Applications in Plant Sciences

Applications in Plant Sciences 2014 2(8): 1400030

Application Article

A neotropical Miocene pollen database employing image-based search and semantic modeling[1]

Jing Ginger Han[2], Hongfei Cao[3], Adrian Barb[4], Surangi W. Punyasena[5], Carlos Jaramillo[6], and Chi-Ren Shyu[2,3,7]

[2]Informatics Institute, University of Missouri, Columbia, Missouri 65211 USA; [3]Department of Computer Science, University of Missouri, Columbia, Missouri 65211 USA; [4]Department of Information Science, Pennsylvania State University–Great Valley, Malvern, Pennsylvania 19355 USA; [5]Department of Plant Biology, University of Illinois, Urbana, Illinois 61801 USA; and [6]Center for Tropical Paleoecology and Archaeology, Smithsonian Tropical Research Institute, Apartado 0843, 03092 Balboa, Ancón, República de Panamá

- *Premise of the study:* Digital microscopic pollen images are being generated with increasing speed and volume, producing opportunities to develop new computational methods that increase the consistency and efficiency of pollen analysis and provide the palynological community a computational framework for information sharing and knowledge transfer.
- *Methods:* Mathematical methods were used to assign trait semantics (abstract morphological representations) of the images of neotropical Miocene pollen and spores. Advanced database-indexing structures were built to compare and retrieve similar images based on their visual content. A Web-based system was developed to provide novel tools for automatic trait semantic annotation and image retrieval by trait semantics and visual content.
- *Results:* Mathematical models that map visual features to trait semantics can be used to annotate images with morphology semantics and to search image databases with improved reliability and productivity. Images can also be searched by visual content, providing users with customized emphases on traits such as color, shape, and texture.
- *Discussion:* Content- and semantic-based image searches provide a powerful computational platform for pollen and spore identification. The infrastructure outlined provides a framework for building a community-wide palynological resource, streamlining the process of manual identification, analysis, and species discovery.

Key words: content-based image retrieval; database; Miocene; pollen morphology; semantics.

Palynologists use the morphological characteristics of pollen and spore grains to identify, classify, count, compare, and log plant diversity within geologic samples from different geographical locations and ages. These data are used to address research questions in areas such as biostratigraphy, paleoecology, biodiversity, climate change, taxonomy, and evolution, and are even increasingly employed in forensics. The potential sample size represented by a fossil pollen sample can be very large, because hundreds to thousands of grains can be preserved in a drop of pollen residue extracted from a geological sample (rock or sediment); however, the classification of samples is still primarily qualitative and manual, based on the visual identification of key morphological features, and requires significant experience and expertise (Faegri et al., 1989; Traverse, 2007).

This manual, intuitive approach to classification (sensu Birks and Peglar, 1980) potentially results in discrepancies in taxonomic identifications due to individual differences in analysts' interpretation of morphological details, familiarity or experience with a given suite of taxa, fatigue, and preservation of fossil pollen material. Morphological similarity among related taxa may also decrease the taxonomic precision of identifications, due to the inability to observe or to define morphological differences (Mander and Punyasena, 2014). Moreover, the intrinsic morphological variability found within pollen grains from even the same species makes it difficult to assess the morphological boundaries of any given fossil species. There are few published studies of how much morphological difference can be consistently recognized among analysts (e.g., Mander et al., 2014). As a result, the recognition and formal naming of new morphotypes rely on a certain degree of consensus from a community of experts. However, with advanced imaging technology, digital microscopic pollen images are being generated with increasing speed and volume, producing opportunities to improve upon the traditional manual identification and sorting of grains and to produce higher throughput approaches to pollen analysis.

There are several public databases and software applications that have been developed to assist palynologists in their identifications. For example, Bush and Weng (2007) designed a downloadable neotropical pollen database as a freeware for neotropical palynology researchers. It provides multiple-access keys to query the database with flexibility

[1]Manuscript received 26 March 2014; revision accepted 1 July 2014.

This research is supported by the National Science Foundation (grant numbers: DBI-1053024 to C.R.S., DBI-1052997 to S.W.P., and EAR-0957679 to C.J.). We thank Alejandra Restrepo, Ingrid Romero, and Carlos D'Apolito for their contributions to our pollen image database. Jacklyn Rodriguez assisted in organizing the sematic labels.

[7]Author for correspondence: shyuc@missouri.edu

doi:10.3732/apps.1400030

 © 2014 Han et al. Published by the Botanical Society of America. This work is licensed under a Creative Commons Attribution License (CC-BY-NC-SA).

and tolerance in missing data attributes. The collection contains pollen images, primarily taken with transmitted light microscopes, from more than 1000 neotropical species. Morphological features, such as pollen shape, pore shape, reticulum shape, and pollen size, can be used to query the database. A second pollen image database, PalDat (http://www.paldat.org), has a similar query structure and Web-based interface and includes both transmitted light images and scanning electron microscopy (SEM) images from ~2200 modern species and 32 fossil ones. Neotoma Paleoecology Database (http://www.neotomadb.org) is another example that provides palynological, paleontological, geological, and geographic data for Pliocene through Holocene sites based on information submitted by collaborating individuals from multiple institutions. Its main purpose is to map spatiotemporal taxa distribution (Grimm et al., 2013).

While these image databases and software applications serve as valuable resources for pollen identification, having to manually label and compare morphology is both time-consuming and subject to the idiosyncrasies of individual analysts. Automated visual content extraction allows analyses to be kept more consistent across multiple sites, and is especially useful when there is an unknown sample with new morphotypes that needs to be compared against existing collections. Previous applications of machine-based classifications for pollen identification have focused on the accuracy of the end classification (Holt and Bennett, 2014) and generally do not provide a mechanism for establishing the community-level consensus of identifications that is required when working with extinct species. For example, Classifynder, developed at Massey University, is a stand-alone system that provides a framework for image acquisition and classification of modern pollen materials (Holt et al., 2011). Its experiments suggested that computer performance in pollen identification and classification was comparable to human experts, but with better consistency. This work can be further extended to image search using extracted visual features of identified grains in both modern and extinct species. Developing a broader platform for capturing and sharing expert knowledge builds on previous machine learning and image database efforts and provides a pathway for making these tools widely accessible.

This study is the result of a long-term collaboration among palynologists, computer scientists, and informaticians in an attempt to develop computational and informatic solutions to streamline the process of palynology analysis for efficient and reliable data management, analysis, and retrieval. To our knowledge, our work is the first attempt to develop an intelligent search engine that utilizes image-based morphological content for grain image retrievals in palynology. We report the following approaches.

First, we applied and extended a suite of image analysis algorithms and toolkits to automate the process of detecting grains from artifacts (debris and organic matter other than pollen and spores, common to fossil palynological slides) and calculated morphological features based on shape and texture. Next, association rule mining (Agrawal et al., 1993) was integrated into our methods to assist experts in trait annotation based on extracted features and a continuously updated expert knowledge base. We then used information retrieval methods (Baeza-Yates and Ribeiro-Neto, 1999) to provide fast and accurate data management and image retrieval. The morphological features identified by our automated analysis were used to determine image semantics (abstract presentations of morphology) that formed the basis of novel tools for automatic semantic annotation, semantic-based

Table 1. Data set of pollen and spore samples from neotropical Miocene.[a]

ID[b]	Taxon	Source	No. of grains	No. of images (mean)
Pollens				
1014	*Clavainaperturites microclavatus*	Hoorn, 1994	6	24 (4.0)
148	*Clavainaperturites clavatus*	Van der Hammen and Wymstra, 1964	7	22 (3.1)
246	*Echiperiporites estelae*	Germeraad et al., 1968	5	18 (3.6)
1430	*Echiperiporites scabrannulatus*	Silva-Caminha et al., 2010	7	24 (3.4)
365	*Grimsdalea magnaclavata*	Germeraad et al., 1968	5	21 (4.2)
254	*Malvacipolloides maristellae*	Müller et al., 1987; Silva-Caminha et al., 2010	7	25 (3.6)
450	*Mauritiidites franciscoi* var. *franciscoi*	Van der Hammen, 1956; Van Hoeken-Klinkenberg, 1964	9	46 (5.1)
451	*Mauritiidites franciscoi* var. *minutus*	Van der Hammen and Garcia, 1966	7	34 (4.9)
511	*Perisyncolporites pokornyi*	Germeraad et al., 1968	7	18 (2.6)
552	*Proxapertites psilatus*	Sarmiento, 1992	7	29 (4.1)
570	*Psilamonocolpites medius*	Van der Hammen, 1956; Van der Hammen and Garcia, 1966	7	35 (5.0)
571	*Psilaperiporites minimus*	Regali et al., 1974	5	19 (3.8)
688	*Retitrescolpites? irregularis*	Van der Hammen and Wymstra, 1964; Jaramillo and Dilcher, 2001	9	32 (3.6)
722	*Retitricolpites simplex*	Gonzalez Guzman, 1967	7	24 (3.4)
767	*Rhoipites guianensis*	Van der Hammen and Wymstra, 1964; Jaramillo and Dilcher, 2001	7	26 (3.7)
Spores				
43	*Echinatisporis muelleri*	Regali et al., 1974; Silva-Caminha et al., 2010	7	28 (4.0)
45	*Magnastriatites grandiosus*	Kedves and Sole de Porta, 1963; Dueñas, 1980	7	24 (3.4)
282	*Kuylisporites waterbolkii*	Belsky et al., 1965	7	25 (3.6)
44	*Crassoretitriletes vanraadshooveni*	Germeraad et al., 1968	6	28 (4.7)
46	*Polypodiisporites usmensis*	Van der Hammen, 1956; Khan and Martin, 1972	5	23 (4.6)

[a]This study collected 397 images of 102 grains from 15 pollen taxa and 128 images of 32 grains from five spore taxa. On average, there were 6.8 grains imaged per pollen taxon and 6.4 grains per spore taxon. There are 3.9 images per pollen grain and 4.1 images per spore grain.

[b]Taxon IDs were adopted from the Smithsonian Tropical Research Institute palynology database (Jaramillo and Rueda, 2013).

image search, and content-based image retrieval by image examples.

MATERIALS AND METHODS

***Pollen images*—**In this study, 525 images from Miocene-aged pollen and spore material were taken from a stratigraphic section of Falcon basin in Venezuela (Aguilera and Carlini, 2010; Quiroz and Jaramillo, 2010). These images represent the 15 pollen taxa and five spore taxa listed in Table 1. Morphological information for each of the taxa was collected from the Smithsonian Tropical Research Institute palynological database (Jaramillo and Rueda, 2013), which contains the morphological descriptions of ~2700 species of neotropical fossil pollen and spores. Images were taken using a Zeiss AxioImager microscope, Plan-Apochromat SF25 (63×, 1.4 NA, oil immersion) lens, and a Zeiss AxioCam ICc 3 digital microscope camera (Carl Zeiss Microscopy GmbH, Göttingen, Germany). This subset of taxa was selected based on its morphological diversity and sample availability at the time of the study. Because overlapping of grains and debris is not uncommon in palynological slides, each sample image was cropped roughly with the grain at the center without intentionally avoiding debris.

***Grain segmentation*—**There are multiple options of image analysis toolkits (Ibañez et al., 2003; Abramoff et al., 2004; Bradski and Kaehler, 2008) to roughly segment a centrally placed object from a field of view (Pham et al., 2000; Gonzalez and Woods, 2002; Armato and MacMahon, 2003; Russ, 2006). We used several of these methods to extract grains from the image background. More refined methods were then conducted to extract morphological features. We detail the process below.

The cropped images of individual grains are RGB color images. However, in computational analysis and machine vision research, this color system is not always the best configuration to represent how human observers perceive content and pattern. Therefore, we converted and separated the original RGB images into three single-channel images using the HSV (*hue*, *saturation*, and *value*) color system (Gonzalez and Woods, 2002). In each image channel, pixel values not only represent part of the color space, but also contribute to segmentation of objects (Ohta et al., 1980) and calculations, representations of advanced visual constructs, such as textural content and shape characteristics. Using only grayscale images limits the ability to segment objects of interest efficiently or extract underlying visual patterns that comprise the image content. While *value* images provided the viewer with detailed texture of the grain, *hue* and *saturation* images allowed us to discriminate between foreground objects and background. Because our goal in grain segmentation was to find reasonable contrast to recognize the grain contour, we merged the *hue* and *saturation* images to reconstruct an intermediate image that displayed better separation of grains from background using the following equation.

$$p_M = \left(\frac{p_H}{180} * w_H + \left(1 - \frac{p_S}{255}\right) * w_S\right) * 255, \qquad 0 \le p_m \le 255 \tag{1}$$

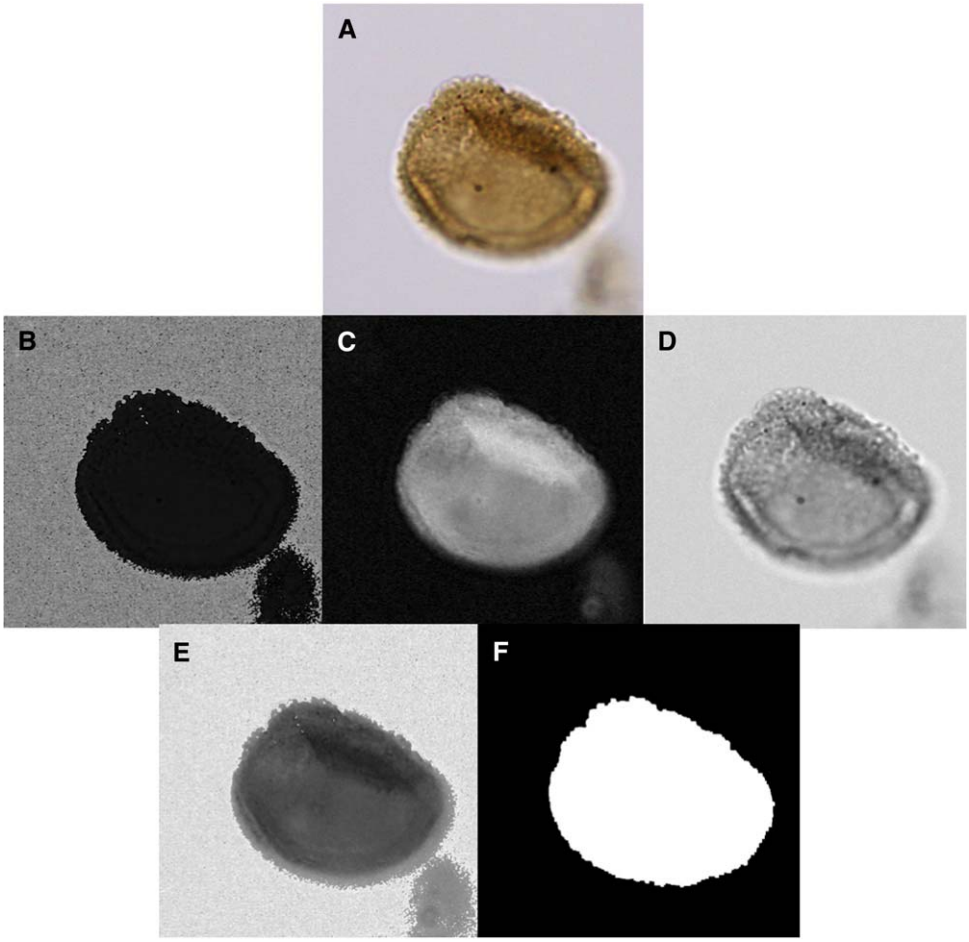


Fig. 1. An example pollen grain (*Clavainaperturites microclavatus*) image segmentation process. The original RGB image (A) is converted from a single RGB image to three single-channel images—hue (B), saturation (C), and value (D). (B) and (C) are then merged using selected weights on pixel values (Eq. 1) to generate an intermediate image (E) for thresholding, morphology operation, watersheding, and connected component operations. This ultimately segments the main grain object (F) from the rest of the image, including background pixels, trivial particles, and debris.

The merged image pixel value (p_M) is a weighted combination of pixel values from *hue* (p_H) and *saturation* (p_S) images at the same pixel location. For example, Fig. 1A is an image of a pollen grain (*Clavainaperturites microclavatus*) that is converted and separated into three single-channel images (Figs. 1B–D) using the HSV color system. The hue image (Fig. 1B) and saturation image (Fig. 1C) are then merged with weights w_H and w_S to produce an intermediate image (Fig. 1E). We tested a sizable sample of images using various weight combinations and observed an influence of weight choices on segmentation performance (Fig. 2). The bigger w_H, the more the hue value was emphasized; therefore, image pixels were separated based heavily on hue, leading to the inclusion of pixels of debris and artifacts. As w_S increased, the more detailed apertures on the grain surface were lost because they were lighter in saturation. Weight values were heuristically chosen as 0.4 for w_H and 0.6 for w_S to produce the most consistent segmentation. To automate the selection of channel-merging weights, a training data set of images with user-defined segmentation would be needed to tune these two parameters. A simulated annealing (SA) algorithm (Kirkpatrick et al., 1983) can then be implemented for automatic parameter selection (Han and Shyu, 2010). This was not done in this study due to limited sample size, but could be implemented with a larger image training data set.

Next, the intermediate image was binarized using Otsu thresholding, which automatically selected a threshold value for binarization (Otsu, 1975). Morphological operations (nonlinear operations related to the shape or morphology characteristics in an image) such as erosion, dilation, opening, and closing (Gonzalez and Woods, 2002) were performed to separate the main body of the grains from any debris or trivial particles that were not of interest in the analysis. Connected components (Gonzalez and Woods, 2002) were identified to represent object candidates, and only the largest one (presumably the grain) was kept. Finally, a watershed algorithm (Vincent and Soille, 1991) was used to separate any remaining particles that were still connected to the main body of the grain. When there is a distinct boundary between grain and overlapped debris based on differences in pixel value of saturation, hue, and intensity or in surface texture, it is also possible to separate the grain from debris using the combination of weights described in the previous paragraph. However, it should be noted that segmentation is still a largely unresolved problem in image analysis research. It is widely recognized in image segmentation that when target objects overlap with debris, their boundaries are blurred and undistinguishable, and segmentation performance has less consistency and accuracy. Human delineation may ultimately be needed to construct a reliable training set for our computer vision program to learn to separate objects from background. However, most of the grain samples in this study minimally overlapped with debris, and efforts were made to confirm accurate segmentation.

***Visual feature extraction*—**Once the pollen or spore grain is segmented, 69 visual features (Table 2) related to global visual characteristics (such as color, pixel value histograms, and textural patterns) and object morphology (such as convexity of convex hull, curvature of contour, and aspect ratio of

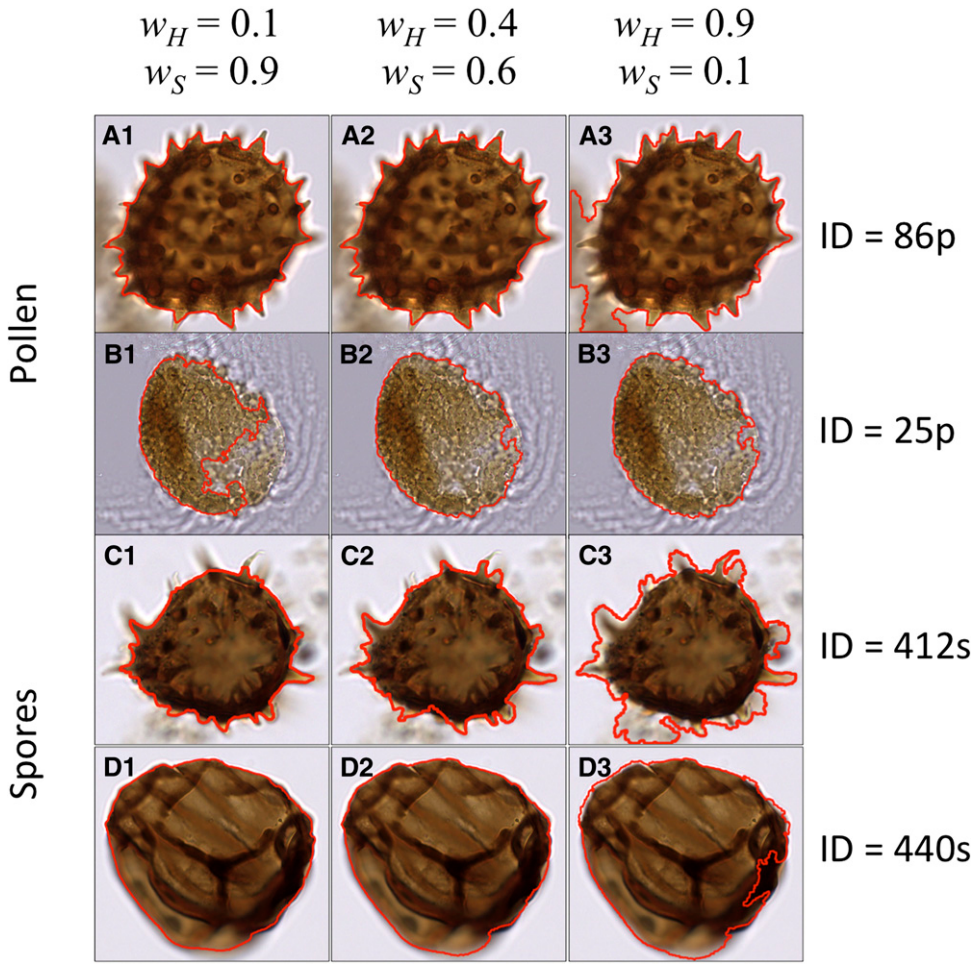


Fig. 2. Weight configuration examples using two pollen grain images (row 1, ID = 86p; row 2, ID = 25p) and two spore grain images (row 3, ID = 412s; row 4, ID = 440s). Three segmentation results (highlighted red contours superposed on original grain images) are shown per each image example using different weight configurations for hue (w_H) and saturation (w_S) channels.

Table 2. Visual features extracted from four single-channel images.

Name	Description	No. of features[a]	Index
Threshold	Otsu[b] threshold	1	color
Mean	Mean pixel value	1	color
SD	Standard deviation of pixel value	1	color
Histogram	1-dimensional histogram with 16 bins	16	color
Size	Grain object size	1	shape
HU	Hu[c] shape descriptors	7	shape
Aspect ratio	Ratio of long edge to short edge of bounding box (Fig. 3C)	1	shape
Compactness	*see Appendix 1*	1	shape
Convexity	*see Appendix 1*	1	shape
Form factor	*see Appendix 1*	1	shape
Roundness	*see Appendix 1*	1	shape
Solidity	*see Appendix 1*	1	shape
Perimeter	*see Appendix 1*	1	shape
Texture	Seven Haralick[d] texture with five step sizes	35	texture
Total		69	

[a] Numbers indicate the value per single-channel image. All features are calculated within segmented grain objects only. Refer to Appendix 1 for detailed calculations.

[b] Otsu, 1975.

[c] Hu, 1962.

[d] Haralick et al., 1973.

bounding box, illustrated in Fig. 3) were extracted from each of the four channels representing the original image: hue, saturation, value, and grayscale. This produced a 276-dimension feature space in which individual images were placed.

***Morphology semantic modeling*—**Palynologists use a common qualitative terminology to describe and compare the morphology of pollen and spores (e.g., Punt et al., 2007). However, complex morphological features that are relatively easy for human experts to detect and describe linguistically are much more challenging for the computer to recognize numerically. To mimic the complex human process of identifying visual patterns, low-level visual features were extracted to represent the visual content in images. Examples of low-level features used in this study include: single-channel histograms (Fig. 4), Hu shape momentum descriptors (Hu, 1962), and texture (Haralick et al., 1973). In some image analysis research domains, such visual patterns are interpreted using high-level abstractions, called *semantics*. The extracted features can describe, to a limited extent, the visual content of grains, but are still not easily interpreted by the human analyst. This is known as the *semantic gap* (Lew et al., 2006). To minimize the semantic gap, mathematical models are constructed using low-level features to map images to high-level trait semantics based on degrees of relevance. Using the mathematical formulas detailed in Barb and Shyu (2010) and Barb and Kilicay-Ergin (2013), each semantic representation is constructed as an association model using the concept of possibilistic *c*-means algorithm (Krishnapuram and Keller, 1993) based on low-level visual features. This process is called semantic modeling.

In this study, there are three morphology semantic categories for pollen images and three for spores, each of which consists of several exclusive semantic labels (Table 3). Using semantic modeling, each semantic label was represented as a semantic model of low-level visual features. Semantic model M_ξ is based on a training data set of images all labeled with semantic ξ. A trained semantic model M_ξ returns a relevance score for each database image for this specific semantic label. To reduce over-fitting issues during semantic modeling and to estimate how well these trained models handle images that lack certain semantic labels, 10-fold cross-validation (Kohavi, 1995) was conducted in this study. In our study, an image was first represented by a multidimensional feature vector, which was then fed into each semantic model to calculate its relevance scores. These relevance scores were then used for automatic semantic annotation and semantic-based image retrieval.

***Image annotation using semantic models*—**Within each category, the higher the relevance score, the larger the possibility that an image has this particular morphology semantic. The model that produces the highest score in each semantic category determines the assignment of semantic labels to an image. In this study, a grain image can be annotated with three semantic labels, each from a different category. For example, the relevance scores for a spore image are {(*pyramidal* = 0.661, *plane-convex* = 0.506, *reniform* = 0.333) $^{lateral\ view}$, (*elliptic* = 0.333, *circular* = 0.921) $^{polar\ view}$, (*radial* = 0.921, *bilateral* = 0.333) ^{symmetry}}. In Category "*lateral view*," *pyramidal* has the highest score compared to *plane-convex* and *reniform*. Therefore, this image can be annotated as having a *pyramidal* shape in lateral view. With the same strategy, this image can also be labeled as *circular* in Category "*polar view*" and *radial* in Category "*symmetry*." In standard practice, newly acquired images are not often labeled. Relevance scores provided by semantic models will be useful for automatic annotation of images with undetermined semantics. Once the models are trained, no human intervention is needed for model selection and image annotation. Confusion matrices were used to visualize annotation performance for individual semantic labels. Average accuracies were calculated for pollen and spore samples in this study.

***Image search using semantic models*—**The relevance scores provided by trained semantic models can be used to search images based on their semantic assignment. Consider ranking images based on their relevance scores of semantic label ξ to be a single semantic image retrieval, its performance can be evaluated using precision, *P*, and recall, *R* (Baeza-Yates and Ribeiro-Neto, 1999).

$$P = \frac{n}{N} \tag{2}$$

$$R = \frac{n}{|t_\xi|} \tag{3}$$

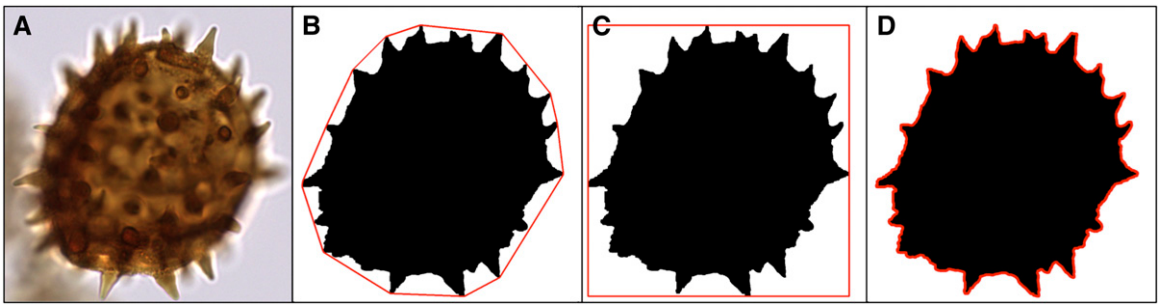
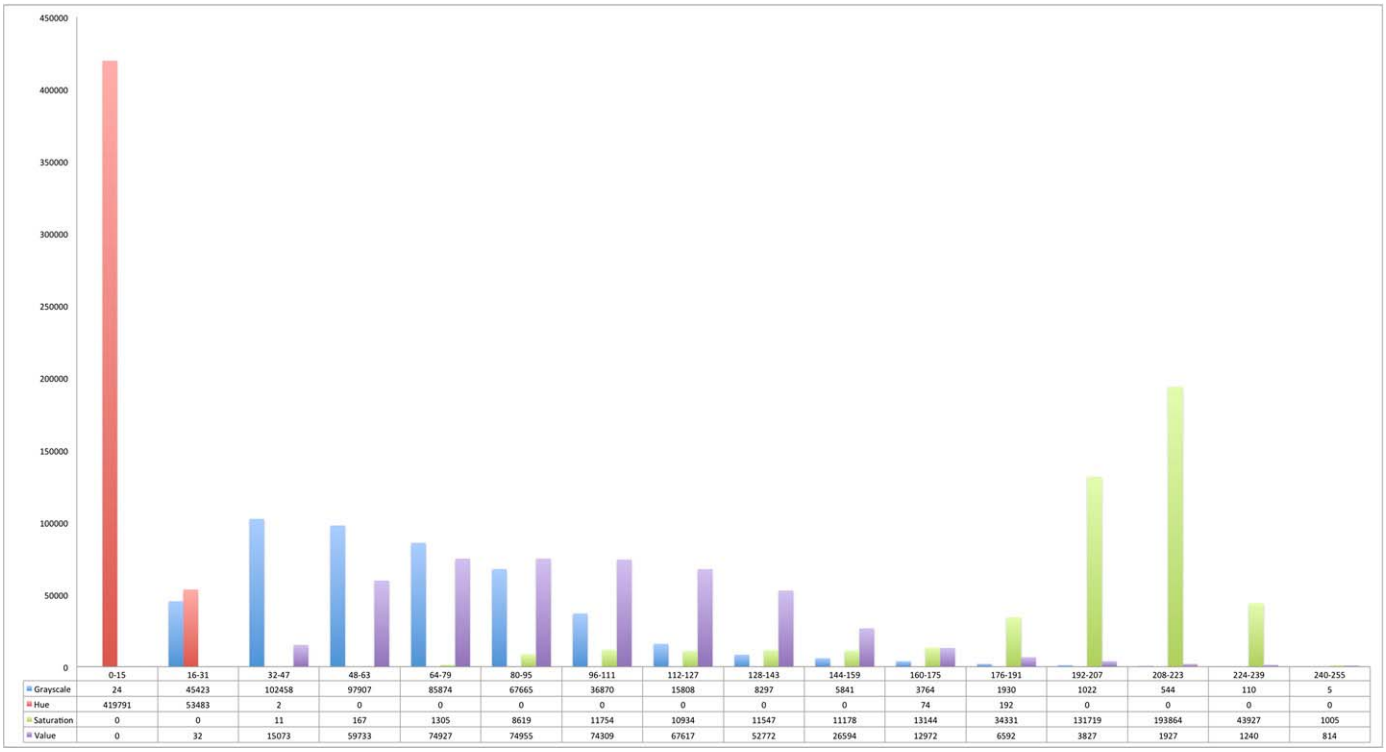


Fig. 3. Image examples of selected features listed in Table 2. (A) Original image; (B) convex hull that encloses binarized pollen grain; (C) bounding box that encloses binarized pollen grain; and (D) contour that traces along the boundary of binarized pollen grain.



	0-15	16-31	32-47	48-63	64-79	80-95	96-111	112-127	128-143	144-159	160-175	176-191	192-207	208-223	224-239	240-255
Grayscale	24	45423	102458	97907	85874	67665	36870	15808	8297	5841	3764	1930	1022	544	110	5
Hue	419791	53483	2	0	0	0	0	0	0	0	74	192	0	0	0	0
Saturation	0	0	11	167	1305	8619	11754	10934	11547	11178	13144	34331	131719	193864	43927	1005
Value	0	32	15073	59733	74927	74955	74309	67617	52772	26594	12972	6592	3827	1927	1240	814

Fig. 4. Example histogram of an example image in four individual channels: grayscale, hue, saturation, and value.

where n is the number of images labeled with semantic ζ in a list T_ζ of top N ranked images and $|t_\zeta| \leq |I|$ is the total number of images labeled with semantic ζ in database I. An image is considered relevant if it is labeled with query semantic ζ. Precision is the fraction of relevant images in result list T_ζ. Recall is the ratio of retrieved relevant images to the total number of relevant images in the database.

When an image database contains hundreds of thousands of images, one wants to see a list of the most relevant images instead of viewing the entire collection. The more relevant the images that appear at the top positions in the list, the better the retrieval. Precision-recall curve, which represents precision as a function of recall rate, can demonstrate how relevant images are distributed in a ranked list. Another evaluation measurement is mean average precision (MAP) score over 10 folds of experiment (detailed in Appendix 1). The higher the MAP score, the more relevant the images that are placed at the top positions in the list.

This semantic-based image search was then extended to include multiple semantic labels. The relevance scores for each semantic were used to calculate an overall relevance score in regard to a set Q of semantics selected by a user (see explanation in Appendix 1).

TABLE 3. Semantic labels used to describe morphology of pollen and spore grains.[a]

Semantic category	Semantic label	No. of images
Pollen		
Equatorial view	Prolate	50
	Spherical	138
	Oblate	25
	Unlabeled	184
Polar view	Elliptic	137
	Circular	140
	Unlabeled	120
Symmetry	Radial	172
	Bilateral	136
	Unlabeled	89
Spore		
Lateral view	Pyramidal	52
	Plane-convex	28
	Reniform	23
	Unlabeled	25
Polar view	Elliptic	23
	Circular	80
	Unlabeled	25
Symmetry	Radial	80
	Bilateral	23
	Unlabeled	25

[a] Morphology semantic terms are adopted from Punt et al. (2007).

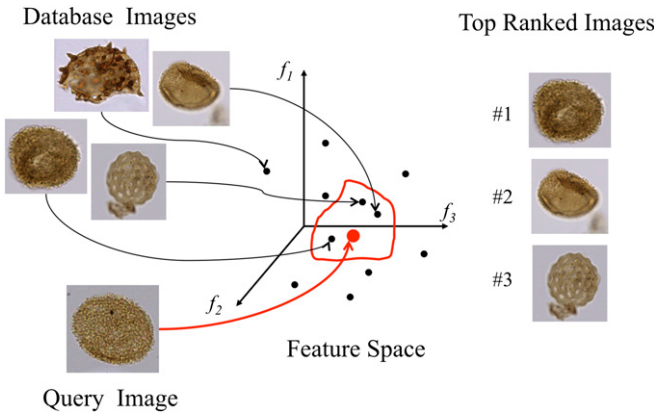


Fig. 5. Hypothetical query by image example in a multiple-dimensional feature space (illustrated here in three dimensions). Each black dot represents a multidimensional feature vector that represents a database grain image. A query image is mapped into the same feature space (red dot). The nearest neighbors are selected and ranked with their corresponding images displayed to the user.

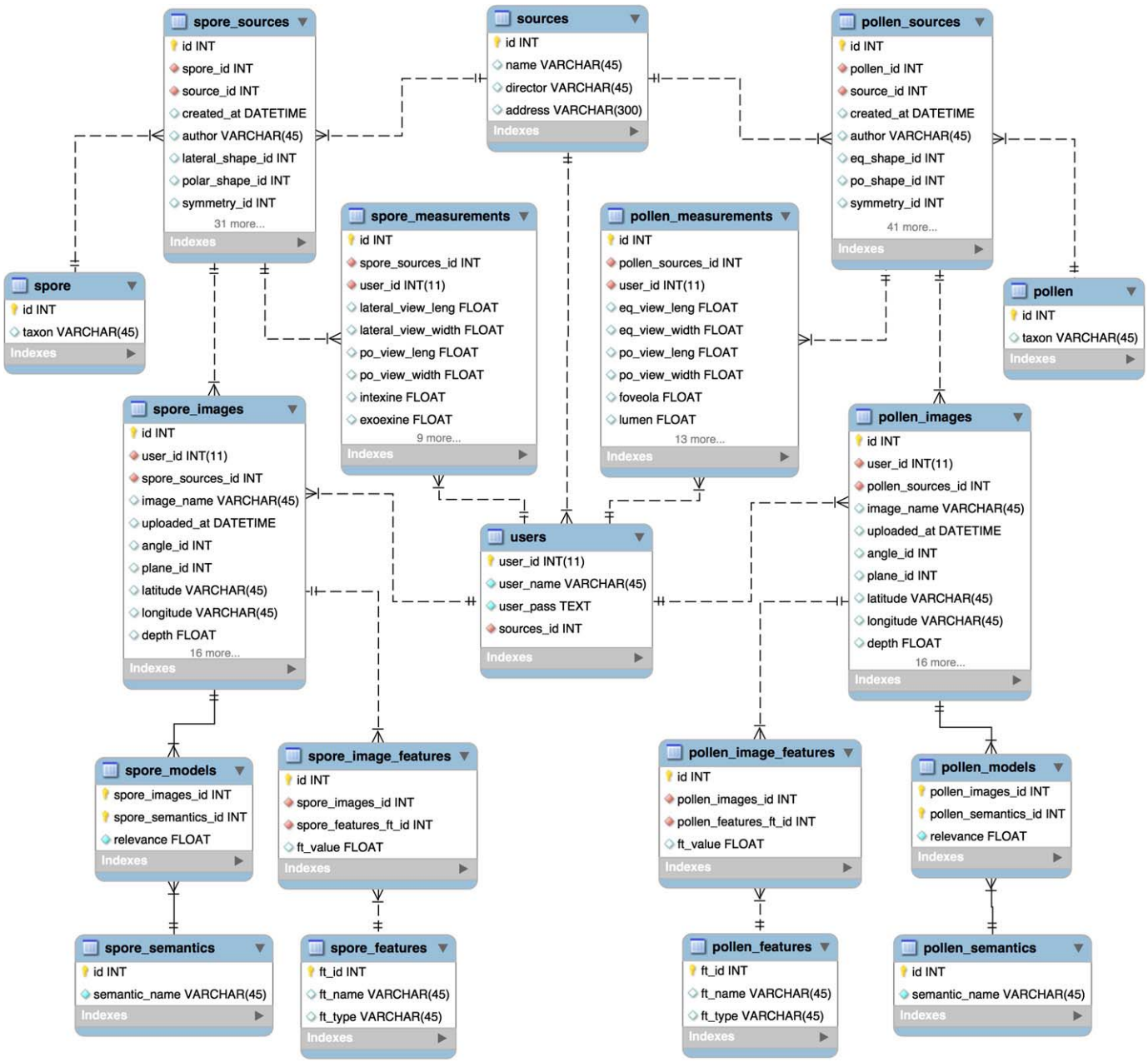


Fig. 6. Entity relation diagram (ERD) of database design. Entities and their relationships are represented as tables with attributes and connected using crow's feet annotation. For example, the relationship between *pollen* and *pollen_images* is a one-to-many identifying relationship. Specifically, one pollen taxon can have multiple images, and each record in *pollen_images* must reference to only one record in *pollen*. There are eight pollen-related tables on the right-hand side and eight spore-related tables on the left-hand side. Tables from both sides share similar structure and reference to two common tables—*users* and *sources*. Note: There are 76 tables and more than 200,000 records in the database. There are 49 attributes in the table *pollen_sources* and 39 in the table *spore_sources*. For simplicity, some auxiliary tables and secondary fields are omitted in this figure. Only the most relevant tables and fields are shown.

***Image search using grain examples*—**The semantic modeling additionally provides a basis for the query of images within the database and the retrieval of the most visually similar pollen grain images. In this way, a newly acquired image can be uploaded into the search engine to find similar types from the database. This allows for the comparison of morphotypes across analysts, potentially improving classification consistency among multiple experts. Using the image itself or, in other words, using image content as a whole for image database search is called *content-based image retrieval* (Smeulders et al., 2000).

The 276-dimensional features used in the initial morphological feature extraction formed a visual content space. These features were used to index the image database for fast retrievals. For simplicity, only three dimensions are depicted in Fig. 5 to demonstrate the concept of content-based image retrieval from a multidimensional feature space. Each data point is a multidimensional vector representing an image in the database. The distance between a query point and a data point defines the similarity between the query image and a database image. In other words, the closer two data points are in the feature space, the more visually similar these two images

are. With this defined similarity measure, images can then be ranked based on their similarity (distance) scores. For a large-scale image database (millions of grains), instead of exhaustively computing distances between the query image and all database images, customized database-indexing structures, such as M-tree (Ciaccia et al., 1997) or Entropy Balanced Statistical (EBS) k-d tree (Scott and Shyu, 2007), can drastically improve the efficiency of retrieval by strategically organizing the indexes of data in a high-dimensional space. In our implementation, we created three M-tree indexes by grouping the 276 visual features into colors, shapes, and textures (Table 2).

Our system provides users with customized weighing options to search grain images. For example, one user may want to see images that are most similar in regard to shape features with less emphasis on color and texture variances. Color, in particular, is a highly variable characteristic as it is mainly controlled by the thermal maturation of the organic matter. Pollen grains are light yellow when thermal maturation is low but change to darker colors as rock maturation increases, reaching a full black when organic matter is over-matured. In this scenario, users can customize their queries with a minimal weight on color index while placing more emphasis on the other two indexes.

In this study, content-based image retrieval was evaluated using precision in the top-ranked images (Baeza-Yates and Ribeiro-Neto, 1999). Precision, as described above, is defined as the ratio of number of relevant images in top k ranked images ($k = 10$ in this study). A result image is relevant when it belongs to the same species as the query image example. Because the contribution of indexes to retrieval performance is not, and should not, be universally fixed across all species, we simulated possible combinations of weights (w) for color (w_c), shape (w_s), and texture (w_t) with an increment of 0.2 (see Appendix 1 for detailed process) using labeled images in our database as a training set. The most suitable weight combinations were identified based on our current database image collection. Their values determined the retrieval precision of each query. Once the most suitable weight combinations were identified for each species, they were presented to users as the initial weights upon which emphases can be adjusted based on users' search preferences. As our database collection grows to include more species and variety in morphology, new weight combinations can be learnt to produce better retrieval results based on the newly populated database.

***Database design for multimodal information integration*—**For an image database to be effectively used in taxonomic classification and customized image retrieval, accurate metadata are as important as novel search algorithms. Our database structure was designed to be flexible for database management across geographically remote sites and allows for sustainable growth over time with the incorporation of new palynological images and the participation of new analysts. Instead of storing data in single files such as spreadsheets or printed catalogs, images and their metadata are stored in a relational database where the shared data structure and data relationships are carefully designed and maintained to avoid duplication or accidental modification. The entity relationship diagram (ERD) illustrates the database structure and its tables with relationships that ensure data integrity and handle dynamic data changes such as insertion, deletion, and update (Fig. 6). In the ERD, tables on the left side are designed for spore taxa, and pollen-related tables are on the right side with the same table structures. In addition, to link multiple research groups for cross-site research, the *sources* table and the *users* table store information of the research teams and palynologists who collected the images. With two relationship tables, *pollen_sources* and *spore_sources*, two sides are linked together to enforce relationship dependencies. With this database as the back end, a Web-based system was built for palynologists to interact with stored data and search for grain images. The system provides not only text-based species search (Fig. 7), but also image searches based on trait semantics (Fig. 8) and visual content (Figs. 9 and 10) with personalized search criteria. The demo website can be accessed at http://www.bioshapes.org/mioceneDB.

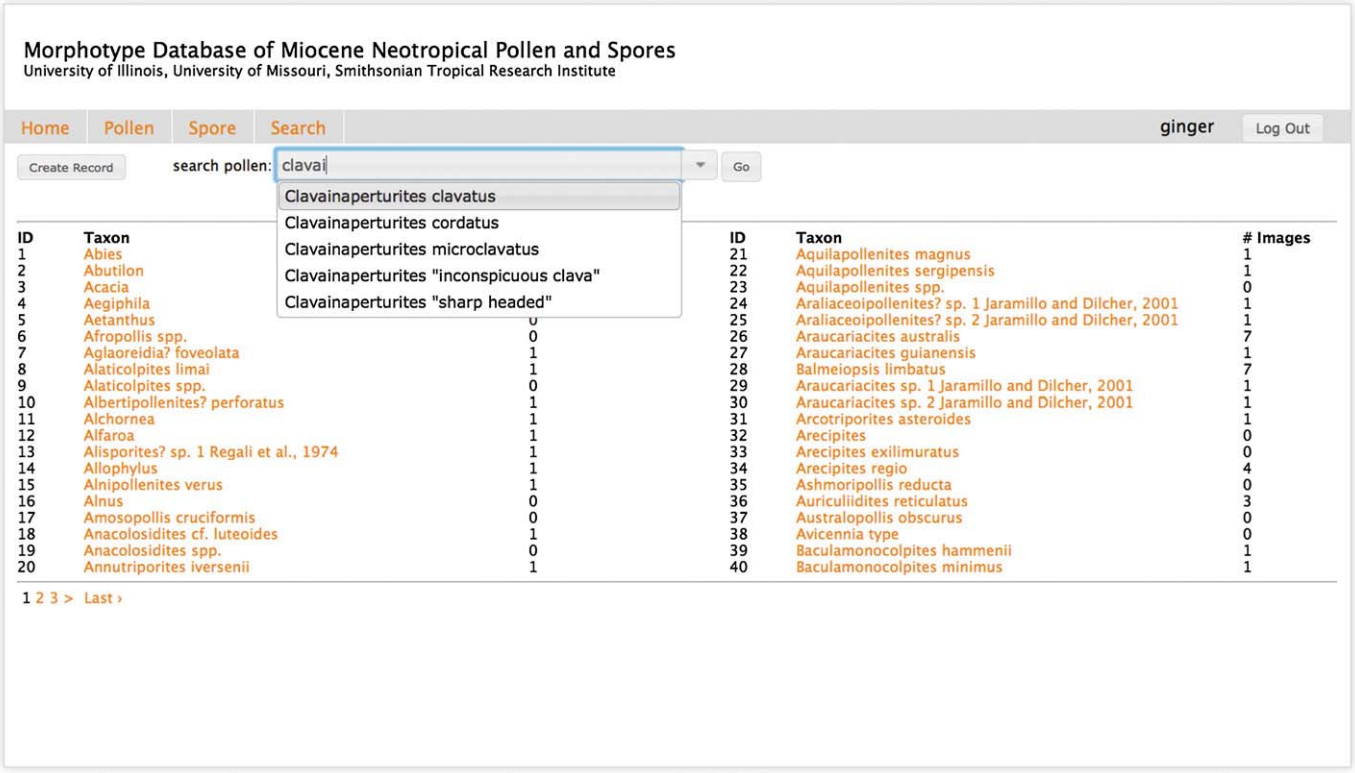


Fig. 7. Searching for pollen taxa by name. All existing taxa in the database are listed on the webpages ordered by taxon ID. Users can choose to search taxa by their scientific names by typing in the text field. Auto-complete hints help users to quickly narrow down the list.

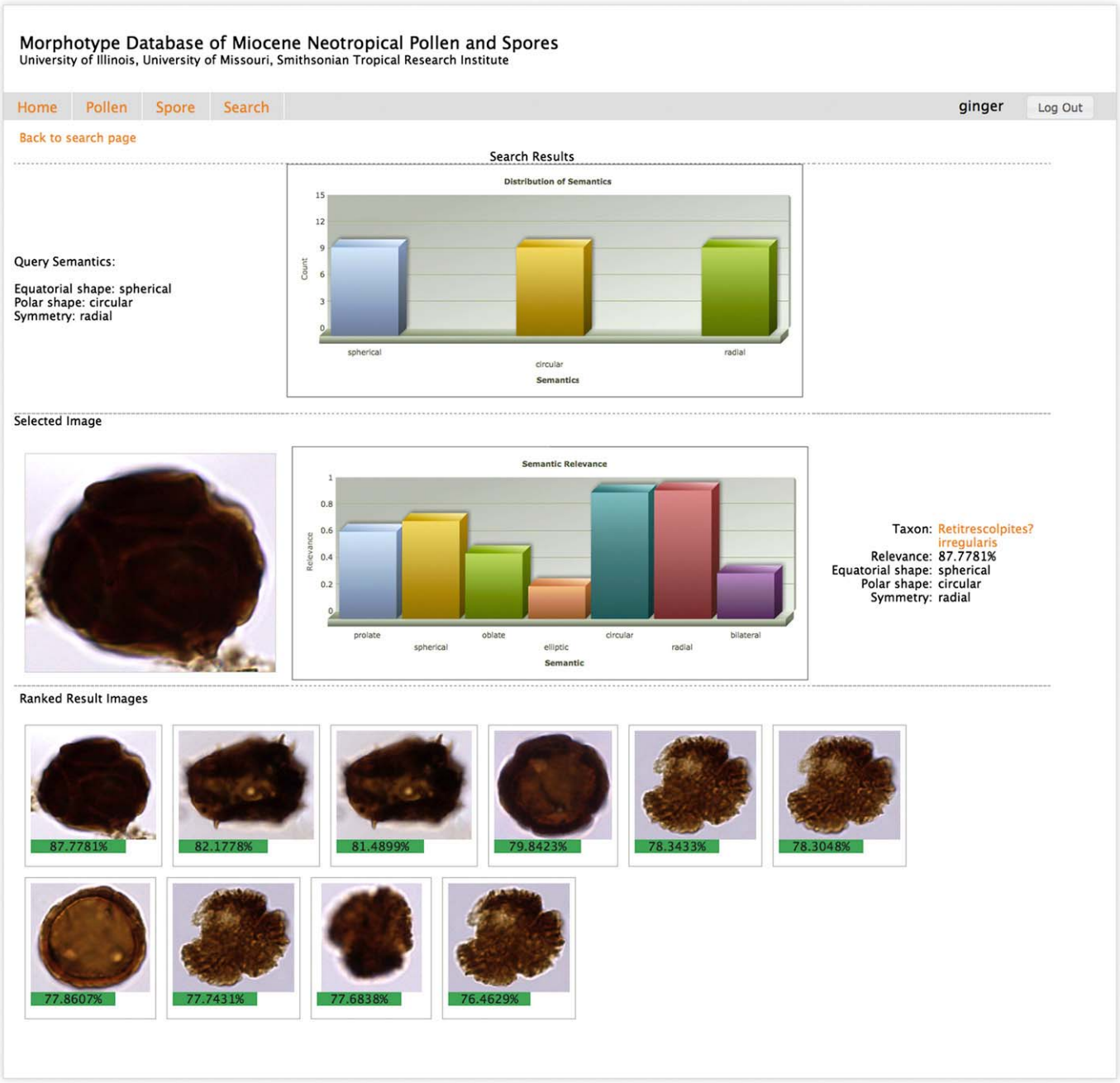


Fig. 8. Searching for pollen images by morphology semantics. Top row: (left) Morphology semantics selected by user and (right) distribution of semantics in result images. Center row: (left) first image in ranked list, (middle) relevance scores calculated by trained semantic models, and (right) additional information about this image, including taxon name, its overall relevance score as regard to user-selected semantics, and its actual semantics annotated and stored in the database. Comparing the actual semantics to the relevance score chart, we can see that spherical has higher relevance than prolate and oblate for equatorial shape semantic, circular is more relevant than elliptic for polar shape semantic, and radial is more relevant than bilateral for symmetry semantic. Bottom row: the ranked result image list with overall relevance score calculated using Eq. 10 in Appendix 1 as regard to user-selected morphology semantics.

RESULTS

Semantic modeling— After 10 folds of association rule mining, semantic models were trained for each semantic label. Each model was composed of several associative rules mapping subspaces of low-level visual features to high-level semantic labels with a certain confidence. For example, one of the rules for model M_{oblate} is

$$\left\{Grayscale\_histogram\_bin2 \in [0.1163, 0.1394] \rightarrow \text{"}oblate\text{"}\right\} = 0.877$$

indicating that it was determined with 87.7% confidence that images with measurements in the feature subspace were labeled with *oblate*. Another rule

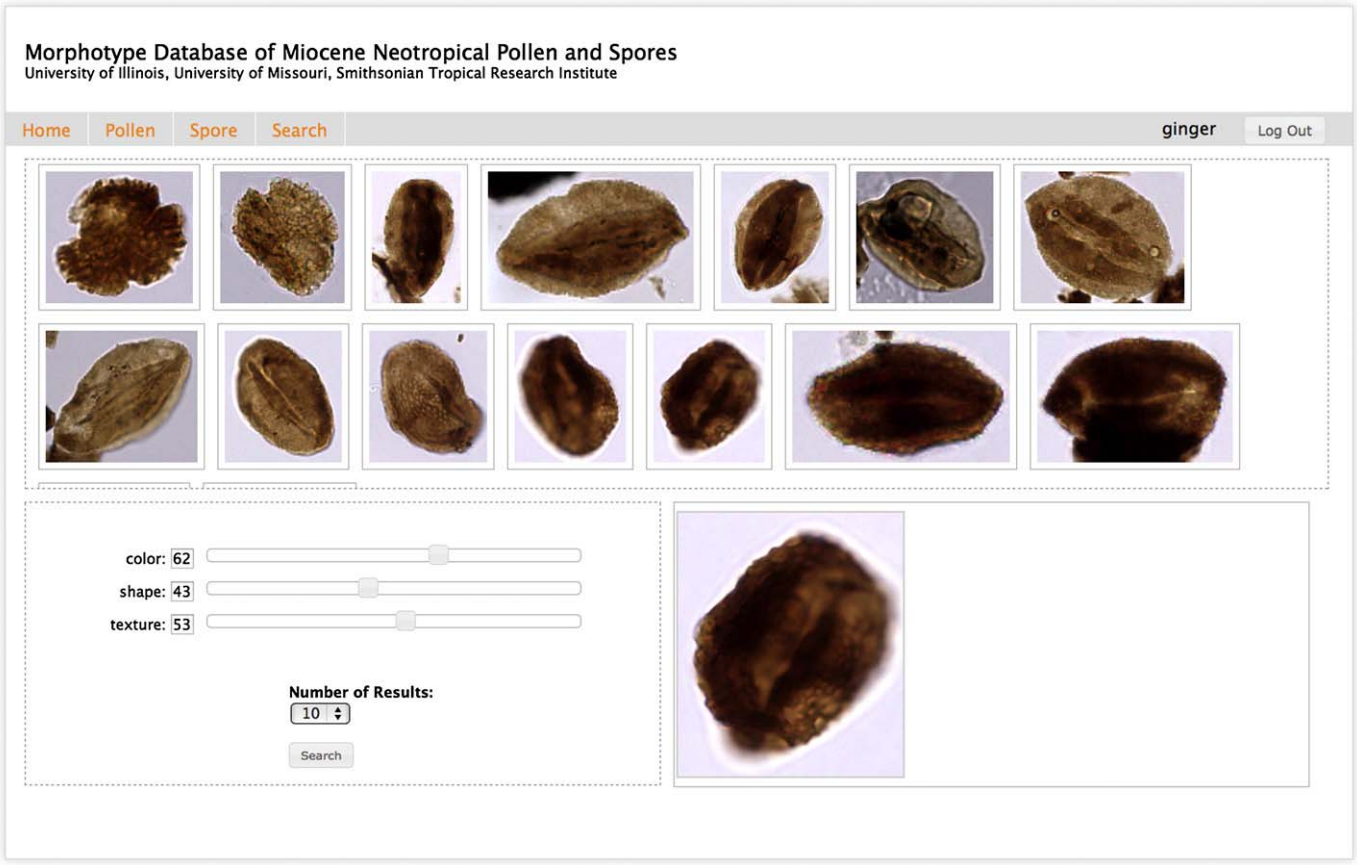


Fig. 9. Searching for pollen images by query image example. Query image example is selected from the example list, and query weights on three indexes (color, shape, and texture) can be adjusted to the user's preference. The weight values range from 0 (left end of the bar, representing no weight) to 100 (right end of the bar, representing the highest amount of emphasis).

$$\{Saturation\_histogram\_bin16 \in [0.0654, 0.1168]$$

$$\wedge Saturation\_(histogram\_bin14\ ) \in [0.0162, 0.1402]$$

$$\wedge Hue\_perimeter \in [0.0677, 0.1151]$$

$$\wedge Saturation\_texture\_step3\_direction5 \in [0.0712, 0.1329]$$

$$\wedge Saturation\_texture\_step5\_direction4 \in [0.1474, 0.1846]$$

$$\rightarrow \text{"}oblate\text{"}\} = 0.766$$

maps a more complex subspace of multiple features to the same semantic label *oblate* with a different confidence of 76.6%.

Image annotation—When trained semantic models are used for automatic semantic annotation, they are evaluated by annotation accuracy. After 10-fold cross-validation, the annotation accuracy is shown in the form of a confusion matrix for pollen and spore morphology semantics (Tables 4 and 5). In a confusion matrix, the value x in a cell (ζ, τ) means that x images with human-annotated semantic label ζ are annotated by computer with semantic label τ. Cell values are meaningful only when ζ and τ are from the same category. In an ideal scenario, we expect all images be annotated with correct semantic labels in each category. Therefore, the confusion matrix should only have nonzero values in cells on the diagonal where row label (human-annotated semantic) and column label (computer-annotated semantic) are the same. In reality, errors cannot be completely eliminated in automatic annotation. For example, in Table 4, among the 25 images that were labeled as *oblate* in Category *equatorial view shape*, 18 images were annotated by computer correctly while the other seven images were annotated as *prolate*. In this case, the accuracy of annotation for the semantic label *oblate* is 18/25 = 72.0%. The average accuracy was 83.9% for pollen semantic annotation and 98.6% for spores.

Semantic-based image retrieval—The performance of the semantic-based image retrieval was evaluated using precision-recall curves and MAP scores (see details in Appendix 1). Figures 11 and 12 are precision-recall curves for morphology semantics of pollen and spores, respectively. In Fig. 12, average precisions calculated from 10 folds of experiments were plotted as functions of recall for all seven semantic models from three categories for spores. Precisions of all semantics were maintained above 80% at 60% recall rate. For pollen images (Fig. 11), despite the steeper

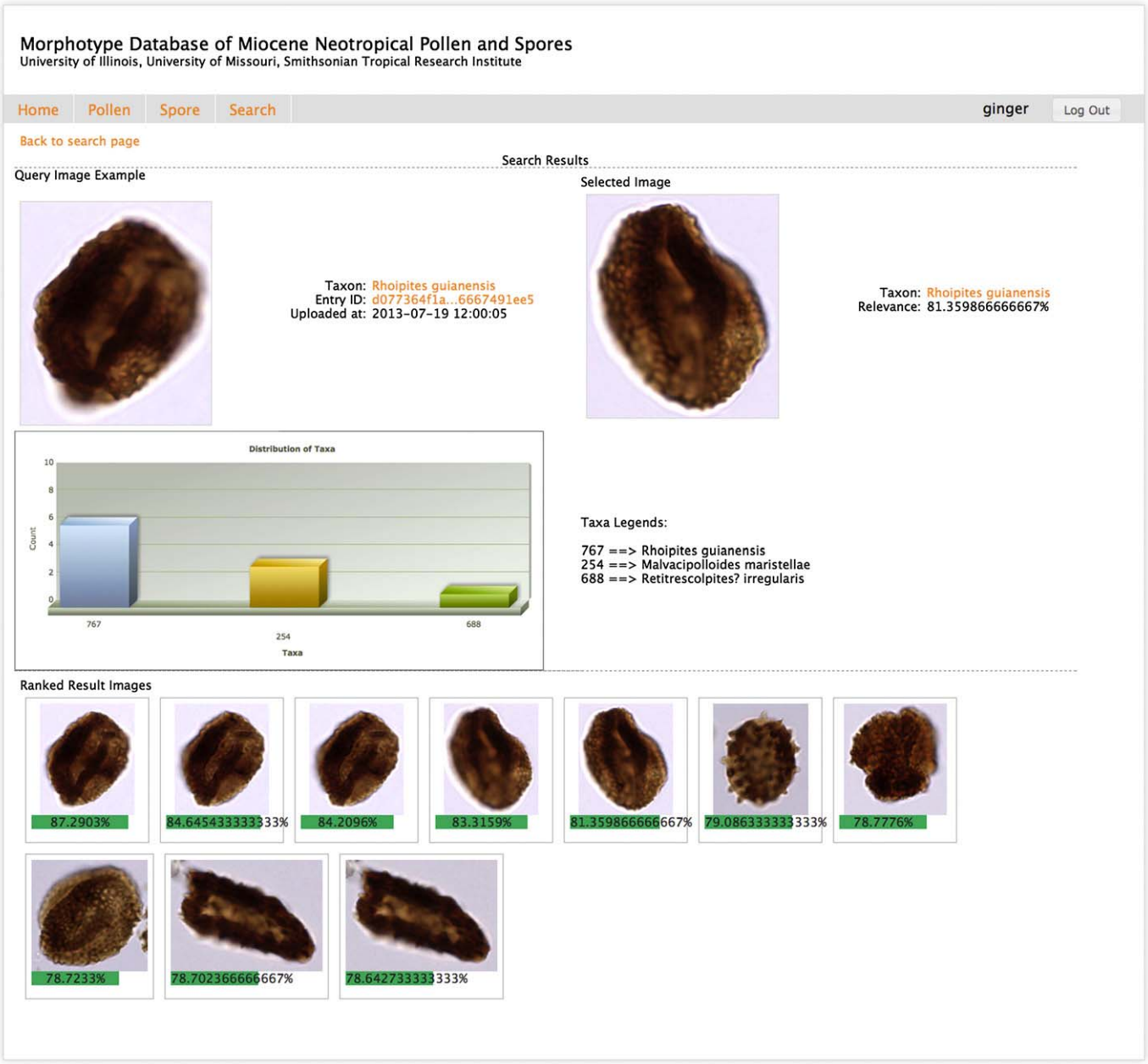


Fig. 10. Search by pollen image example result page. Top row: (left) search example and (right) the fifth result in the ranked list. Center row: distribution of taxa count from results. Bottom row: ranked result image list with their similarity measures against the query image example (top left). The bar chart in the center row indicates that there is a mixture of taxa in the result images.

precision-recall curves, all precisions were still maintained above 60% until the recall rate of 60%. The differences in these results are understandable, because the pollen image samples in this study were distributed in 15 distinct taxa, and it was therefore more challenging for semantic models to find association rules of feature subspaces that were generalized for all images. Queries using pollen images had an average MAP score of 0.81/1.00, and queries using spore images had an average MAP score of 0.93/1.00 (Table 6). The average search time for pollen and spore images was less than 0.2 s and as short as 65 ms.

Figure 8 shows the result page of a pollen image search using multiple morphology semantics. The semantic-based image search engine calculates overall relevance scores using Eq. 10 in Appendix 1, which uses each image's three relevance scores provided by semantic models for *spherical* equatorial view shape, *circular* polar view shape, and *radial* symmetry. The database images are ranked based on their calculated overall relevance scores, and the top 10 most similar images are displayed. It is not required that retrieved images must have all three morphologies labeled. As long as their relevance scores are significant, the overall relevance still satisfies the search criteria.

Table 4. Confusion matrix of pollen image trait semantic assignment.

	p	*s*	*o*	*e*	*c*	*r*	*b*	Accuracy (%)
p	50	0	0					100
s	45	90	3					65.2
o	7	0	18					72.0
e				133	4			97.1
c				23	117			83.6
r						129	43	75.0
b						8	128	94.1

Note: *p* = *prolate*; *s* = *spherical*; *o* = *oblate*; *e* = *elliptic*; *c* = *circular*; *r* = *radial*; *b* = *bilateral*.

Table 5. Confusion matrix of spore image trait semantic assignment.

	p	*v*	*r*	*e*	*c*	*r*	*b*	Accuracy (%)
p	47	5	0					90.4
v	0	28	0					100
r	0	0	23					100
e				23	0			100
c				0	80			100
r						80	0	100
b						0	23	100

Note: *p* = *pyramidal*; *v* = *plane-convex*; *r* = *reniform*; *e* = *elliptic*; *c* = *circular*; *r* = *radial*; *b* = *bilateral*.

Content-based image retrieval—On the content-based image retrieval interface (Fig. 9), a user first picks an image as an example and then adjusts emphasis on three trait semantic categories using sliding bars. On the search result page (Fig. 10), a list of top-ranked database images that are most similar to the image example is displayed.

Content-based image retrieval is a much more complex image retrieval method than those by key words and semantic labels. It is worth mentioning that species-level classification is the most challenging classification task in palynology (Punyasena et al., 2012; Mander and Punyasena, 2014; Holt and Bebbington, 2014), and images that are visually similar based on their content do not necessarily belong to the same taxon. As a result, even though a list of the visually most similar images is ranked and returned, the precision value calculated by judging the variation of species can be much lower if multiple species are presented in the resulting list.

Table 7 lists the top 10 best-performing universal weight combinations that yielded average retrieval precisions of 57.8% and 72.3% for pollen and spores, respectively. This means that if all species were treated the same, using a one-fit-for-all weight combination, the search results retrieve a list of database images with 57.8% belonging to the same species as the query image. The retrieval precision using such universal weight combinations for spore images in the database is 72.3% on average.

In contrast, if instead of choosing universal weight combinations for all images, different weight combinations were made for individual species, accuracy improves. Table 8 shows the top choices of weight combinations for individual species and their average retrieval precisions. For images of some taxa, selected weights on trait semantics could be drastically different from others. For example, to get an average precision of 77.1% for all *Clavainaperturites microclavatus* (ID = 1014) images, it is best to emphasize shape heavily and reduce weights on color and texture. These customized weight selections become handy if the user has a small number of targeted species in mind during the search.

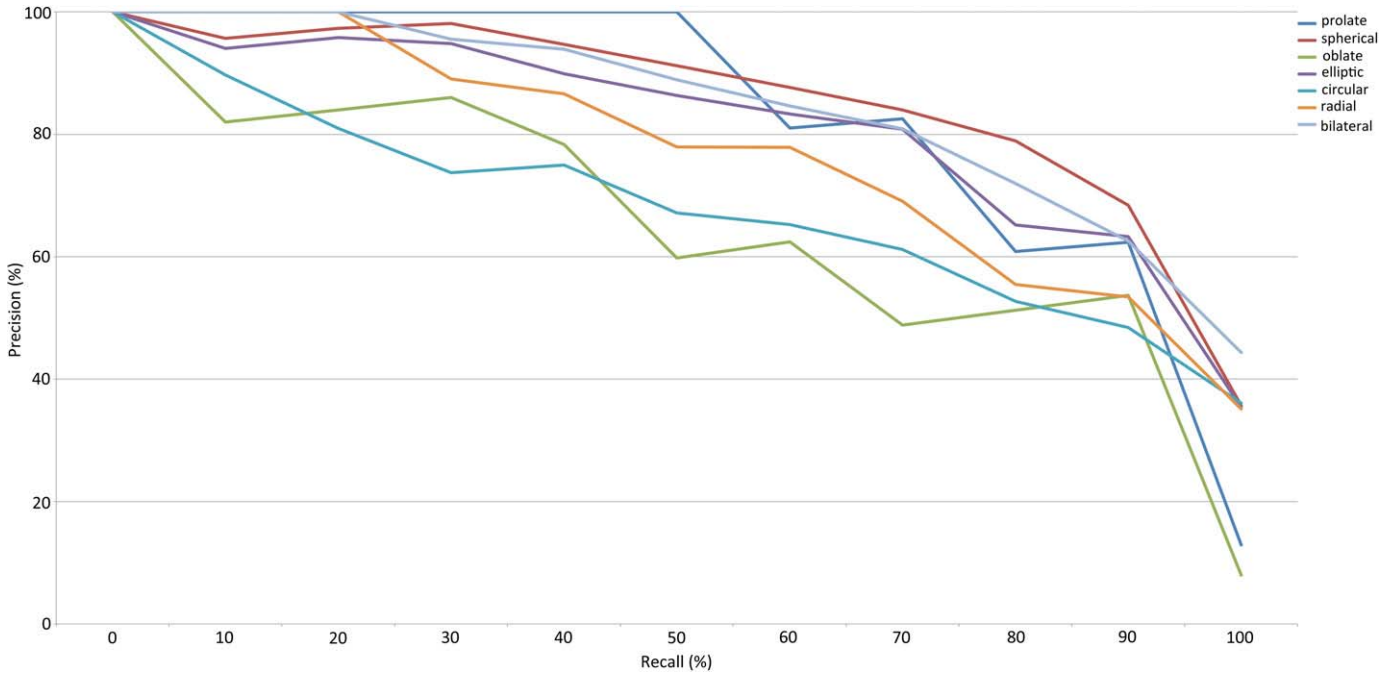


Fig. 11. Precision-recall curves of morphology semantics for pollen images. As the number of images retrieved increases, recall values gradually approach 100% while precision values gradually decrease due to retrieval of nonrelevant images.

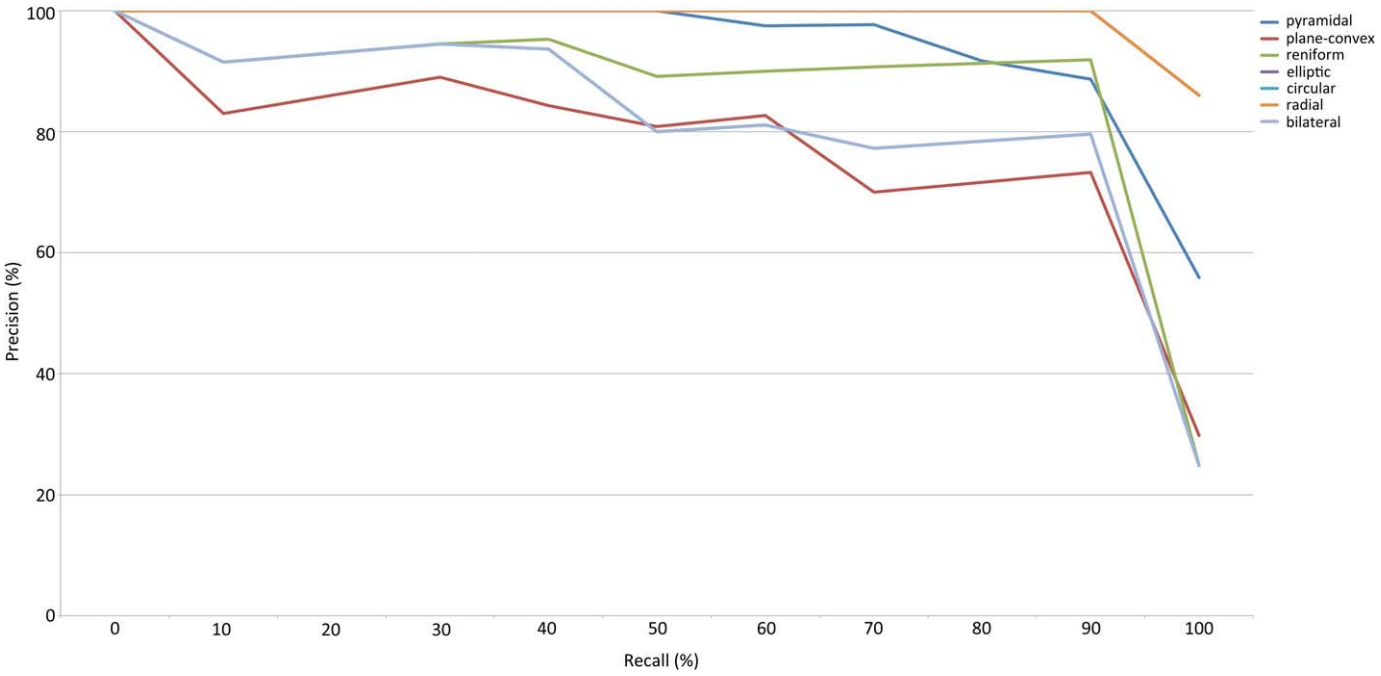


Fig. 12. Precision-recall curves of morphology semantics for spore images.

DISCUSSION

The ultimate goal of this research is to use informatics tools to assist palynological study by increasing speed and efficiency, reducing inter- and intra-observer inconsistency and labor intensity, and eventually assisting in the determination of new species. The database-driven application that we have presented integrates the analysis of image content, grain object morphology, morphology semantic modeling and annotation, and Web-based user-computer interaction for multimodal information integration.

Morphological semantic modeling uses mathematical equations that are optimized to discover association models in visual feature space and to find the best-fit semantic label to describe an image. In this study, morphology-related semantics were used to describe the grains with differences in shape and symmetry (as viewed in both polar and equatorial positions). With an average annotation accuracy of 83.9% for pollen semantics and 98.6% for spore semantics, the semantic models built for morphology annotation demonstrably produced good predictions based on grain morphology that can later be reviewed and validated by palynologists.

Table 6. Mean average precision (MAP) scores for semantic models trained over 10-fold cross-validation.[a]

Semantic category	Semantic label	MAP
Pollen		
Equatorial view	Prolate	0.86
	Spherical	0.88
	Oblate	0.73
Polar view	Elliptic	0.83
	Circular	0.70
Symmetry	Radial	0.86
	Bilateral	0.79
Spore		
Lateral view	Pyramidal	0.98
	Plane-convex	0.84
	Reniform	0.94
Polar view	Elliptic	0.88
	Circular	1.00
Symmetry	Radial	1.00
	Bilateral	0.88

[a] Morphology semantic terms are adopted from Punt et al. (2007).

The degree of morphological difference between morphospecies, sample size, and evenness of image representation all contributed to the robustness and reliability of automatic semantic annotation. For example, the high annotation accuracy scores for the spore images in this study (Table 5) are likely a result of the morphological diversity among the five spore taxa included in this trial analysis, which made classification relatively trivial. While the semantic labels for pollen *polar view shape* (Table 4) had the best annotation consistency, the group of equatorial semantics had less consistency for pollen images, especially for the semantic label *spherical*. This observation, as well as the high accuracy for all spore samples, indicates that a broader selection of species and image samples is needed for large-scale studies. However, we should emphasize that the semantic labels used in this study are a small subset of the features that can eventually be incorporated into an automated system for annotating morphology. Our semantic models can, and will, be expanded to include additional features known to be critical for accurate pollen and spore identifications, including pore and colpus number and arrangement, surface ornamentation, and grain size. A full list of possible semantic terms is illustrated in Punt et al. (2007).

In semantic image retrieval, the system's average MAP scores of 0.81 for pollen image samples and 0.93 for spore image samples demonstrate that trained semantic models can retrieve images based on their estimated relevance of trait semantics. Our system provides image retrieval by multiple

Table 7. Top 10 best-performing weight combinations sharing similar retrieval performance for all pollen and spore species in the database.

Top 10 performance with universal weight combinations	Pollen		Spore	
	Weights (color_shape_texture)	Average precision (%)	Weights (color_shape_texture)	Average precision (%)
1	0.4_0.6_0.6	57.9	0.2_1.0_0.0	72.4
2	0.8_1.0_0.8	57.9	0.0_0.2_0.0	72.3
3	0.6_1.0_0.8	57.9	0.0_0.4_0.0	72.3
4	0.6_0.6_0.8	57.8	0.0_0.6_0.0	72.3
5	0.2_0.2_0.2	57.8	0.0_0.8_0.0	72.3
6	0.4_0.4_0.4	57.8	0.0_1.0_0.0	72.3
7	0.6_0.6_0.6	57.8	0.2_1.0_0.4	72.2
8	0.8_0.8_0.8	57.8	0.4_1.0_0.4	72.2
9	1.0_1.0_1.0	57.8	0.2_0.8_0.0	72.0
10	0.2_0.4_0.4	57.8	0.2_0.8_0.4	72.0

trait semantics, which returns a list of images ranked based on their overall relevance to selected semantics. In the end, however, it is up to experts to judge the similarity and closeness of retrieved images and subsequently form a knowledge base for the palynology community. Our image database provides a potential platform for capturing those expert classification decisions.

Table 8. Best-performing weight combinations for each taxon and their average retrieval precisions. The taxon *Retitricolpites simplex* (ID = 722) has been used as an example. Each of its 24 images were used as query images, searched against the database, and the top 10 most similar images in feature space were retrieved. All 215 weight combinations were used for each image, yielding a total of $215 \times 24 = 5160$ queries. For each image as a query image, maximal precision was identified. There could be multiple weight combinations (n out of 215) that produced the same maximal precision for the same query image. All of these weight combinations were considered candidates. The candidates that occurred most frequently were the final candidates. In this example, there were four (#Candidates) candidate weight combinations that were identified to produce maximal precisions in nine (#Occurrence) individual query occasions. The average precisions using each of four candidates across all 24 images were calculated, and the candidate with the highest average precision was the top choice.

ID[a]	#Candidate	#Occurrence	Top choice[b]	Avg. precision (%)
Pollen				
722	4	9	0.8_0.2_0.0	50.8
767	2	17	0.4_0.6_0.2	73.1
688	7	6	0.0_0.4_0.2	63.4
570	2	24	0.0_0.6_0.2	59.1
571	21	10	1.0_0.6_0.2	74.7
552	1	16	0.2_0.8_1.0	73.1
451	8	14	0.2_0.0_0.0	61.2
511	1	9	0.2_0.0_1.0	68.9
450	1	20	1.0_0.4_0.8	67.2
254	5	16	0.2_0.0_0.0	49.6
1430	13	13	0.4_0.2_0.2	62.1
365	2	10	0.0_0.2_0.6	49.5
148	2	8	0.8_0.2_0.6	40.5
246	14	15	0.4_0.2_0.2	38.9
1014	3	11	0.2_1.0_0.2	77.1
Spore				
46	3	15	0.2_0.6_0.0	89.1
44	6	11	0.2_1.0_0.00	65.0
282	2	12	0.2_0.2_0.4	78.4
45	9	8	0.2_0.4_0.2	58.3
43	5	24	0.0_0.2_0.0	91.1

[a] Taxa IDs are provided in Table 1.

[b] Weight combinations are of format $w_c\_w_s\_w_t$ for color, shape, and texture.

Using content-based image retrieval, or image-based search, to search an image database provides an alternative nontextual way to describe morphology, compare visual content, and visually discover matches for unknown morphotypes. The performance of content-based image retrieval relies on the separation of samples and clusters in the feature space, training sample size, diversity, and on the balance in image representation among categories. Those taxa that have low precisions in Table 8 have higher variances in visual appearances due to factors such as variability in imaging acquisition settings and preservation quality of the grain. In this study, cautions were made to ensure consistency in image sample generation. Yet, ideal consistency is not always guaranteed when data samples are contributed from multiple sources with customized sample preparation protocols and experiment settings. As data collection proceeds with future contributions from multiple researchers, training data will be further enhanced by the expansion of the number and diversity of image representations, which will support a more robust and generalized basis for image retrieval. On the other hand, an increase in data sample variety also introduces uncertainty and inconsistency in semantic modeling and retrieval performance. The future expansion and implementation of this database will require that extra precautions be taken to overcome these potential shortcomings.

Additionally, we plan to incorporate the metadata generated from image acquisition devices, such as scanner configuration parameters, resolution, lighting, and other experimental variables. Using this additional information, images contributed by different researchers can be normalized, producing a standard for comparing images taken with different imaging conditions and more consistent performance in image analysis and, consequently, in retrieval and annotation. Human-in-the-loop annotation and validation is another strategy to help improve data sample quality and model robustness, as it is not always possible to replicate human expert knowledge with computerized programs.

CONCLUSIONS

Our data model provides a platform for the larger palynological community to collaboratively share and compare the morphological variations of formal and informal morphotypes. Content-based image retrieval allows palynologists to find similar images that may exist within the image database, assisting in the identification of novel types. The ability to

compare images with subtle visual differences potentially allows for finer delimitations of morphospecies and a more consistent taxonomy among analysts. By using the semantic modeling function in place of manual labeling, researchers can automatically annotate new grain images with best-fit morphology semantic labels for learning and discovery. This potentially speeds up the discovery and establishment of novel types. The value of such a database will only increase as the diversity of images and annotations increases, because the search results and semantic modeling potentially will become more robust and more generalized with training on larger and more heterogeneous image data.

LITERATURE CITED

Abramoff, M. D., P. J. Magalhães, and S. J. Ram. 2004. Image processing with ImageJ. *Biophotonics International* 11: 36–42.

Agrawal, R., T. Imielinski, and A. Swami. 1993. Mining association rules between sets of items in large databases. Proceedings of ACM SIGMOD International Conference on Management of Data, 207–216. Washington, D.C., USA.

Aguilera, O. A., and A. A. Carlini. 2010. Urumaco and Venezuelan paleontology: The fossil record of the Northern Neotropics. Indiana University Press, Bloomington, Indiana, USA.

Armato, S. G., and H. MacMahon. 2003. Automated lung segmentation and computer-aided diagnosis for thoracic CT scans. *International Congress Series* 1256: 977–982.

Baeza-Yates, R., and B. Ribeiro-Neto. 1999. Modern information retrieval. ACM Press, New York, New York, USA.

Barb, A., and C.-R. Shyu. 2010. Visual-semantic modeling in content-based geospatial information retrieval using associative mining techniques. *IEEE Geoscience and Remote Sensing Letters* 7: 38–42.

Barb, A., and N. Kilicay-Ergin. 2013. Genetic optimization for associative semantic ranking models of satellite images by land cover. *ISPRS International Journal of Geo-Information* 2: 531–552.

Belsky, C. Y., E. Boltenhagen, and R. Potonié. 1965. Sporae dispersae der Oberen Kreide von Gabun, Aquatoriales Afrika. *Paläontologische Zeitschrift* 39: 72–83.

Birks, H. J. B., and S. M. Peglar. 1980. Identification of *Picea* pollen of Late Quaternary age in eastern North America: A numerical approach. *Canadian Journal of Botany-Revue Canadienne de Botanique* 58: 2043–2058.

Bradski, G., and A. Kaehler. 2008. Learning OpenCV: Computer vision with the OpenCV library. O'Reilly Media, Inc., Sebastopol, California, USA.

Bush, M. B., and C. Weng. 2007. Introducing a new (freeware) tool for palynology. *Journal of Biogeography* 34: 377–380.

Ciaccia, P., M. Patella, and P. Zezula. 1997. M-tree: An efficient access method for similarity search in metric spaces. Proceedings of the 23rd VLDB Conference, 426–435. Athens, Greece.

Dueñas, H. 1980. Palynology of Oligocene-Miocene strata of borehole Q-E-22, Planeta Rica, Northern Colombia. *Review of Palaeobotany and Palynology* 30: 313–328.

Faegri, K., P. E. Kaland, and K. Krzywinski. 1989. Textbook of pollen analysis, 4th ed. John Wiley & Sons Ltd., Hoboken, New Jersey, USA.

Germeraad, J. H., C. A. Hopping, and J. Müller. 1968. Palynology of Tertiary sediments from tropical areas. *Review of Palaeobotany and Palynology* 6: 189–348.

Gonzalez, R. C., and R. E. Woods. 2002. Digital image processing, 2nd ed. Prentice Hall, Upper Saddle River, New Jersey, USA.

Gonzalez Guzman, A. E. 1967. A palynologic study on the upper Los Cuervos and Mirador formations (lower and middle Eocene, Tibú Area, Colombia). E. J. Brill, Leiden, The Netherlands.

Grimm, E. C., J. Keltner, R. Cheddadi, S. Hicks, A.-M. Lézine, J. C. Berrio, and J. W. Williams. 2013. Pollen databases and their application. *In* S. A. Elias and C. J. Mock [eds.], Encyclopedia of Quaternary science, 831–883. Elsevier, Amsterdam, The Netherlands.

Han, J. G., and C.-R. Shyu. 2010. Improving retrieval performance in medical image databases using simulated annealing. *AMIA Annual Symposium Proceedings Archive* 2010: 276–280.

Haralick, R. M., K. Shanmugam, and I. H. Dinstein. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (6): 610–621.

Holt, K., G. Allen, R. Hodgson, S. Marsland, and J. Flenley. 2011. Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. *Review of Palaeobotany and Palynology* 167: 175–183.

Holt, K. A., and M. S. Bebbington. 2014. Separating morphologically similar pollen types using basic shape features from digital images: A preliminary study. *Applications in Plant Sciences* 2(8): 1400032. doi:10.3732/apps.1400032

Holt, K. A., and K. D. Bennett. 2014. Principles and methods for automated palynology. *New Phytologist* 203: 735–742.

Hoorn, C. 1994. An environmental reconstruction of the palaeo-Amazon River system (Middle-Late Miocene, NW Amazonia). *Palaeogeography, Palaeoclimatology, Palaeoecology* 112: 187–238.

Hu, M. K. 1962. Visual pattern recognition by moment invariants. *I.R.E. Transactions on Information Theory* 8: 179–187.

Ibañez, L., W. Schroeder, L. Ng, and J. Cates. 2003. The ITK software guide. Kitware, Clifton Park, New York, USA.

Jaramillo, C. A., and D. L. Dilcher. 2001. Middle Paleogene palynology of central Colombia, South America: A study of pollen and spores from tropical latitudes. *Palaeontographica Abteilung B* 258: 87–213.

Jaramillo, C., and M. Rueda. 2013. A morphological electronic database of Cretaceous-Tertiary fossil pollen and spores from northern South America V. Colombian Petroleum Institute and Smithsonian Tropical Research Institute. Website http://biogeodb.stri.si.edu/jaramillo/palynomorph/ [accessed 15 July 2014].

Kedves, M., and N. Sole de Porta. 1963. Comparación de las esporas del género *Cicatricosisporites* R. Pot y Gell, 1933 de Hungria y Colombia. Algunos problemas referentes a su significado estratigráfico. *Boletín Geológico UIS* 12: 51–76.

Khan, A. M., and A. R. H. Martin. 1972. A note on genus *Polypodiisporites* R. Potonié. *Pollen et Spores* 13: 475–480.

Kirkpatrick, S., C. D. Gelatt Jr., and M. P. Vechi. 1983. Optimization by simulated annealing. *Science* 220: 671–680.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence* 14(2): 1137–1145.

Krishnapuram, R., and J. M. Keller. 1993. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems* 1: 98–110.

Lew, M. S., N. Sebe, C. Djeraba, and R. Jain. 2006. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications* 2: 1–19.

Mander, L., and S. W. Punyasena. 2014. On the taxonomic resolution of pollen and spore records of Earth's vegetation. *International Journal of Plant Sciences* 175: in press.

Mander, L., S. J. Baker, C. M. Belcher, D. S. Haselhorst, J. Rodriguez, J. L. Thorn, S. Tiwari, et al. 2014. Accuracy and consistency of grass pollen identification by human analysts using electron micrographs of surface ornamentation. *Applications in Plant Sciences* 2(8): 1400031. doi:10.3732/apps.1400031

Müller, J., E. Di Giacomo, and A. Van Erve. 1987. A palynologic zonation for the Cretaceous, Tertiary and Quaternary of northern South America. *AASP Contribution Series* 19: 7–76.

Ohta, Y.-I., T. Kanade, and T. Sakai. 1980. Color information for region segmentation. *Computer Graphics and Image Processing* 13: 222–241.

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

Pham, D. L., C. Xu, and J. L. Prince. 2000. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering* 2: 315–337.

Punt, W., P. P. Hoen, S. Blackmore, and A. Le Thomas. 2007. Glossary of pollen and spore terminology. *Review of Palaeobotany and Palynology* 143: 1–81.

Punyasena, S. W., D. K. Tcheng, C. Wesseln, and P. G. Mueller. 2012. Classifying black and white spruce pollen using layered machine learning. *New Phytologist* 196: 937–944.

Quiroz, L. I., and C. A. Jaramillo. 2010. Stratigraphy and sedimentary environments of Miocene shallow to marginal marine deposits in the Urumaco Trough, Falcon Basin, western Venezuela. *In* M. Sanchez-Villagra, O. Aguilera, and A. A. Carlini [eds.], Urumaco and Venezuelan palaeontology: The fossil record of the northern Neotropics, 153–172. Indiana University Press, Bloomington, Indiana, USA.

Regali, M., N. Uesugui, and A. Santos. 1974. Palinologia dos sedimentos Meso-Cenozoicos do Brasil. *Boletim Tecnico da Petrobras* 17: 177–191.

Russ, J. C. 2006. The image processing handbook. CRC Press, Boca Raton, Florida, USA.

Sarmiento, G. 1992. Palinología de la Formación Guaduas: Estratigrafía y sistemática. *Boletín Geológico Ingeominas* 32: 45–126.

Scott, G. J., and C. R. Shyu. 2007. Knowledge-driven multidimensional indexing structure for biomedical media database retrieval. *IEEE Transactions on Information Technology in Biomedicine* 11: 320–331.

Silva-Caminha, S. A. F., C. A. Jaramillo, and M. L. Absy. 2010. Neogene palynology of the Solimões basin, Brazilian Amazonia. *Palaeontographica Abteilung B* 284: 13–79.

Sklansky, J. 1982. Finding the convex hull of a simple polygon. *Pattern Recognition Letters* 1: 79–83.

Smeulders, A. W., M. Worring, S. Santini, A. Gupta, and R. Jain. 2000. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22: 1349–1380.

Traverse, A. 2007. Paleopalynology. Springer, Dordrecht, The Netherlands.

Van der Hammen, T. 1956. Description of some genera and species of fossil pollen and spores. *Boletín Geológico (Bogotá)* 4: 103–109.

Van der Hammen, T., and T. A. Wymstra. 1964. A palynological study on the Tertiary and Upper Cretaceous of British Guayana. *Leidse Geologische Mededelingen* 30: 183–241.

Van der Hammen, T., and C. Garcia. 1966. The Paleocene pollen flora of Colombia. *Leidse Geologische Mededelingen* 35: 105–114.

Van Hoeken-Klinkenberg, P. M. J. 1964. A palynological investigation of some Upper-Cretaceous sediments in Nigeria. *Pollen et Spores* 6: 209–231.

Vincent, L., and P. Soille. 1991. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13: 583–598.

Appendix 1. Supplementary definitions for visual feature calculation and algorithms for image retrieval.

Selected features listed in Table 2 are defined and demonstrated as follows:

Area: Total number of pixels in identified grain object or size of connected component.

Perimeter: Summation of distances between each pair of neighboring points on object contour (Fig. 3D).

Max. diameter: The maximal distance between two points on object contour.

Convex hull (Sklansky, 1982) length: Convex hull (Fig. 3B) perimeter.

Convex hull area: Total number of pixels contained inside a convex hull.

Compactness:

$$\frac{\sqrt{4 \times area / \pi}}{\max diameter} \tag{4}$$

Form factor:

$$\frac{4 \times \pi \times area}{perimeter} \tag{5}$$

Roundness:

$$\frac{4 \times area}{\pi \times (\max diameter)^2} \tag{6}$$

Convexity:

$$\frac{convex\ hull\ length}{perimeter} \tag{7}$$

Solidity:

$$\frac{area}{convex\ hull\ area} \tag{8}$$

Mean average precision (MAP) is the mean of the average precision (AP) scores over multiple queries.

$$AP_\zeta = \frac{1}{|t_\zeta|} \sum_{k=1}^{|I|} P\left(T_\zeta(I,k)\right) \tag{9}$$

Specifically, all database images are first ranked using relevance scores calculated by M_ζ. At each position k in the ranked list T_ζ, precision is calculated using Eq. 2 where $N = k$ and n is the number of relevant images counted until cutoff k. When the k-th image is not relevant, $P = 0$. The precisions at each position are then averaged to yield an AP score for semantic ζ. In this fold of modeling of M_ζ, an AP score is generated. Finally, the mean of AP scores for semantic ζ over 10 folds of modeling is calculated.

Image search using multiple semantic labels—Once the semantic models were trained, database images could be searched based on their relevance scores of multiple morphology semantics. For example, a set Q of query semantics is selected out of all N available semantics to query the database. Each image's overall relevance to this query is calculated using Eq. 10. All database images are then ranked based on such relevance score. The top k most relevant images are eventually returned to the user (Fig. 8).

$$s = \frac{s_{rel} * s_{irr}}{p * s_{rel} + (1-p) * s_{irr}} \tag{10}$$

$$s_{rel} = \frac{1}{|Q|} \sum_{\substack{i=1 \\ i \in Q}}^{N} r_i \tag{11}$$

$$s_{irr} = 1 - \max_{i \notin Q} r_i \tag{12}$$

Specifically, the overall relevance score (s) is a weighted combination of the average relevance score (s_{rel}) and the irrelevance score (s_{irr}). To calculate the average relevance score (s_{rel}) of an image, its relevance scores (r_i) for each selected semantic in Q are averaged. The irrelevance score (s_{irr}) is the opposite of the maximal relevance score calculated for those semantics that are not in Q. In Eq. 10, p is a system adjustment penalty to balance the scores of relevant and nonrelevant semantics. It is heuristically set to be 0.002 in this study.

Evaluation of content-based image retrieval—To find the most-suitable weight combinations, a series step was performed using permutated experiment results for each species in the current data set.

$$w \overset{\text{def}}{=} (w_c, w_s, w_t) \in \left[(0,0,0),(1,1,1)\right] \tag{13}$$

1. Each image, $\boldsymbol{t}^{\theta}$, from species θ was used as a query image for 215 (= 6^3-1) times to search against the entire database. Precisions, p_i^t ($1 \le i \le 215$), were calculated for each query.

2. The weight combinations (w_k^t) that produced the highest precision (p_{max}^t) were identified for each query image, composing a set of candidates, $W^t = \left\{ w_k^t \middle| p_k^t = p_{max}^t \right\}$.

3. For all images from the same species θ, their sets of weight combinations identified in step 2 were joined, and the most frequently occurring combinations were considered candidates for the most-suitable weight choices. If there were multiple candidates with the same number of occurrences, the one that yielded the highest average precision across all images in this species was considered the top choice.