Supplementary Material

Ying Cao * Antoni B. Chan † Rynson W.H. Lau ‡

City University of Kong Kong

1 Panel Clustering Results

Additional panel clustering results are presented in Figure 1.

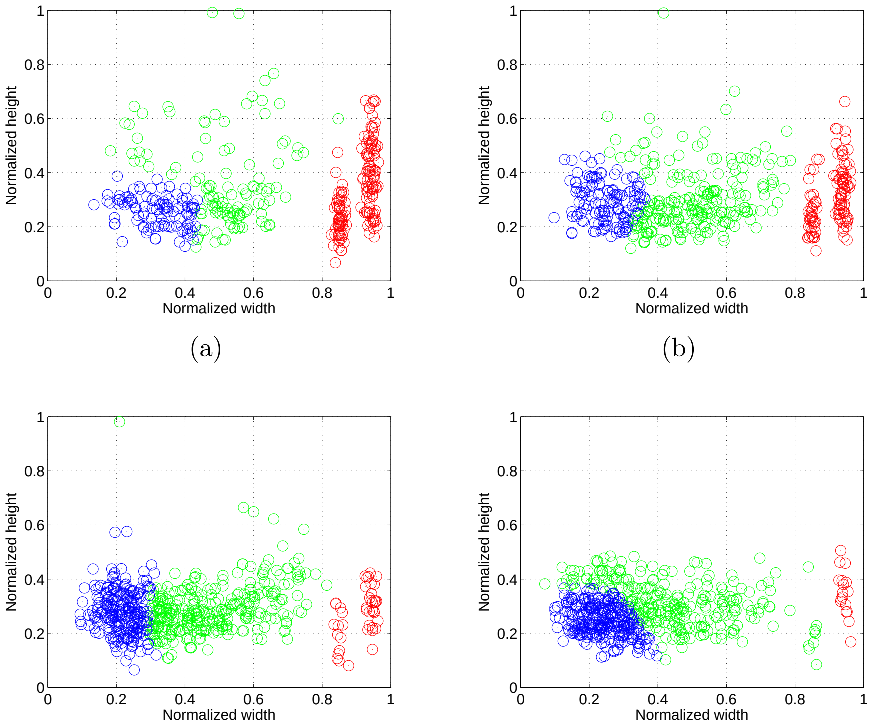

Figure 1: Panel clustering results. (a) clustering of the panels from 4-panel pages; (b) clustering of the panels from 5-panel pages; (c) clustering of the panels from 7-panel pages; (d) clustering of the panels from 8-panel pages. All the panels are extracted from "*Fairy Tail*".

2 Automatic Extraction of Spatial Division Instances

Let L_i be a spatial division instance (SDI) label, and $\{N_i, \boldsymbol{X}_i\}$ be the corresponding number of rows/columns and the splitting configuration. Given the coordinates of the panel vertices in

*yingcao4@student.cityu.edu.hk
†abchan@cityu.edu.hk
‡rynson.lau@cityu.edu.hk

a manga page, we aim to extract the SDIs, denoted by $\{\mathcal{L}_i, N_i, \boldsymbol{X}_i\}$. To do this, we recursively cluster the panels into a number of rows/columns according to their spatial positions, and obtain a SDI, denoted by a tuple $(\mathcal{L}_i, N_i, \boldsymbol{X}_i)$, at each recursive clustering level. At each level, N_i is the number of the clusters formed, and $\boldsymbol{X}_i$ is computed as a tuple of normalized heights/widths of bounding boxes of the clusters. The clustering is composed of two steps. First, we segment panels into several clusters according to the line-to-line distance between upper borders for row clustering (left border for column clustering) of the panels. To compute the line-to-line distance, we extend both borders to reach the edges of the page, and sample n points along these extended lines in evenly spaced intervals. The distance is computed as the average distance between the n pairs of corresponding points on both lines. We cluster the panels by finding groups of panels whose inter-panel distance are below a threshold. A panel that is far away from all other panels is treated as an isolated panel. Second, the clusters are refined by merging clusters whose convex hulls overlap. We then test whether isolated panels obtained in the previous step belong to any existing cluster. If the centroid of an isolated panel lies in the convex hull of a particular cluster, we assign this isolated panel to this cluster. Otherwise, the isolated panel is an independent panel spanning the entire row or column.

3 Derivation of Closed-Form Solutions to Two Steps of the Alternating Solver

Step 1: Calculating spatial transformation $T_i = \{s_i, \mathbf{t}_i\}$ for each image geometry.

T_i can be obtained by setting the derivatives of $\mathbf{E}_i$ w.r.t. $\{s_i, \mathbf{t}_i\}$ to zero,

$$\hat{s}_i = \frac{(4w_i+\lambda)S_{uv} - w_i S_u^T S_v}{(4w_i+\lambda)S_{uu} - w_i S_u^T S_u}, \quad \hat{\mathbf{t}}_i = \frac{w_i S_{uu} S_v - w_i S_u S_{uv}}{(4w_i+\lambda)S_{uu} - w_i S_u^T S_u}, \tag{1}$$

where $S_{uv} = \sum_{j=1}^{4} \mathbf{u}_{ij}^T \mathbf{v}_{\gamma(i,j)}$, $S_{uu} = \sum_{j=1}^{4} \|\mathbf{u}_{ij}\|^2$, $S_u = \sum_{j=1}^{4} \mathbf{u}_{ij}$, and $S_v = \sum_{j=1}^{4} \mathbf{v}_{\gamma(i,j)}$.

Step 2: Calculating layout mesh $\mathbf{V}$. The analytic solution for the optimization problem in Eq. 15 is obtained by minimizing the following Lagrange function,

$$L(\mathbf{V}, \boldsymbol{\lambda}) = \alpha\|\mathbf{A}\mathbf{V} - \mathbf{c}\|^2 + \beta\|\mathbf{V} - \mathbf{V}_0\|^2 + \sum_{i=1}^{l} \lambda_i(\boldsymbol{m}_i^T\mathbf{V} - b_i), \tag{2}$$

where $\lambda_i, i = 1, \ldots, l$ are non-negative Lagrange multipliers. where $\mathbf{A}$ is a $8n$-by-$2m$ matrix whose elements are,

$$A_{i,j} = \begin{cases} w_i^2 & \text{if } j\text{-th element in } \mathbf{V} \text{ is x- or y-coordinate} \\ & \text{of } k\text{-th vertex of } l\text{-th panel} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

and $k = \lceil (i \; mod \; 8)/2 \rceil$ and $l = \lceil i/8 \rceil$, and $\mathbf{c} \in \mathbb{R}^{8n}$ is a vector with elements c_i as $w_i^2 x$ or $w_i^2 y$, where x and y are the coordinates of k-th vertex of l-th image geometry.

Setting $\nabla_{\mathbf{V}} L(\mathbf{V}, \boldsymbol{\lambda}) = 0$ yields,

$$\mathbf{V} = (\alpha\mathbf{A}^T\mathbf{A} + \beta\mathbf{I})^{-1}(\alpha\mathbf{A}^T\mathbf{c} + \beta\mathbf{V}_0 - \frac{1}{2}\sum_{i=1}^{l} \lambda_i \boldsymbol{m}_i), \tag{4}$$

The Lagrange multipliers are calculated by substituting Eq. 4 into Lagrange function above,

$$L(\boldsymbol{\lambda}) = \rho - \sum_{i=1}^{l}(Z_i + b_i)\lambda_i + \sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j D_{ij}, \tag{5}$$

where $\mathbf{H} = (\alpha\mathbf{A}^T\mathbf{A} + \beta\mathbf{I})^{-1}$, $\mathbf{R} = \alpha\mathbf{A}^T\mathbf{c} + \beta\mathbf{V}_0$, and

$$\rho = \alpha\|\mathbf{AHR} - \mathbf{c}\|^2 + \beta\|\mathbf{HR} - \mathbf{V}_0\|^2 \tag{6}$$

$$Z_i = \boldsymbol{m}_i^T\mathbf{H}[\alpha\mathbf{A}^T(\mathbf{AHR} - \mathbf{c}) + \beta(\mathbf{HR} - \mathbf{V}_0) - \mathbf{R}] \tag{7}$$

$$D_{ij} = \tfrac{1}{4}(\alpha K_{ij} + \beta P_{ij} - 2H_{ij}), \tag{8}$$

with $\mathbf{K} = \mathbf{H}^T\mathbf{A}^T\mathbf{AH}$, $\mathbf{P} = \mathbf{H}^T\mathbf{H}$, $K_{ij} = \mathbf{K}(nmi, nmj), P_{ij} = \mathbf{P}(nmi, nmj)$, and $H_{ij} = \mathbf{H}(nmi, nmj)$, where nmi and nmj are the indexes of non-zero elements in $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$, respectively. The optimal $\boldsymbol{\lambda}$ is obtained by setting $\nabla_{\boldsymbol{\lambda}}L(\boldsymbol{\lambda}) = 0$,

$$\boldsymbol{\lambda} = \mathbf{F}^{-1}\mathbf{C} \tag{9}$$

where $\mathbf{F}$ is a l-by-l matrix with entries $\mathbf{F}(i,j) = D_{ij}$, and $\mathbf{C} = (Z_1 + b_1, \cdots, Z_l + b_l)^T$. Given the optimal $\boldsymbol{\lambda}$, the optimal $\mathbf{V}$ are calculated by substituting into Eq. 4.

4 Evaluation of Inter-panel Semantics Annotation

Figure 2 illustrates the importance of the inter-panel semantics annotation.

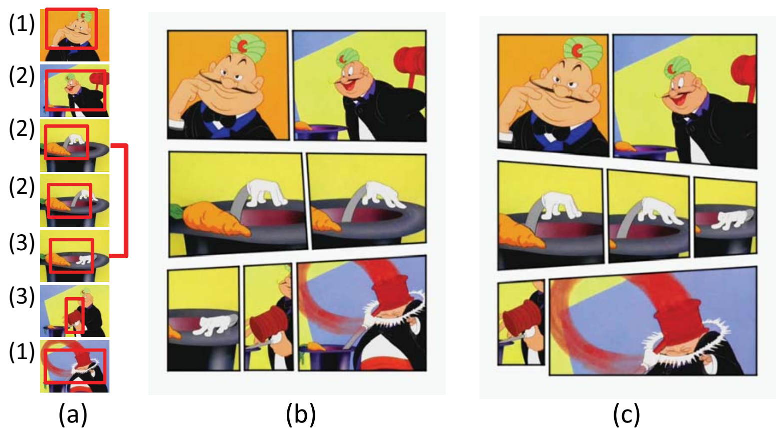

Figure 2: With a sequence of artworks (a) ("*Case of the Missing Hare*" (1942) in public domain), resulting layouts without and with inter-panel annotation are shown in (b) and (c), respectively. Note that, compared to (b) that has a visual break at the right boundary of the second row, (c) is a more reasonable layout, where semantically-related panels (three consecutive panels connected by red line in (a)) can be easily interpreted as an entity.