Simple but long-lasting: A specimen imaging method applicable for small- and medium-sized herbaria

AtsukoTakano[1], Yasuhiko Horiuchi[2], Yu Fujimoto[3], Kouta Aoki[2], Hiromune Mitsuhashi[1,4], Akira Takahashi[1,4]

1 Museum of Nature and Human Activities, Hyogo, 6 chome, Yayoigaoka, Sanda, Hyogo 669-1546, Japan, **2** The Field, NPO corporation, Showa-dai 2-27-2, Takatsuki, Osaka 569-1041, Japan, **3** Department of Geography, Nara University, Misasagi 1500, Nara, Japan, **4** Institute of Natural and Environmental Science, University of Hyogo, 6 chome, Yayoigaoka, Sanda, Hyogo 669-1546, Japan.

Corresponding author: *Atsuko Takano* (takano@hitohaku.jp)

Introduction

The digitization of museum collections has recently become a major topic of discussion. For example, the Naturalis Biodiversity Center is in the process of digitizing 37 million objects. The Finnish Museum of Natural History are running mass digitization programs for their entire collections. Furthermore, major herbaria around the world—such as Harvard University, the Royal Botanic Garden Edinburgh, the Royal Botanic Gardens, Kew, the French Muséum national d'histoire naturelle vascular plant herbarium, and the Meise Botanic Garden —have also conducted digitization. China and the United States promote specimen digitization as the national projects.

Compared to the aforementioned countries, Japan is rather behind. About 10 million herbarium specimens are housed in universities and museums in Japan, and less than 2% of those were digitized and opened to the web. The largest herbarium database in Japan with specimen images is the Shimane University virtual herbarium (http://tayousei.life.shimane-u.ac.jp/harbarium/). Currently, ca. 100,000 specimen images are available online.

To facilitate digitization specimens in Japan, we explored a simple and inexpensive imaging method and developed a custom-made digitization system for herbarium specimens using a mirrorless interchangeable-lens camera (MILC) and a large bank light system. The system we have developed is not only simple and inexpensive, but requires minimal space, could be managed by part-time workers, and makes it possible to easily obtain multiple standardized digital files of several kinds. This system is

suitable for small and medium-sized herbaria where staff, space, and budget are limited, as well as for larger herbaria with larger numbers of staff members and stations. It may also be applied when digitizing other kinds of collections, including entomological, mineralogical, and fossil collections.

Materials and methods

As our target for digitization, we chose the Shoei Junior College herbarium collection, which was previously known as SHO and is currently one of the collections at the Museum of Nature and Human Activities, Hyogo. The collection consists of ca. 250,000 specimens of vascular plants.

To complete digitization within a limited time and budget, we decided to use the minimum acceptable quality point (MAQP) for the obtained images before developing equipment. Our terms were as follows:

1. Images should be usable and suitable for long-term storage.
2. Images should have enough resolution to withstand expansion up to 150% on a display monitor and be capable of withstanding life-size, high-definition printing.
3. Images should have applicable OCR for data transcription from the specimen label. To increase the accuracy of the OCR output, images obtained should have sharp margins, and be flat with minimal distortion.
4. Imaging should be finished within two to three years using the same hardware (camera and lens) to keep the quality of all images consistent. Hardware lifespan is generally in the range of three to five years; that of a digital camera may be shorter.

After some trial and error, we chose the SONY α6300 (ILCE6300), APS-C sensor, digital e-mount camera, and an FE 35 mm F.2.8 ultra- compact wide-angle lens for the Sony e-mount full frame (Samyang Optics SYIO35AF-E 35 mm F/2.8). Specimen images obtained using this combination of camera and lens are 5100×3500 pixels or ca.25 MB in size.

To record the color of herbarium specimens precisely at reasonable cost, LED light with a high color rendering property was selected (039 SH50 Pro-S LED Lamp, China), and diffuser film was also chosen to be durable enough for frequent changes over a long period (Savage Translum^{TM}, U.S.A). A large light bank system was designed to apply sufficient light above the specimen in a manner similar to a skylight.

We constructed our own copy stand for imaging specimens, adapted a lightweight aluminum frame structure with excellent durability and practical use strength for reducing camera shake (Yuki corp., Aichi Japan). On the copy stand, we put a mark to

indicate where to place herbarium sheets and the GIN-ICHI Silk Gray Card.

Results

The imaging system we developed costs ca. 314,000 JPY in total (US $2,778), including 170,000 JPY for the MILC camera, lens, adaptor and monitor, 90,000 for the bank light system (incl. LED lights, ramp folders, diffusion film, etc.), and 54,000 for the camera stand.

Specimen images obtained using our method are of a quality suitable for OCR output. All procedures, from pre-digitization curation to storage and archiving of images, were performed by two part-time workers between 9 am and 5 pm each day, four days a week. The speed of imaging depended on how many specimens needed conservation or clean-up before and/or after imaging. Therefore, we sometimes obtained only 1,200 images (300 specimens) per day. In contrast, for woody or large herbaceous specimens without roots (and associated soil), imaging ran smoothly, and we could obtain up to 4,000 images (1,000 specimens) per day. From the start of guidance and training on Nov 10, 2017 to July 4, 2018, a total of 73,180 specimens were imaged and stored as RAW and JPEG files. It means, 571 herbarium specimens were imaged per day on average.

Discussion

Digitization of herbarium specimens benefit both curators and stakeholders: For stakeholders, it becomes possible to access a digital voucher for each specimen remotely via the internet, and for curators, it reduces the need for specimen handling and makes semi-automated label data extraction by OCR possible. Further, crowdsourcing the manual data entry of specimen labels can be considered, given remote access to specimen images. Specimen processing, from mounting until manual data entry, can be facilitated, updated, and automated wherever possible as technology develops. The imaging of herbarium specimens is the first important step in this process.

Acknowledgements

We would like to give sincere thanks to Prof. Dr. Kunio Iwatsuki, Director Emeritus, Museum of Nature and Human Activities, Hyogo (HYO) & Professor Emeritus, The

University of Tokyo, for his donation to HYO and warm encouragement for this project. Thanks also go to two reviewers and Dr. Sandra Knapp, the handling editor of this paper for their critical suggestions and constructive comments. Enago (www.enago.jp) provided the English language review for this paper. Moriya Sangyo Co. Ltd. donated some consumables. Part of the study was also supported by JSPS (no. 26440227 to A.T.).