Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Supplementary materials for

Lvhan PAN, Guodao SUN, Baofeng CHANG, Wang XIA, Qi JIANG, Jingwei TANG, Ronghua LIANG, 2023. Visual interactive image clustering: a target-independent approach for configuration optimization in machine vision measurement. *Front Inform Technol Electron Eng*, 24(3):355-372. https://doi.org/10.1631/FITEE.2200547

1 Mathematical concepts in data sampling

For generalization, we illustrate concepts in data sampling by mathematical expressions. Table S1 shows the symbols and their description. A parameter set $\boldsymbol{P}$ can be expressed in matrix form in Eq. (S1), and a parameter element as an element in the set corresponds to a row in the matrix. Each column of $\boldsymbol{P}$ represents a parameter, and the default order of the columns is consistent with the order of the parameters in the algorithm. The number of rows of $\boldsymbol{P}$, that is, the size of $\boldsymbol{P}$, equals $\prod_{j=1}^{N}|\boldsymbol{v}^j|$.

$$\boldsymbol{P}=\begin{bmatrix}\boldsymbol{v}_{\min}^1 & \boldsymbol{v}_{\min}^2 & \cdots & \boldsymbol{v}_{\min}^N\\ \boldsymbol{v}_{\min}^1 & \boldsymbol{v}_{\min}^2 & \cdots & \boldsymbol{v}_{2}^N\\ \vdots & \vdots & & \vdots\\ \boldsymbol{v}_{\max}^1 & \boldsymbol{v}_{\max}^2 & \cdots & \boldsymbol{v}_{\max}^N\end{bmatrix}. \tag{S1}$$

Table S1 Mathematical symbols for concepts in data sampling

Symbol	Description
N	Number of parameters
$\boldsymbol{v}^j$	The vector representing the range of the j^{th} parameter in ascending order $(\boldsymbol{v}_{\min}^j, \boldsymbol{v}_2^j, \cdots, \boldsymbol{v}_{\max}^j)$
$\boldsymbol{v}_{\min}^j$	Minimum of the j^{th} parameter
$\boldsymbol{v}_{\max}^j$	Maximum of the j^{th} parameter
$\lvert\boldsymbol{v}^j\rvert$	Number of values of the j^{th} parameter
$\boldsymbol{P}$	A parameter set

2 Relationships among all concepts in data processing

We provide a figure to illustrate the relationship among all concepts in data processing (Fig. S1). In data sampling, each image generates a parameter set $\boldsymbol{P}$ that is represented as a matrix. The columns of the matrix are parameters selected in data sampling, and a row of the matrix is regarded as a parameter element for the algorithm. We first categorize each parameter element as success or failure according to whether the algorithm is completely executed or not (criterion_1). We then categorize success into the filtered parameter set or the dirty parameter set depending on whether the results satisfy the measurement accuracy requirement or not (criterion_2).

3 Edge filtering

In our experiments, we find that our tactics (method 1 and method 2) are effective in generating appropriate initial clusters for subsequent interactive clustering. As shown in Fig. S2, two lonely clusters

(the quadrangle and the pentagon) with a single glyph are generated after spectral clustering using all edges (Fig. S2a), providing an unsuitable initiation that leads to time-consuming interactive clustering. In contrast, our tactics can generate clusters that have basically the same number of glyphs (Fig. S2c). Our tactic can also reveal special patterns, which is beyond the capability of method 1 (for instance, Glyph 15 in Fig. S2c has more edges than Glyph 15 in Fig. S2b), indicating that the image is a non-essential element in the cluster.

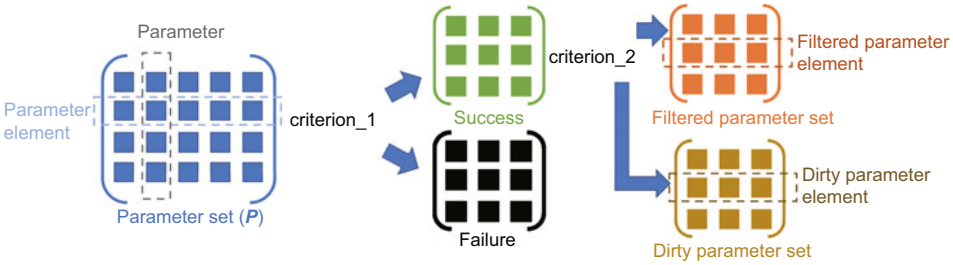

Fig. S1 Relationships among all concepts in data processing

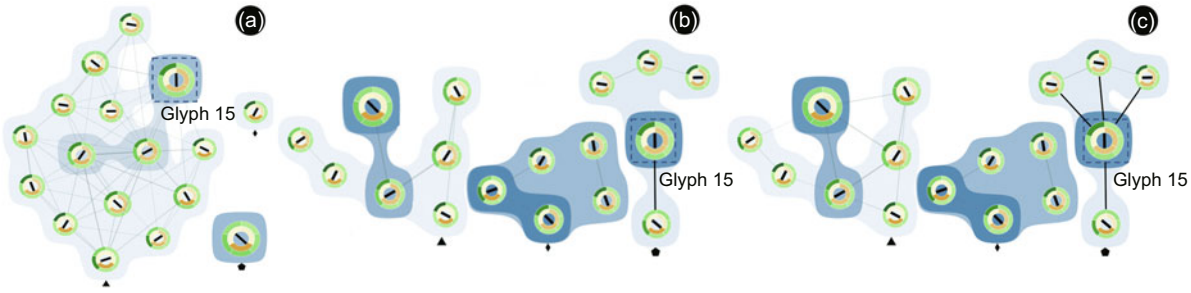

Fig. S2 Initial clustering results with different edge-filtering methods by spectral clustering: (a) all edges are reserved; (b) edges are filtered with method 1 ($p = 30\%$); (c) edges are filtered with method 1 and method 2 ($p = 30\%$) (Glyph 15 and its edges are highlighted in all results to show the difference in the three methods)