and facilitate downstream analysis, for instance association tests of traits with putative causal factors in the genome. In this work, we propose an unsupervised machine learning method to quantify shape variations of a given object class depicted...
guppy fish, human labor is costly and error-prone. Obtaining accurate non-trivial measurements on large datasets where interpretable features and results. Thus, our approach facilitates unannotated images in a fully unsupervised fashion while retaining discoveries and objectively quantifies deformation phenotypes on selection/binarization on each image required. Instead, ShapePheno priori known are aligned to a template. In contrast to previous studies, our method does not require explicit knowledge of the template a priori; neither is supervision like setting of landmarks or outline selection/binarization on each image required. Instead, ShapePheno discovers and objectively quantifies deformation phenotypes on unannotated images in a fully unsupervised fashion while retaining interpretable features and results. Thus, our approach facilitates obtaining accurate non-trivial measurements on large datasets where human labor is costly and error-prone.

In Section 2, we present a case study of our method on guppy fish, Poecilia reticulata. The individuals in this dataset are subject to variation in appearance and shape. Interestingly, both appearance and shape variability have previously been shown to exhibit considerable genetic components (Tripathi et al., 2009). In Section 2.1, we discuss how a graphical model based on Markov random fields (MRFs) can be used to simultaneously learn the unknown template and recover smooth mapping fields, performing a flexible variant of deformable registration. We decompose the mapping fields into a sum of a technical translation component and shape-related deformation fields, both specific to each image. The overall setup of our probabilistic model largely follows the jigsaw model (Kannan et al., 2009). Under this model, a set of $N$ observed images $I_i, i=1,...,N$ is explained by a common latent template image $T$. The training images are explained as function of the template through learned mapping vectors $L_i$ between pixels in the template and observed pixels in each image $i$. The coordinate mapping accounts for an overall shift of the image with respect to the template, as well as local deformations, compressing or stretching specific parts of each image to match the common reference (Section 2.2). Subsequently, once the template and the deformation fields are learned, we extract quantitative traits from the information captured in the deformation fields. For this purpose, we employ linear dimensionality reduction (Section 2.3), yielding a compact set of features that explain the major axes of variation in the deformation fields for each image. For comparison, we also show how our model can be used to quantify length vectors within images, which can be directly related to established manual measurements of shape traits. Both types of features can be used for downstream analyses.

Summary overview of model parameters: In our model, we assume a set of $N$ images $I_i$ and corresponding mapping fields $L'_i$ of dimension $(d_x \times d_y)$ for
The size of template image parameterized by an energy function for translation (\( E_T \)) and the smoothness prior are given as: the prior on the translation component of \( E_T \) is a function of mapping distance. At maximum allowed deformation \( \rho \), the smooth deformation fields \( L_i \) are decomposed into a translation and deformation component with individual smoothness priors. The specific modeling choices will be discussed in Section 2.2.

The joint probability under our model is

\[
P(T, L^T | \mu_T, \nu_T, \omega_t, \omega_d, \omega_L, \alpha, \beta) = P(T) \prod_{i=1}^{N} P(T_i | \mu_T, \nu_T) P(L_i | \omega_t, \omega_d, \omega_L, \alpha, \beta) .
\]

The likelihood of the observation model corresponding to Equation (4) is a Gaussian mixture model independent for each image pixel \((x, y)\):

\[
P(T_i | \mu_T, \nu_T) = \sum_{i=1}^{N} \mathcal{N}(\mu_T(x_i, y_i), \nu_T(x_i, y_i)) \text{ Normal-Gamma}.
\]

Smoothness of the mapping fields \( L_i \) is encouraged through the choice of a Markov random field prior that couples neighboring mapping offsets in each image

\[
P(L_i) \propto \exp \left( - \sum_{(x,y), (x',y') \in N} E(l_i(x,y), l_i(x',y')) \right).
\]

2.2 Design choices for MRF energy functions

Since the alignment of template and observed images requires a combination of translation and deformation, we choose \( L_i \) to be the sum of a translation \( T_i \) and a deformation \( D_i \). In this stacked two-layer Markov random field, \( P(L_i) \) accounts for the global shift of images to the template and \( P(D_i) \) specifies the prior probability of local deformations (see also Fig 2b). The joint prior probability of the mapping-field components can be expressed as

\[
P(L_i, T_i, D_i) = P(T_i) P(L_i | T_i) P(D_i | T_i, L_i) = P(T_i) P(L_i | T_i) P(D_i | T_i, L_i).
\]

where \( \delta(x) \) denotes the Dirac delta function. Accounting for both prior contributions, the effective energy term in Equation (4) becomes \( E_T = E_T + E_D \). Inference in the full model is done iteratively within the mapping updates, by first keeping \( L_i = 0 \) fixed and updating \( T_i \). Next, \( L_i \) is kept at the learned value while updating \( T_i \). Both update steps can be done following the standard jigsaw inference (Section 2.2 and Kannan et al. 2007). In the following, we will explain the modeling choices of each mapping prior separately.
2.2.1 Translation (rigid) model: Having defined a template that is of equal size as the images, the goal is to register images to it. In order to allow the shift field $L_i$ to incorporate translation behavior, we employ a Pott's Model prior with energy function $E_{\text{V}}(L_i) = \beta_0 \delta_0 (L_i)$. Here, the cost parameter $\gamma_0$ is set to large value, such that all mappings $L_i$ are forced to take on identical values, solely accounting for a constant overall image shift.

2.2.2 Deformable model: For the deformable prior $P(L_i)$, we employ a non-rigid smoothness prior that encourages smooth deformation fields. In contrast to the rigid Pott's model, the energy costs is distant-dependent, favoring short-range deformations. More specifically, the energy function $E_{\text{V}}$ scales linearly with a particular choice of distance norm of order $\alpha$.

$$E_{\text{V}}(L_i) = \frac{1}{2} \frac{\gamma}{\alpha} \sum_{i} \| L_i - L_0 \|_a^2 \quad \text{for} \quad |L_i - L_0| \leq \rho$$

Here, $\rho$ denotes the maximum permitted range of deformation, and $\alpha$ is a (or order) where $\alpha \in [1, 2]$ for $\alpha \geq 2$ and scaling parameter $\gamma_0 = \frac{\gamma}{\alpha}$. Intuitively, one can imagine this function to apply elastic bands connecting pixels (in our case points) to the various components with the $E_{\text{V}}$ part of the mapping costs being the equivalent of the elastic potential of the bands between all pieces. Figure 2c shows the energy function for two choices of $\alpha$.

2.2.3 Robustifying deformable registration: We use various constraints on the registration to further improve the robustness of our method against noise and non-standardized images and reliably produce good results. We apply our deformation fields on pixel blocks, meaning that we constrain groups of pixels of block-size to obtain the same mapping via the prior $\tilde{V}$ shown in Figure 4b. This leads to piecewise smooth deformation fields. Additionally, we constrain the parameter $\rho$ of the deformation field itself. This makes large jumps prohibitively expensive and drives the model to use smoothly varying local deformation patches. A positive side effect of this constraint is a significant boost in computational efficiency and robustness against outlier-mistakes since the solution space is reduced. We also robustify alignments against appearance outliers with a mixture of densities used as the observation model in Equation 2, which allows for a background class.

2.3 Feature representation for deformation maps

The deformation fields $L_i$ at pixel resolution, described in Section 2.2, capture the relevant information to explain local shape deformations of the samples in each image. Comparing deformation fields with non-equal objects at non-equal positions is hard, since we face the problem of correspondence. However, in our framework this problem can be elegantly circumvented using the common template all images are aligned to. To render individual deformation maps comparable between images, we first project these maps from observation space into the common reference coordinate system on the template. For this purpose, we apply the same mapping operator we used for image pixels now to the mappings themselves $D'_{SP} = L_i$, resulting in deformation field representations in the object-centered template space. Thus, we obtain an easily interpretable shift-corrected field $D'$ of deformation for each image-mapping $L_i$.

2.3.1 Low-rank representations of deformation fields: In order to extract meaningful features from the high-dimensional deformation fields, we reduce their dimensionality by representing individual deformation fields via a set of coefficients over a small number of bases $\Phi$ obtained via standard principal component analysis (PCA) (Fig 3a). Prior to running PCA, we can apply a binary mask to the template to select only relevant template regions for consideration as variance components. The resulting individual bases can be visualized and constitute local deformation factors.
This cross (157 Quare × Cumaná) has been subject of a genotyping project to establish a comprehensive genetic map and to initiate conventional QTL (quantitative trait locus) mapping. The raw images were rescaled such that the ratio of pixels to physical length is constant across the dataset. Due to rescaling, the image size was variable, ranging between 75 × 226 and 83 × 250 pixels size. To account for images taken at slightly varying distance, we chose to embed all images according to original size of the fish into empty images of the chosen format (83 × 250). For this particular experiment, there were few outliers and thus setting the outlier ratio \( \pi = 0 \) yielded good results. For each of the 388 individuals, the dataset included matching genotype information, covering a total of 1063 genome-wide single nucleotide polymorphisms (SNPs). After filtering, removing rare SNPs with a minimum rare allele frequency <5%, we obtained 814 polymorphisms that were considered for analysis.

### 3.2 Experimental settings

We chose the following parameters for the deformation model: \( \gamma_0 = 40, \rho = 10 \) and \( \alpha = 1 \), which resulted in robust registration in a series of test runs. We applied the deformation field to 2 × 2 pixel blocks in order to locally tie together image pixels to correct for appearance differences and to prevent excessive local deformation. The normal-gamma prior parameters (Equation (5)) were set such that the prior reflects the first and second empirical moments of the distribution of the raw image pixels (see also Kannan et al. 2009). We ran a Python-based parallelized implementation of ShapePheno on an 8-core Intel Xeon machine where the full dataset could be run to convergence within 3 days. After convergence, we manually segmented the template fish from the background template to facilitate all downstream analyses (clustering and association mapping on foreground information only).

### 3.3 Shape phenotype determination

The converged ShapePheno model yielded a sharp template that resembles an average fish and mapping fields \( L_i \) for every image in the dataset (Fig. 2). The model perceives the shapes of the fish in individual images as locally stretched or smoothly distorted versions of the template and smoothly bypasses appearance differences that would counteract shape alignment. This suggests that the deformation fields \( L_i \) capture shape information corrupted by noise stemming from the difference in sizes of images and the background color similarity to the fish corpora. Next, we used linear dimension reduction (Section 2.3) to determine the corresponding deformation factors of the converged model. Figure 2 depicts the first three PCA-bases. These three main deformation features appear to divide the fish into the anterior and posterior part. Inflated anterior parts at the belly region as well as distorted posterior trunks are the main sources of shape variation. We also observed that local structure in the bases matches appearance features of the template that get distorted frequently. These findings reflect the set-up of the experiment in Tripathi et al. 2009, in agreement with our expectations, as the parents were originally chosen to exhibit these shape differences and the offspring shows strong variation at these features. Supplementary Figure S1 provides examples of inference results for extreme outliers within the data, here a singleton shape-mutant in our training set. Since the method is unsupervised, it requires shape mutants to be well-represented in the data in order to model their shape accurately.

### 3.4 Quantification of geometric measurement accuracy

After the qualitative evaluation of the reconstructed shape template, we next characterized the accuracy of the shape representation captured by the model in a quantitative manner. For this purpose, we used the converged model to automatically measure geometric distances in images (Section 3.2). We comparatively evaluated eight geometric trait measurements (described in Figs 1 and 2), whose choice was motivated by primary analyses of the Guppy dataset Tripathi et al. 2009. Manual quantification was done on 50 individuals from our dataset chosen at random, measuring all 8 geometric distances in each raw image by 3 independent experts, as well as using the fully automated approach provided by the ShapePheno model. We assessed the correlation between manual and automated measurements, comparing the ShapePheno prediction to the mean of the manual quantification runs (Fig. 3b). Encouragingly, all automated geometric measurements were in good agreement with the corresponding manual annotation. The correlation score for pairs of corresponding automated and manual measurements ranged between 0.65 (A2 versus M2) and 0.84 (A6 versus M6) with a mean correlation score of 0.76.

To better understand the magnitude of the variation between automated and manual measurements, we also considered the pairwise correlation between two of the three manual runs (Fig. 5b), yielding comparable results. Pairwise correlations here ranged from 0.81 (M7) up to 0.96 (M1) with a mean correlation score of 0.87. This suggests that ShapePheno captures true variability in images and yields high levels of accuracy when used to quantify geometric measurements in place of a human expert. Detailed scatter plots, showing the correlations between manually and automatically determined traits are shown in Figure 2.

From either of the correlation analyses, it was also notable that geometric measurements correlated well with each other, reflecting the biological relatedness of growth-phenotypes that underlie the geometric measurements under consideration. In contrast to this observation, the correlation to the new PCA-deformation phenotypes described in Section 3.5 was weak, which shows that they capture orthogonal aspects of shape variation and hence are complementary to geometric measurements.

### 3.5 Clustering of populations based on deformation traits

We clustered populations of guppy fish according to their characteristic local deformation patterns, without any prior
quantitative traits complement established measurements. Weak correlation between PCA-L1–L8). The results show that the reproducibility of manual expert labels is between two manual quantification runs by an expert user (M1–M8 versus L1–L8). The results showed the reproducibility of manual expert labels is similar to automated measurements by ShapePheno, suggesting the model is able to achieve human-like results. Weak correlation between PCA-deformation phenotypes and geometric phenotypes shows that these new quantitative traits complement established measurements.

Fig. 5. Quantitative evaluation and comparison of the geometric traits as determined by ShapePheno (A1–A8), manually measured counterparts (B1–B8), and novel PCA-deformation phenotypes (PX1–PX8). (a) Correlation between automatic geometric measurements, manual measurements and PCA-deformation phenotypes. (b) Correlation between two manual quantification runs by an expert user (M1–M8 versus L1–L8). The results showed the reproducibility of manual expert labels is similar to automated measurements by ShapePheno, suggesting the model is able to achieve human-like results. Weak correlation between PCA-deformation phenotypes and geometric phenotypes shows that these new quantitative traits complement established measurements.

Fig. 6. Scatter plots with error bars, showing the relationship between manual measurements and automated geometric measurements as obtained by ShapePheno. Shown is data for each of the 8 length phenotypes, where each point corresponds to an instance of the 50 samples chosen for quantification. Error bars show ±1SD, estimated separately for each quantification approach. For manual quantification, error bars correspond to the empirical variation between three independent annotation runs. For automated quantification, uncertainty estimates stem from the variation of the 15 nearestneighbor assignments on the template to the selected point and measuring their SD. The green diagonals show the expected ideal correlation.

knowledge of their genetic constitution. Morphometric prototypes for the guppy have previously been determined from hand-annotated images and correlated to sex and environmental factors (Hendry et al., 2009). We clustered deformation fields according to a linear kernel between low-rank projections of D′ (as described in Section 3.3.1) using affinity propagation (Frey and Dueck, 2007), a non-parametric clustering technique that uses deformation kernel values as inputs and yields a flexible number of clusters |C|. Reconstructing the mean low-rank vector field of each cluster given by its embedding in deformation space yields cluster-specific morphological deformation bases. Figure 7 provides a comparative overview of three characteristic clusters, indicating that independent factors can influence the shape of the anterior and posterior trunk of the examined guppy fish. Deformation bases correspond to low-rank projections of the cluster means, where only the characteristic local deformations per cluster are considered as shared elements of cluster members. The shape seen in an example image representing the median deformation field per cluster corresponds to our expectation given the profiles. Different clusters portray significant variability between their profiles, such as the regional focus and the extent of expected local deformation.

3.6 Association study of shape factors to genotype

Finally, we performed a genome-wide association study using the previously learned phenotypes and their measurements. The phenotypic measurements y are the per-image coefficients w′ of PCA deformation bases Φ (Sections 3.1 and 3.2). We used a linear model that assesses how well a particular phenotypic value is modeled when genetic factors are taken into account, compared to when they are ignored. The relevant quantity is the log-odds (LOD) score,

\[
\log_{10} \left( \prod_{j=1}^{N} \frac{P(y_j | \theta)}{P(y_j | \theta_{bck})} \right)
\]

where \( x_j \) is a SNP measurement and \( y_j \) the phenotypic expression value for the \( j \)-th individual. The terms \( \theta, \theta_{bck} \) are parameters for the genetic and background models, respectively. We thus obtain LOD score plots over a large genomic region to obtain an association plot. We used Storey’s method (Storey and Tibshirani, 2003), a variant of Benjamini Hochberg, to assess genome-wide significance.

Although the available data has sparse genetic marker coverage, we still obtained statistically meaningful peaks as can be seen in Figure 8. Previous genetic QTL mapping in overlapping data has suggested markers of the proximal region of linkage group 12 (LG12) as relevant for size and body shape traits in male guppies, and in addition phenotypic sex has impact on these traits (Tripathi et al., 2009). Among the significant hits in our mapping, Markers 398 (lod 7.7 on LG12) and 442 (lod 10.3, LG12) are found in the proximal region of LG12 while marker 229 (lod 11.9, LG12) is the most distal and closest to the putative male sex-determining locus. Depending on the trait analyzed, significant QTL were suggested within a region spanning ~6 cM (~7 Mb) (Tripathi et al., 2009) in cross 157. Marker 442 was supported as a QTL for area of the posterior trunk for cross 158 (Tripathi, 2009). Additional loci were detected with good statistical support, in agreement with the observation that co-factors on various linkage groups contribute to complex traits.

4 DISCUSSION

We have proposed a generative probabilistic model that extracts deformation phenotypes by registering images to a latent, learned template in an unsupervised fashion. Our method presents a novel, clean framework for researchers to quantify and describe subtle local deformation patterns and use them for downstream analyses, like clustering or genetic association tests. We applied our method to a bioimaging task, where we discovered significant deformation patterns in images of guppy fish. We also showed that ShapePheno can be used for automated quantification of geometric measurements and showed good correspondence to manually labeled data.

More important than accurate geometric traits, ShapePheno yielded deformation fields that characterize the variability in shape and could be used to identify low-rank PCA factors...
of shape variability. While simple distance measurements inter-
correlate strongly, the deformation phenotypes we propose describe
orthogonal shape factors and are thus novel holistic descriptors
of shape. We showed practical utility of these PCA-deformation
phenotypes in the context of clustering, grouping the data into
clusters exhibiting characteristic deformation. We also performed a
GWAs with the same traits, which yielded biologically sound results
in agreement with previous results on geometric approximations of
shape (see Tripathi et al. (2009) and unpublished observations of
C.D.). We are convinced that comprehensive genomic analyses on
larger datasets can be performed by using this method with a rigorous
treatment of image acquisition, higher image resolution and higher
marker density.

Unsupervised extraction and quantification of subtle
morphological phenotypes, as done here, is the logical next
step in automated image analysis. The relevance of these new types
of methods is expected to rise quickly as dataset sizes increase,
providing the necessary statistical power to identify and quantify
complex phenotypic variation.

Funding: T.K. was supported by a Microsoft Research Cambridge
stipend and O.S. was supported by a fellowship from the Volkswagen
Foundation.

Conflict of Interest: none declared.

REFERENCES

divergence in the cuman guppy—a possible case of incipient speciation. J. Evolution.
Biol., 17, 1238–1254.

New York.


Science, 315, 972–976.

Fuchs, F. et al. (2010) Clustering phenotype populations by genome-wide RNAs and

Hendry, A.P. et al. (2006) Parallel evolution of the sexes? Effects of predation and
habitat features on the size and shape of wild guppies. J. Evolution. Biol., 19,
741–754.

Kannan, A. et al. (2007) Clustering appearance and shape by learning jigsaws. In
Scholkopf, B. et al. (eds) Advances in Neural Information Processing Systems. MIT

Ning, F. et al. (2005) Toward automatic phenotyping of developing embryos from videos.

Pau, G. et al. (2010) Ebimage—an R package for image processing with applications to
cellular phenotypes. Bioinformatics, 26, 979–981.

24, 1827–1836.

Saafedini, S. et al. (2010) As-rigid-as-possible mosaicking and serial section registration
of large system datasets. Bioinformatics, 26, 357–363.

T. Karaletsos et al.


