Supplementary data

1. Literature Review Details

The literature review data could be found in the excel file: <Supplementary-data\_Size\_Orientation\_Form\_review.xlsx>.

* 1. Query methodology:

1. The different query tested are presented in Table 1 , with their associated csv files corresponding to the query results.
2. Sort by relevance, Influencing factors for sorting by Scopus RELEVANCE:
   * Number of times the term is present in a document.
   * Importance/specificity of the word (e.g. Experiment = common word vs xenomorphic, less weight in the calculation.)
   * Where the word is found. Higher relevance if the word is in the title, abstract or keywords.
   * The earlier the word appears in the document, the more weight it has.
   * Proximity: the closer the words in a query are to each other the higher the score.
   * If all the words of a query are in the same field of a document the higher the score will be.
   * Factors that are not considered: The number of clicks, the date of publication, the number of citations
   * references: <https://service.elsevier.com/app/answers/detail/a_id/14182/supporthub/scopus/kw/relevance/>
3. Limit of Scopus sets at 2000 documents (limit of Scopus export), the last pages of articles are not relevant to the study after checking
4. Citation per documents mapping (VOS viewer)
5. Minimum number of citations 0
6. Document selected 2000
7. Citations type by documents, links are citations in-between documents
   1. Explanation of the different steps:

* Step1, base request (2044 articles) obtained after a study of the keywords used in the literature and comprehension of Scopus queries.
* Step 2 add cells and texture, does not change that much but good article not linked with cells but texture showed up (2060 articles).
* Step 3, removes the form adjectives in the query: 42 262 articles in English, test on the 2000 first. The query is noisier but the same principal articles showed up (42 262 articles).
* Step 4, the texture within the query to get a broader result than step 1 and 2 but less than step 3 (2127 articles).

1. Notes on datasets

All the images datasets are present in the compressed file <images\_dataset.zip>.Table 2 shows the various remarks regarding these datasets and presents the corresponding descriptor table for each dataset. The files <loadings\_PCA1\_form.csv> and <loadings\_PCA1\_galena.csv> are the loadings of the first PCA made using all the descriptors computed on the corresponding image dataset. From these two files, it is possible to generate the scree plot and visualize the eigenvector that helps select the features presented in the second PCA (Table 3 and 4 within the manuscript).

Table 1. Query made on Scopus (02.05.2023).

|  |  |  |
| --- | --- | --- |
| **Step** | **Query** | **CSV file** |
| 1 | ( TITLE-ABS-KEY ( "mathematical analysis" OR "numerical analysis" OR descriptor\* OR factor\* OR "Image analysis" OR "computer vision" OR "Image processing" ) AND TITLE-ABS-KEY ( grain\* OR particle\* ) AND TITLE-ABS-KEY ( orientation OR zonation OR distribution OR spacing OR arrangement OR proximity OR boundary OR round\* ) AND TITLE-ABS-KEY ( shape OR size OR form ) AND TITLE-ABS-KEY ( idiomorphic OR xenomorphic OR circular OR elongated OR rounded OR acicular OR anastomosing OR rosette-like OR "aspect ratio" OR "edge quality" OR curvature OR angularity ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) ) | step1\_scopus-query\_result.csv |
| 2 | ( TITLE-ABS-KEY ( "mathematical analysis" OR "numerical analysis" OR descriptor\* OR factor\* OR "Image analysis" OR "computer vision" OR "Image processing" ) AND TITLE-ABS-KEY ( grain\* OR particle\* OR cell\* ) AND TITLE-ABS-KEY ( orientation OR zonation OR distribution OR spacing OR arrangement OR proximity OR boundary OR round\* OR texture ) AND TITLE-ABS-KEY ( shape OR size OR form ) AND TITLE-ABS-KEY ( idiomorphic OR xenomorphic OR circular OR elongated OR rounded OR acicular OR anastomosing OR rosette-like OR "aspect ratio" OR "edge quality" OR curvature OR angularity ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) ) | step2\_scopus-query\_result.csv |
| 3 | ( TITLE-ABS-KEY ( "mathematical analysis" OR "numerical analysis" OR descriptor\* OR factor\* OR "Image analysis" OR "computer vision" OR "Image processing" ) AND TITLE-ABS-KEY ( grain\* OR particle\* ) AND TITLE-ABS-KEY ( orientation OR zonation OR distribution OR spacing OR arrangement OR proximity OR boundary OR round\* OR texture OR "aspect ratio" OR curvature OR angularity ) AND TITLE-ABS-KEY ( shape OR size OR form ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) ) | step3\_scopus-query\_result.csv |
| 4 | ( TITLE-ABS-KEY ( "mathematical analysis" OR "numerical analysis" OR descriptor\* OR factor\* OR "Image analysis" OR "computer vision" OR "Image processing" ) AND TITLE-ABS-KEY ( grain\* OR particle\* ) AND TITLE-ABS-KEY ( orientation OR zonation OR distribution OR spacing OR arrangement OR proximity OR boundary OR round\* OR texture ) AND TITLE-ABS-KEY ( shape OR size OR form ) AND TITLE-ABS-KEY ( idiomorphic OR xenomorphic OR circular OR elongated OR rounded OR acicular OR anastomosing OR rosette-like OR "aspect ratio" OR "edge quality" OR curvature OR angularity ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) ) | step4\_scopus-query\_result.csv |

Table 2. Notes on images datasets.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset label** | **Notes** | **Image Number** | **CSV file** |
| Test1\_images | Ideal shapes: reference dataset. The orientation data isn’t valid. | 3900 | Test1\_data.csv |
| Test2\_images | Low Perlin noise intensity. The orientation data isn’t valid. | 3900 | Test2\_data.csv |
| Test3\_images | High Perlin noise intensity. The orientation data isn’t valid. | 3900 | Test3\_data.csv |
| Test4\_images | Low roundness intensity. The orientation data isn’t valid. | 3900 | Test4\_data.csv |
| Test5\_images | High roundness intensity. The orientation data isn’t valid. | 3900 | Test5\_data.csv |
| Test8\_images | Ideal shapes for PCA validation. The orientation data isn’t valid. | 325 | Test8\_data.csv |
| Orientation\_images | The orientation data is valid, only rectangle and ellipse with random axis length. Orientation is measured before and after a given rotation. | 600 | Orientation-test\_data.csv |
| Galena\_test | Real grain images, 4 characteristic grain for PCA validation. | 4 | Galena\_test\_data.csv |
| Galena\_training | Real grain images, used for the PCA training. | 580 | Galena\_training\_data.csv |