Report on trial of SatScan tray scanner system by SmartDrive Ltd.

Vladimir Blagoderov, Ian Kitching, Thomas Simonsen & Vincent S. Smith
Natural History Museum, Sackler Biodiversity Imaging Lab
London, UK, March 8 – April 9, 2010

Background

The recent House of Lords Science and Technology Committee Report on Taxonomy and Systematics[1] acknowledged the magnitude of the challenge facing taxonomists as they seek to document biodiversity on a global scale. The Committee noted that *"the rate of progress by the UK taxonomic institutions in digitising and making collections information available to be disappointingly low. Unless a more strategic view is taken of how they can contribute to the development of the field of biodiversity informatics, there is a significant risk of damage to the international reputation of major institutions such as The* [Natural History Museum] *NHM"*. The NHM holds more than 78 millions specimens in its collection. Current efforts to digitise these specimens are fragmented into a series of small-scale labour intensive projects, usually imaging single specimens, one at a time. A more coordinated effort to digitise this collection on an industrial scale offers the prospect of dramatically reducing the time taken to digitise the entire collection. Such an effort could provide worldwide specimen level access to the collection through the internet and dramatically improve the collections utility for the public and scientific community.

Aims

Smartdrive Ltd.[2] has developed a prototype imaging system, SatScan[tm], that captures digitised images of large areas while keeping smaller objects in focus at very high resolution. The system was set up in the Sackler Biodiversity Imaging laboratory of Natural History Museum on March 8, 2010 for a one-month trial. A series of projects imaging parts of the entomological, botanical and palaeoentomological collection were conducted to assess the systems utility for museum collection management and biodiversity research. The technical and practical limitations of the system were investigated as part of this process.

System Description

SatScan[tm] by Smartdrive (http://www.smartdrive.co.uk) is a combination of hardware and software that provides automated capture of lower resolution images, which are then assembled ("stitched") into a larger panoramic image, generating an extremely high-resolution final image. A camera with the attached lens is moved in two dimensions along precision rails positioned above the imaged object. This method maximizes depth of field of the captured images and minimizes distortion and parallax artefacts.

Methodology

In most cases we acquired original tile images during working hours at the Museum (8 am-6 pm) and conducted batch-stitching overnight to produce the final stitched images.

Configuration tested
- Basler A631FC 1/2" CCD camera
- Edmund Optics 0.16x telecentric lens #NT56-675

---

[1] http://www.publications.parliament.uk/pa/ld200708/ldselect/ldsctech/162/162.pdf
[2] http://www.smartdrive.co.uk/

Output specification
- Field of view: 35.5 x 27.5 mm
- Original images: 1280 x 960
- Final images: up to 21000 x 21000
- Resolution: ~35 px/mm
- Minimal resolved structures: 0.06 - 0.1 mm
- Depth of field: 10 - 80 mm
- File formats: 24bit BMP or LZW-compressed tiff
- File size (15000 x 14000): ~780Mb (BMP), 340Mb (TIFF)
- Exposure: 1-1000 ms
- Scanning time (45 x 50 cm): 5-7 min, depending on exposure
- Stitching time, 200-400 tiles: 5:30-9:30 min

Example images
- A sample drawer of smaller insects (Diptera: Ditomyiidae)
  http://sciaroidea.info/node/44309
- A sample drawer of medium-size insects (Heteroptera: Coreidae)
  http://sciaroidea.info/node/44310

System at work:
- http://www.youtube.com/watch?v=sQXQN1FEZxk

Results

Given an average capture time of 5-7 minutes, a trained operator can process up to 60 drawers of specimens each day. Up to 90 final pictures can be stitched in 12 hours (e.g. overnight).

Timing:
- Average scanning time (for all tests)
  7:49 min (without stitching)
- Avg. time of acquiring final stitched image, per drawer (Apidae, Sciaroidea & Oecophoridae)
  12:05 min
  n.b. We encountered the fewest errors in these tests and consider this time to be typical for a trained operator and bug-free software.
- Average time of acquiring final stitched image, per drawer (for all tests):
  17:26 min
  The increase of average time is due to the large size of the drawers and numerous stitching errors during scanning of the British lichens collection.
- Total time spent on the project.
  142:37 hr.

File sizes
- Image size for typical drawer (450 x 400 mm; 15819 x 14617 px, 165 tiles)
  108 Mb (JPEG)
  337 Mb (LZW compressed TIFF)
  661 Mb (BMP)
- Largest scanned unit: Palaeontology drawer (550 x 540 mm; 19708 x 19479 px, 300 tiles)
  520 Mb (LZW compressed TIFF)

When published online, *Zoomify* technology[3] can be used to reduce amount of transferred data (see exemplar images above). This method allows users to download and view in their browser a high resolution image of a region of interest only. A typical Zoomified image would be 0.8x the size of the corresponding compressed TIF file.

---

[3] http://www.zoomify.com/

Lighting

Twelve fluorescent tubes provide sufficient light for short exposures (20-40 ms). With the aperture completely open, the amount of light can be reduced but at the cost of sacrificing some depth of field. The high sides of some drawers resulted in a degree of marginal shadowing. In cases where specimens are position very close to the edge of a draw this shadow can obscure parts of affected specimens.

Biological specimens

A number of test scans were conducted with different types of entomological, botanical and palaeoentomological specimens.

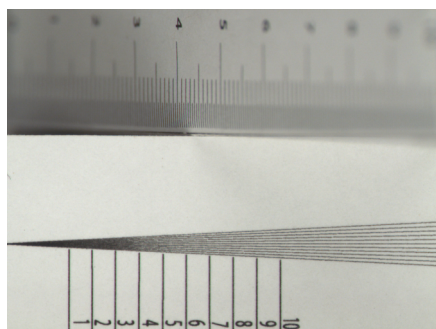**Table 1. Biological specimens imaged in the current trial**

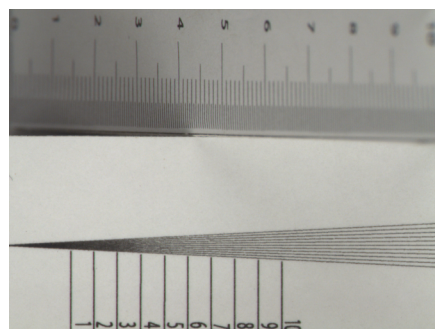|	# of files	Disk space (original & final files)
Original files (total)	102,352	354 Gb
Stitched files (total)	491	207 Gb
Including: Apidae types	51	14.4 Gb
Coleoptera, various	13	3.66 Gb
Morphinae, types	15	5.05 Gb
Sphingidae, glass bottomed drawers	6	3.0 Gb
Oecophoridae	87	34.5 Gb
Asilidae	44	23,6 Gb
Sciaroidea	97	28.9 Gb
Fossil Diptera (Isle of Wight)	6	3.05 Gb
Smith's British lichen collection	84	31,8 Gb

Technical tests: DoF/Resolution:

A series of test shots was conducted to investigate the relationship between aperture, Depth of Field (DoF) and resolution for the given lens/camera combination. DoF varies from 6 mm with the aperture fully open to 80 mm when it is completely closed. There is no significant loss of resolution until aperture dial reaches a dot on the lens housing (~4 EV stops); after this resolution and picture sharpness drops dramatically. Resolution was identified on a test printout (see figures). The smallest resolvable structure was identified as a minimal width of a resolvable black/white pair in the picture.

Table 2. Aperture, depth of field and resolution of tested system
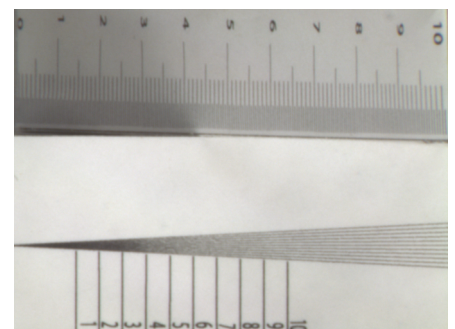
Aperture	Exposure, ms	DoF, mm	Resolution	~Smallest resolvable structure, μm
Completely opened	11	6	3-4	56
	21	9	3-4	56
	31	14	4	59
Dot	41	17	4	59
	74	30	4-5	65
	184	44	5	69
	374	55	6	78
Completely closed	810	80	8	98

Fig. 1. Aperture completely open   **Fig. 2.** Aperture set at the dot (5.6?)   **Fig. 3.** Aperture completely closed

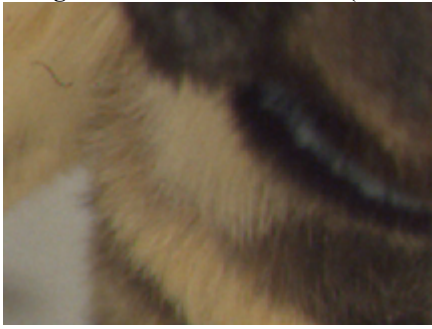**3. Figs. 4-6. Loss of resolution (100% crop of photo of Acherontia atropos (Lepidoptera: Sphingidae))**

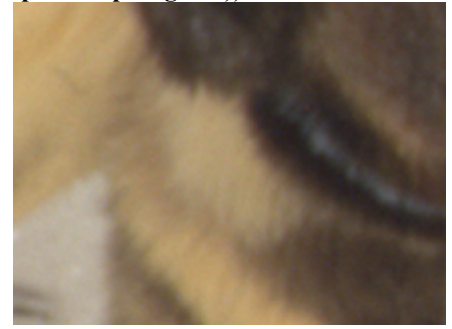**Fig. 4.** Aperture completely open **Fig. 5.** Aperture set at the dot (5.6?) **Fig. 6.** Aperture completely closed

Problems encountered

1. Dynamic range of images, as displayed on the PC LCD monitor, was unacceptable at the start of the trial. Very light and very dark areas could not be satisfactorily displayed in the same set of images. Setting the gamma correction to 2.2 solved the problem.
2. Maximum scanning area ~ 500 x 600 mm – insufficient for some larger drawers.
3. Errors during scanning. It seemed that errors could sometimes be generated while scanning empty areas of uniform background, in which there is insufficient visual information for stitching adjacent tiles.
4. Errors during stitching. Such errors require some response from operator and can be very troublesome. In particular, if such an error occurred during an overnight batch stitching, the entire sequence was halted, resulting in stitching of tile sets in the job list after that which generated the error not being done. This was particularly disruptive if it occurred early in a batch job.
5. Focusing is possible only by adjusting lens/camera or object height, which requires a lot of time with the current set-up.
6. Inconvenient access to scanning area.
7. Large size of final images prevents from quick preview.
8. Due to the large size of the stitched images, disk space ran out very quickly, but there are no safeguards and an inexperienced operator may encounter a permanent error.
9. Determination of the region of interest (scanning area) is not sufficiently flexible and fast.

Conclusion

General impression and utility

The system is best suited for capturing images of drawers containing numerous, uniformly positioned, medium-size biological specimens. Excellent results were achieved with drawers containing large and medium-size beetles, moths and butterflies. For such objects, sufficient information is usually preserved to allow identification. Objects smaller than 10 mm could not be imaged so adequately as to retain any taxonomic information, although such images could be used in other ways. Specimen labels and barcodes (when not obscured by the specimen) could be easily read from the digitised image.

SatScan could be used for industrial-scale imaging of biological collections. The collections of the department of Entomology contain approximately 85,000 drawers that could be usefully imaged at the current level of resolution with this system, out of the 135,000 draws in the whole entomology collection. Quick calculations show that this work could be completed in ~2024 person-days (ten person-years) using one system.

Our tests of the system have identified a number of research, collection management and public engagement applications:

Research uses:
- Acquiring a large number of images for use with automated identification software
- Morphometric analysis of individual specimens
- Specimen population counts
- Manual identifications
- Studies on colour pattern variations

Collection management applications:
- Accurate specimen count of the entire collection
- Collections audit and security
- Improving accessibility to the collection
- Saving curators' time by reducing trivial requests through use of images posted on-line
- Improving curation: updating identifications and taxonomy through crowd sourcing
- Encouraging typification, discovery of unrecognized/unlabelled types and checking existing types
- Populating KE EMu

Public engagement:
During the test period, the system was on display in the Sackler's Biodiversity Imaging Lab forming an external part of the Explore tour of the Darwin Centre. Its use generated significant interest among the general public (judging from the number observers we saw) owing to the number of moving parts and the highly visual nature of the output (images, assembled and stitched on a large panoramic screen). A system professionally placed with live output being displayed to the public could be used to stimulate public interest in the natural history and to engage their participation in crowd sourcing activities (e.g. data transcription) through citizen science projects. Possibly crowd sourcing activities include extracting data from specimens and labels, cropping images and numbering specimens, even identification of specimens.

Caveats
The major impediment to large-scale use of SatScan at the NHM is the lack of a method to unambiguously link specimens and the digitised image together (e.g. via a unique number, probably expressed as a barcode). Many specimens in the NHM collection (especially in Entomology) lack a unique unambiguous identifying number. Even when this is present, it is not necessary visible without manual manipulation of the specimen. Without a solution to this problem, moving specimens between draws (e.g. during standard curation activities) would break the vital link between the specimen image and its physical location in the collection. This represents a large and possibly critical barrier to the wide-scale application at the NHM. Possible solutions include attaching physical identifiers to specimens (e.g. barcodes) at the point of digitisation or automatically / manually attaching virtual identifiers as a digital layer on specimen images (possibly backed up by the physical identifiers as dictated by curation activities). A more radical solution could involve wholesale digital curation of specimens (i.e. freezing the recuration of specimens in collection draws after they have been digitised or redigitising draws after recuration activities).

Additional limitations concern the web space necessary to mount and backup digitised collections the internet; development of the software necessary to support the workflow (transferring images and associated metadata from the point of digitisation to the web); the association of images with existing metadata (e.g. information held in the KeEmu database); the association of images with new metadata (e.g. transcribed label information). Another concern, although outside the scope of this trail is that the dorsal images captured by the SatScan system are of limited utility for certain parts of biological collections (e.g. palaeontological or mineralogical collections). Similarly images of spirit collections (dorsal or otherwise) containing collection 'lots' are likely to have limited scientific or public value, although even these can still have a role in collection management.

Recommendations for improvement[4]

Software
- SatScan: manual input of scan starting point
- SatScan: manual input of row/columns number
- Analyze: skip folders with stitching errors and proceed directly to next folder; produce the list of skipped folders
- Analyze: prompt to select saving folder in the beginning
- Analyse: automatic move of folder with original images to backup location on a separate drive after successful stitching
- Analyse: change default file type at start-up
- Analyse : option to save small JPEG (1/5 resolution) as well as TIFF or BMP

Hardware and ergonomics
- Better access to scanning area from front, say through a counterbalanced "sash window" arrangement
- Stopper rail more easily adjustable and solidly fixed

Imaging system
- Calibration of camera/lens combination for each aperture to determine DOF/front-back focus/resolution

Possible long-term further developments

Software and hardware
- Remote aperture control from computer
- SatScan: maximum size of ROI to be increased

Software
- Scanning and stitching reduced to take together ~5 min on average (12 min at present)
- Automatic determination of number of specimens in the image
- Automatic cropping of individual images

Hardware
- Adjustable camera holder for different lens/camera combinations
- Adjustable bed for focusing (~5cm up/down)
- Book holder (cradle)
- Scanning area increased to 1.0 x 0.7m

Imaging system
- Option for different lens/camera combinations
- Resolution increased to 150 px/mm (current compact cameras achieve this, the current video capture camera has 35 px/mm)

Next steps
The system opens the door to a complete change in the way we digitise much of the collection. However, significant challenges remain. Software that supports the workflows required to assign and print identifiers (virtually and or physically) has yet to be developed and this must integrate with our existing systems (notably KeEmu and DAMS). We recommend the development of a project proposal that addresses these challenges, with a view to development of a pan European EU funded project.

<div align="right">

Vladimir Blagoderov, Ian Kitching
Thomas Simonsen & Vincent S. Smith
Natural History Museum, London
May 6th 2010

</div>

---

[4] Some of the points listed here have been addressed by SatScan during the final testing of the machine and preparation of this document. See the addendum by SatScan for further information.

Executive Summary

The SatScan system facilitates the capturing of a very large number of good quality images in a very short time. Large parts of the NHM collection could be digitised in dorsal view extremely quickly. These images have a wide range of uses across research, collection management and public engagement. Scalability of the system is limited by our desire to assign unique identifiers (a number and/or a barcode) to specimens, and the cropping of these images. Without these identifiers digitised images will have limited long term value. The assignment of specimen level identifiers is potentially labour intensive.

Options for assigning identifiers were not investigated as part of this trail but include the use of physical labels on each specimen (with significant resource implications and a significant volunteer workforce) and the use of virtual identifiers (as a virtual layer over the image, perhaps automatically assigned, and possible coupled with physical labels attached to specimens as dictated by recuration activities). Intuitive software (with a web interface) needs to be developed to facilitate this process, including support for cropping of an image and the automatic assignment and printing of an identifier label. On-demand assignment of identifiers would allow us to prioritize the digitisation but it will represent a significant change to the way we curate our collections and would require sustained and ongoing support from Collection Management. Additional concerns relate to the amount of storage space required to manage images, connection with existing digital systems and the utility of dorsal images for certain parts of the collection. These problems need to be addressed as part of a larger scale study.

Vladimir Blagoderov
Ian Kitching
Thomas Simonsen
Vince Smith

Natural History Museum, London
May 6th 2010