Automated workflows using Quantitative Colour Pattern Analysis (QCPA): A guide to batch processing and downstream data analysis

Cedric P. van den Berg[1,2 †], Nicholas D. Condon[3], Cara I. Conradsen[1], Thomas E. White[2], Karen L. Cheney[1]

[1] School of Biological Sciences, The University of Queensland, Australia

[2] School of Life and Environmental Sciences, The University of Sydney, Australia

[3] Institute for Molecular Bioscience, The University of Queensland, Australia

† Current address: School of Biological Sciences, The University of Bristol, UK

Corresponding author:

Cedric van den Berg, c.vandenberg@bristol.ac.uk

Abstract

Animal and plant colouration presents a striking dimension of phenotypic variation, the study of which has driven general advances in ecology, evolution, and animal behaviour. Quantitative Colour Pattern Analysis (QCPA) is a dynamic framework for analysing colour patterns through the eyes of non-human observers. However, its extensive array of user-defined image processing and analysis tools means image analysis is often time-consuming. This hinders the full use of analytical power provided by QCPA and its application to large datasets. Here, we offer a robust and comprehensive batch script allowing users to automate many QCPA workflows. We further provide a set of useful R scripts for downstream data extraction and analysis. These scripts will empower users to exploit the full analytical power of QCPA and facilitate the development of customised semi-automated workflows. Such quantitatively scaled workflows are crucial for exploring colour pattern space and developing ever-richer models of visual perception in animals other than humans. These advances will, in turn, facilitate the testing of hypotheses on the function and evolution of vision and signals, which are otherwise computationally unfeasible.

Keywords: Open access, ImageJ, java, R software, automation, animal colouration

1. Introduction

Understanding the perception of visual information by non-human observers is crucial to studying the ecology and evolution of animal and plant colouration. The last two decades have seen the creation and widespread adaptation of tools and methods that allow researchers to simulate or approximate aspects of animal vision, such as colour contrast perception (e.g. Endler and Mielke, 2005; Gawryszewski, 2018; Kemp et al., 2015; Renoult et al., 2017; Vorobyev and Osorio, 1998) and spatial vision (Caves and Johnsen, 2018; Godfrey et al., 1987). These advances coincide with the steady development of colour pattern analyses (e.g. Chan et al., 2018; Endler, 1991; Endler, 2012; Stoddard et al., 2014; Van Belleghem et al., 2018; van den Berg et al., 2020) as well as their integration into increasingly comprehensive collections of tools and functions such as Pavo (Maia et al., 2019) or the Multispectral Image Calibration and Analysis toolbox (MICA) (Troscianko and Stevens, 2015).

Quantitative Colour Pattern Analysis (QCPA) (van den Berg et al., 2020b) is a recent and powerful addition to this landscape. It provides a dynamic analytical framework for analysing visual scenes through the eyes of ecologically relevant observers and is integrated into the MICA toolbox (Troscianko and Stevens, 2015). It allows users to choose preferred tools and analyses uniquely suitable for analysing spatiochromatic information. Briefly, it achieves this by taking full-spectrum images as its input (thereby generating '.mspec' files) before applying a suite of models which can consider the spectral, spatial, and temporal sensation of non-human viewers, to ultimately give image-based and numerical outputs that summarise the structure of visual scenes and stimuli (van den Berg et al., 2020b).

Preparing and processing calibrated images in the framework, however, remains tedious. This is due to the time needed to translate calibrated digital images into cone catch images, user-guided input for identifying regions of interest (ROIs), and the subsequent application of image processing and analyses. Saving and labelling multiple output files, as well as extracting data for subsequent statistical analysis, also remains a tiresome process. As a result, applying the QCPA at quantitative scales, such as analysing hundreds or thousands of images at multiple viewing distances, is nearly impossible in its original form.

An image currently needs hours of manual, repetitive work to obtain the full numerical output from the various types of analyses in the QCPA framework. This is especially true when considering multiple viewing distances and regions of interest (ROIs). Consequently, to date, most studies that have used the framework do not consider more than a few dozen individual observations with a limited subset of available image statistics (e.g. Nokelainen et al., 2021; Rodríguez-Morales et al., 2021). This is problematic as quantifying ecologically relevant effects, such as colour pattern variability in natural populations or the analysis of behavioural data, requires sufficiently large sample sizes to provide adequate statistical power. This holds particularly true for a method producing hundreds of image statistics from a single observation, such as the QCPA, where the number of available image statistics easily outnumbers observations (van den Berg et al., 2022). Considering all possibly relevant image statistics and deducing a relevant set of parameters is often the desired approach instead of pre-emptively narrowing down the number of considered image statistics (van den Berg et al., 2022). However, such deductive approaches require adequate statistical solutions and remain an intriguing challenge in using colour pattern space (Stoddard and Osorio, 2019; van den Berg et al., 2020b; van den Berg et al., 2022).

The QCPA framework is, therefore, in urgent need of automation to facilitate its complete application to large datasets. However, unlike other image analysis tools available to visual ecologists, such as PAVO (Maia et al., 2019), MICA is an ImageJ (Schneider et al., 2012) plugin written in Java (Arnold et al., 2005). It is intentionally tailored towards accessibility via graphical user interfaces (GUIs) rather than command line prompts. Unlike R software (R Core Team, 2021) or Python (van Rossum, 1995), frequently taught in undergraduate and HDR Biology degrees worldwide, Java is not a programming language that biologists are readily familiar with. Therefore, writing scripts and macros for a complex network of GUI-guided tools in an unfamiliar language is a hurdle for most users of the QCPA framework seeking to batch-process large volumes of image data. This creates the potential for unintentional gatekeeping by researchers with access to programmers familiar with writing scripts in JAVA and an adequate understanding of the framework's mechanisms. The result is a hidden custom of informally traded scripts, which undermines fast and equal access to methods in the field.

Here, we present a dynamic batch-processing extension to QCPA, which allows users to apply almost the entire QCPA framework or selected parts of it to large datasets via a flexible, GUI-guided input. We further provide a set of R scripts that efficiently extract numerical data from the

consecutively stored output files for downstream analysis. We describe the functionality of both—the QCPA batch script and complementary R scripts—and provide detailed worked examples. While this script does not provide a complete solution for every possible application of the QCPA, we hope it will offer a viable solution to most. This batch script should also help researchers customise their automated pipelines, stimulating the open exchange of programming solutions via platforms such as open-access publications, open-access platforms such as GitHub or the dedicated user forum for the MICA toolbox.

2. QCPA batch script

2.1. Intended use

The QCPA batch script is intended to quantify spatiochromatic information of an animal against its visual background. As such, the script can also be applied to images without an animal or object of interest or an animal or object itself. The batch script is intended for users with moderate experience in the manual application of QCPA.

The script allows the user to choose the following desired analytical outputs individually: 1) Colour Adjacency Analysis (CAA), Visual Contrast Analysis (VCA), Boundary Strength Analysis (BSA) and particle analysis; 2) Local Edge Intensity Analysis (LEIA); 3. Colour Maps; and 4) GabRat. For a detailed discussion of these analyses, please see the original publications and their modifications for QCPA as listed in van den Berg et al. (2020).

Notably, the script requires cone mapping functions derived from calibrated cameras (several standard profiles included in the MICA toolbox) and currently does not permit using chart-based models. While covering many analyses available in QCPA and the MICA toolbox, more broadly speaking, it does not provide a comprehensive library of automation. However, the script is designed to be easily modified by the user. Such modifications are explicitly invited, and we encourage their open sharing on community platforms.

2.2. Data preparation

Image processing in QCPA is specific to a variety of user choices specifying the properties of the observer's visual system, the light environment in which a picture was taken as well as the intended target light environment, observer viewing distance, and a variety of parameters and processing choices that vary depending on the desired analyses. The batch script uses three approaches to minimise the need to repeat the number of times such input is required:

1. Pre-defined folder structure
2. Auxiliary files
3. Batch processing graphical user interface (Batch GUI)

2.2.1 pre-defined folder structure

The batch script requires input data to be organised in a specific way. This refers to the naming of files and the folder structure. For example, the script expects individual observations (i.e., animals or objects) to be grouped within an overarching folder (i.e., species or site) and file naming needs to be uniform. This allows the script to reliably detect individual observations while allowing the user to structure their data according to sites or taxa (see the provided datasets for specific examples).

2.2.2 Auxiliary files

The batch script uses the MICA toolbox's pre-existing approach to associate several files with each image to be analysed. In addition to the already existing need for a folder with the specified ROIs and the corresponding .mspec of each image, the batch script requires two additional text files specifying the rotation of each image and the cone mapping function (see the manual for details). These allow users to standardise the orientation of all images before analysis and permits the use of multiple cone mapping functions within the same dataset (e.g., multiple lighting environments).

2.2.3. Batch GUI

The batch script allows the user to specify which of the available analyses they want to conduct and what settings they want to use for them. The Batch GUI consists of a central interface and a suite of dedicated, tool-specific GUIs activated by user choices. These tool-specific GUIs provide helpful information and allow running all analyses with independent settings. Furthermore, the batch script remembers past user input and will pre-fill the previous choices by the user, helping to test settings and maintain track of chosen settings in case of incomplete runs. See the manual and worked examples for detailed instructions on how to use them. Users familiar with QCPA and MICA toolbox will find many GUIs similar or identical to existing ones. And indeed, this extension to the QCPA uses modified existing scripts within the QCPA and MICA toolbox.

2.3. Workflow recommendations

The Batch GUI, manual and worked examples contain a suite of recommendations in addition to those in van den Berg et al. (2020) and the empiricalimaging.com website that users might find informative. We recommend that the user validate the numerical accuracy of the batch script output on a small dataset before analysing large datasets. This also allows the user to confirm that the folder structure and auxiliary files have been arranged correctly. As the processing of large datasets can take many hours, days, or even weeks, we recommend that the user consider running sub-sections of larger datasets on different instances, such as multiple computers, servers, or cloud-based computing facilities. This will ultimately save time and prevent data loss while helping to detect faulty files in large datasets faster.

Like all approaches to visual modelling, image analysis with QCPA requires many different input choices by the user. Keeping a record of these choices is crucial for two reasons. First, it allows the user to keep track of settings, facilitating record-keeping and collaboration. Second, it provides for the publication of repeatable research. The latter is often an issue in visual modelling studies, as modelling choices are often poorly documented (White et al., 2015). To this end, the QCPA batch script keeps various detailed log files in the data output. These can easily be added to the supplementary information of any publications or uploaded/shared as part of the data.

3. R script library

We provide a set of tailored R-scripts enabling the extraction and compilation of batch QCPA output data into .csv tables that can be used for downstream analysis. These are formulated and commented on in a way that is aimed to facilitate modification by the user.

Specifically, these include:

1. Compiling data from CAA, VCA, BSA & particle analysis.

2. Compiling data from LEIA

3. Compiling data from GabRat

5. Normalising data

6. Preliminary data visualisation

4. Worked examples

We provide detailed worked examples with test data and corresponding toolbox files, such as cone mapping models. These aim to highlight the functionality of the batch script and allow users to familiarise themselves with examples related to their needs.

4.1 Trichromatic example without UV (sea slugs seen by a triggerfish, *Rhinecanthus aculeatus*)

Here, we showcase the use of the batch script by applying all modules in the script to investigate how a triggerfish (*R. aculeatus*) perceives two species of cryptic sea slugs (*Aphelodoris varia, Aplysia sp*.) from the east coast of Australia (Nelson Bay, NSW) photographed underwater (Fig. 1).

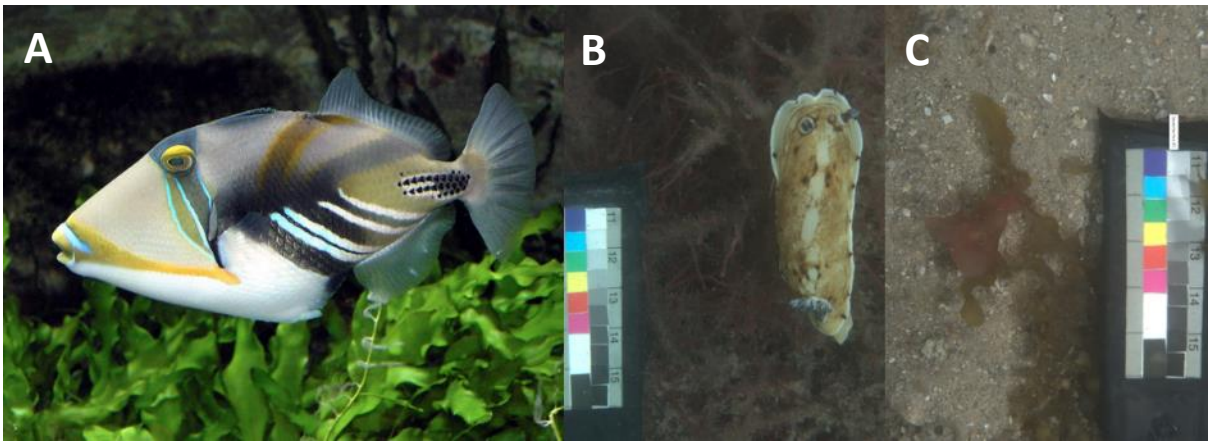

Figure 1 A) *Rhinecanthus aculeatus* B) *Aphelodoris varia* C) *Aplysia sp.*

4.2 Tetrachromatic example with UV (spiders seen by a bluetit, *Parus major*)

This example uses all batch script modules to analyse cryptic spiders (*Tamopsis brisbanensis*) found on eucalypt trees in Sydney, NSW (Australia) (Fig. 2).

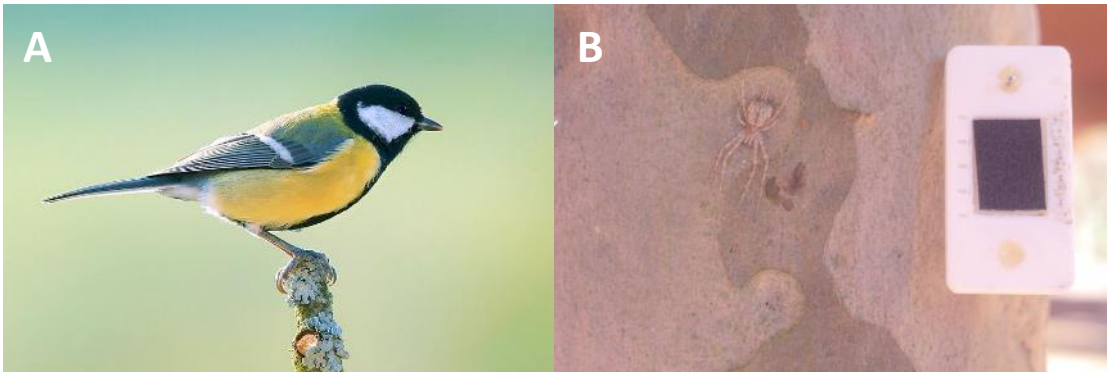

Figure 2 A) *Parus major* B) *Tamopsis brisbanensis*

To access these worked examples and corresponding explanations and instructions, please see the electronic Supplement.

5. Discussion

We provide a user-friendly, open-source batch processing extension to QCPA, allowing for the semi-automated analysis of large datasets within the QCPA framework. This greatly facilitates the full-scale use of the analytical power of the framework as intended by its original design (van den Berg et al., 2020b). The open-source code of the script is presented and structured with the deliberate purpose of enabling users to create customised versions of the batch script, which might be more suitable to their intended purposes. We further provide a set of R-scripts aiding in the downstream analysis of the data generated by the batch script. These scripts are presented and structured in a way that facilitates users to customise them to their needs. A detailed manual guiding further supplements the batch script as to how to prepare & analyse the image data. Lastly, the user is provided with three worked examples and corresponding test data, covering many contexts in which users are likely to use the script.

The script enables the semi-automated analysis of large image datasets, computing hundreds of colour pattern descriptors at multiple viewing distances for dozens, hundreds, or even thousands of observations. This has previously been impossible due to the time-consuming manual processing of images in QCPA. As QCPA and other colour pattern analyses using visual modelling approaches (e.g. Maia et al., 2019; Troscianko and Stevens, 2015) merely relate to approximations of early-stage visual processing (e.g. Endler and Mielke, 2005; Marr, 2010; van den Berg et al., 2020a; van den Berg et al., 2020b; Vorobyev and Osorio, 1998), considering the context and task-dependent relationships with and between a broad range of descriptors is critical (see van den Berg et al., 2022 for discussion). Thus, the presented batch script provides a crucial step towards such analytical approaches.

As QCPA provides a mainly conceptual approach towards the combination of visual modelling and colour pattern analyses, automating workflows and advancing customisability represent essential stepping stones towards the continued development and refinement of the framework and its contribution towards alternative methods. By providing an inviting, well-annotated, intuitive, open-access and open-source environment, we hope to facilitate the application of and, importantly, eventual modifications and additions to the framework and its combined use with other existing methodologies. Cumulatively, this philosophy will not only advance the development and testing of methods in the field but, crucially, will facilitate the increased critical understanding of and familiarity with existing methodology among members of the field. On a more tangible note, this script can easily be adjusted to enable the analysis of different data structures or include processing steps and analyses available in QCPA or the MICA toolbox currently not considered in the script.

6. Conflict of Interest

The authors declare no conflict of interest.

7. Funding

This work was supported by the Australian Research Council (FT190199313 to K.L.C), Hermon Slade Foundation grant to TW (HSF20082), and UQ Research Stimulus (Allocation Two) Fellowship to Cedric. Nicholas Condon is supported as a CZI Imaging Scientist by grant number 2020-225648 from the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation.

9. Supplementary data

Batch script ImageJ plugin, Batch script manual, worked examples guide:

https://github.com/cedricvandenberg/QCPA-batch-script

Test data for the worked examples:

https://doi.org/10.48610/3cdcc1f

10. Author contributions & acknowledgements

CB – Overall concept, testing, coding, original draft, manuscript writing & review, Funding, NC –coding batch script, Manuscript review, CC – R scripts, Manuscript review, TEW – Manuscript review, Funding, KLC – Original concept, Funding, Manuscript review.

We thank Alfonso Aceves for providing the test data for the UV examples.

11. References

Arnold, K., Gosling, J. and Holmes, D. (2005). *The Java programming language*. Addison Wesley Professional.

Caves, E. M. and Johnsen, S. (2018). AcuityView: An r package for portraying the effects of visual acuity on scenes observed by an animal. *Methods Ecol. Evol.* **9**, 793–797.

Chan, I. Z. W., Stevens, M. and Todd, P. A. (2018). PAT- GEOM: A Software Package for the Analysis of Animal Patterns. *Methods Ecol. Evol.* 2041–210X.13131.

Endler, J. A. (1991). Interactions between predators and prey. In *Behavioural Ecology* (ed. Krebs, J. R.) and Davies, N. B.), pp. 169–196. Oxford: Blackwell Scientific.

Endler, J. A. (2012). A framework for analysing colour pattern geometry: Adjacent colours. *Biol. J. Linn. Soc.* **107**, 233–253.

Endler, J. A. and Mielke, P. W. (2005). Comparing entire colour patterns as birds see them. *Biol. J. Linn. Soc.* **86**, 405–431.

Gawryszewski, F. M. (2018). Color vision models: Some simulations, a general n -dimensional model, and the colourvision R package. *Ecol. Evol.* **8**, 8159–8170.

Godfrey, D., Lythgoe, J. N. and Rumball, D. a (1987). Zebra stripes and tiger stripes: the spatial frequency distribution of the pattern compared to that of the background is significant in display and crypsis. *Biol. J. Linn. Soc.* **32**, 427–433.

Kemp, D. J., Herberstein, M. E., Fleishman, L. J., Endler, J. A., Bennett, A. T. D., Dyer, A. G., Hart, N. S., Marshall, N. J. and Whiting, M. J. (2015). An integrative framework for the appraisal of coloration in nature. *Am. Nat.* **185**, 705–724.

Maia, R., Gruson, H., Endler, J. A. and White, T. E. (2019). pavo 2: New tools for the spectral and spatial analysis of colour in r. *Methods Ecol. Evol.* 0–2.

Marr, D. (2010). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. The MIT Press.

Nokelainen, O., Scott-Samuel, N. E., Nie, Y., Wei, F. and Caro, T. (2021). The giant panda is cryptic. *Sci. Rep.* **11**, 1–10.

R Core Team (2021). R: A Language and Environment for Statistical Computing.

Renoult, J. P., Kelber, A. and Schaefer, H. M. (2017). Colour spaces in ecology and evolutionary biology. *Biol. Rev.* **92**, 292–315.

Rodríguez-Morales, D., Tapia-McClung, H., Robledo-Ospina, L. E. and Rao, D. (2021). Colour and motion affect a dune wasp's ability to detect its cryptic spider predators. *Sci. Rep.* **11**, 1–9.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675.

Stoddard, M. C. and Osorio, D. (2019). Animal Coloration Patterns: Linking Spatial Vision to Quantitative Analysis. *Am. Nat.* **193**, 000–000.

Stoddard, M. C., Kilner, R. M. and Town, C. (2014). Pattern recognition algorithm reveals how birds evolve individual egg pattern signatures. *Nat. Commun.* **5**, 1–10.

Troscianko, J. and Stevens, M. (2015). Image calibration and analysis toolbox – a free software suite for objectively measuring reflectance, colour and pattern. *Methods Ecol. Evol.* **6**, 1320–1331.

Van Belleghem, S. M., Papa, R., Ortiz-Zuazaga, H., Hendrickx, F., Jiggins, C. D., Mcmillan, W. O. and Counterman, B. A. (2018). Patternize : An R package for quantifying color pattern variation. *Methods Ecol. Evol.* **9**, 390–398.

van den Berg, C. P., Hollenkamp, M., Mitchell, L. J., Watson, E. J., Green, N. F., Marshall, N. J. and Cheney, K. L. (2020a). More than noise: Context-dependant luminance contrast discrimination in a coral reef fish (*Rhinecanthus aculeatus*). *J. Exp. Biol.* **223**, jeb.232090.

van den Berg, C. P., Troscianko, J., Endler, J. A., Marshall, N. J. and Cheney, K. L. (2020b). Quantitative Colour Pattern Analysis (QCPA): A comprehensive framework for the analysis of colour patterns in nature. *Methods Ecol. Evol.* **11**, 316–332.

van den Berg, C. P., Endler, J. A., Papinczak, D. E. J. and Cheney, K. L. (2022). Using colour pattern edge contrast statistics to predict detection speed and success in triggerfish (*Rhinecanthus aculeatus*). *J. Exp. Biol.*

van Rossum, G. (1995). Python tutorial, Technical Report CS-R9526. *Cent. voor Wiskd. en Inform.*

Vorobyev, M. and Osorio, D. (1998). Receptor noise as a determinant of colour thresholds. *Philos. Trans. R. Soc. B Biol. Sci.* **265**, 351–358.

White, T. E., Dalrymple, R. L., Noble, D. W. A., O'Hanlon, J. C., Zurek, D. B. and Umbers, K. D. L. (2015). Reproducible research in the study of biological colouration. *Anim. Behav.* **106**, 51–57.