MANGA-SPECIFIC FEATURES AND LATENT STYLE MODEL FOR MANGA STYLE ANALYSIS

Wei-Ta Chu and Wei-Chung Cheng

National Chung Cheng University, Chiayi, Taiwan

ABSTRACT

A latent style model describing manga styles based on the proposed manga-specific features is constructed to facilitate novel style-based applications. Two manga-specific features, i.e., screentone features showing texture and shade, and panel features showing panel arrangement, are firstly proposed to describe manga pages. Based on the latent Dirichlet allocation technique, we discover latent style elements embedded in manga documents, which are described by visual words derived from manga-specific features. Distributions of style elements are then used to measure similarity between manga documents, and facilitate the development of various style-based applications. Experimental results show that the features and models especially designed for describing manga styles yield promising performance and could bring many potential extensions.

Index Terms— Manga-specific features, latent style model, style-based applications

1. INTRODUCTION

Many mangas (Japanese comics) are published every year, which not only build a big entertainment market but also serve as an important medium to convey knowledge and culture. Tremendous amounts of manga pages, however, give rise to significant challenges of efficient access. Although many techniques and features have been proposed for content-based image retrieval or visual search, fewer manga-specific studies have been conducted to discover the space formed by *style elements*, which could be characterized by drawing and layout arrangement.

In this work we propose novel access scenarios based on *manga styles*. We thought that manga styles can be described in a style space, where is conceptually constituted by distributions of style elements from different modalities, such as line drawing, screentone, and panel arrangement. Figure 5 shows sample manga pages produced by different artists, where we can easily perceive style difference. We design features especially designed to describe manga pages, and then discover latent style elements based on the latent Dirichlet allocation technique. With the latent style model, the style distribution of given manga pages can be unveiled and serves as a new representation, which facilitates innovative ways of efficient access like style-based retrieval.

Contributions of this paper are summarized as follows.

- Manga-specific features: Other than conventional visual features, we propose features specific to manga pages, including bag of screentone primitive, and panel arrangement feature. Associated with line features that were recently proposed, manga pages are jointly described from three different perspectives, and style elements can be well discovered.
- Latent style model: With manga-specific features, we describe manga pages as documents, and adopt the latent Dirichlet allocation technique to discover latent style elements. Such computational model provides a systematic approach to conduct manga style analysis.
- Applications: We propose innovative ways to efficient access manga pages. Given a set of manga pages, we are able to retrieve manga pages produced by the same artist, or produced by the artists of the same art movement. The proposed method is also demonstrated to be able to do finer analysis such as style evolution of the same artist.

2. RELATED WORKS

Manga analysis. Current manga analysis works can be roughly grouped into three categories: character detection, panel segmentation, and feature design. We briefly review few of them due to space limitation. Sun et al. [1] extracted local features from each panel, and conducted feature matching to verify whether two panels contain the same character. Takayama et al. [2] extracted skin color and jaw contour to identify face of human-like characters. They also extracted hair color and hair quantity, associated with skin color, as features for face recognition. Cao et al. [3] designed a data-driven approach to learn artist's panel arrangement style. As panel segmentation is an important basic step for style elements extraction, recently the same research team proposed a robust panel extraction method for manga [4]. Although there have been tremendous amounts of visual features proposed for natural images, few features have been proposed especially for mangas. Chu and Chao [5] proposed description for line segments detected from character's face and conducted statistical analysis to show its effectiveness in image classification.

Style analysis. In terms of visual analysis, style refers to the distinctive manner with which entities like artworks, architecture, and clothing can be grouped or recognized. Styles usually can be easily perceived, but are hard to be described in a computational way. In the visual analysis community, currently there are works related to calligraphy [6], paintings [7], and clothing [8] [9], and few studies have been conducted for mangas.

3. MANGA FEATURES

Although line-based features have been shown to be descriptive for manga image classification [5], detecting character's face is not a trivial task, and thus extracting line-based features from faces is not reliable in all cases. In this work, we further propose two features to be important complementary information.

3.1. Screentone Features

Screentone is a technique to apply texture or shade to objects or scene in mangas. It is often used to present shadow of charac-

Fig. 1: Left and middle: samples of different screentones applied to the same character; right: Sample patches of six screentone primitives.

ters, fold or lap of clothes, and texture of character's hair. Basically, screentone is constituted by densely placed low-intensity pixels. With different setting on density and pixel's intensity, various screentones can be constructed. Figure 1 shows two sample images with different screentone settings to present distinct styles.

To detect screentone in a panel, we first binarize each pixel by checking whether its intensity value is larger than a predefined threshold, say 200 in this work. The set of pixels with lower intensity values tend to present screentone. We then subsequently employ the erosion and dilation operations to filter out small fragments, and view the final set of pixels forming regions containing screentone.

Two screentone features are proposed: the ratio of screentone area to the whole panel area (s_1), and the bag of screentone primitives (s_2). For s_1, we calculate the ratio of screentone area to the whole panel area. To extract s_2, we divide the panel into nonoverlapping 40×40 patches. If more than 80% of pixels in this patch are with low intensity values, this patch is viewed as a screentone patch. The Gabor wavelet transform with four orientations and three scales is applied to each screentone patch, and average and standard deviation of transform coefficients in each frequency band are then calculated and concatenated to be the patch's texture feature vector. Inspired by the bag of visual words model, feature vectors of screentone patches in the training dataset are collected, and then the affinity propagation algorithm [10] is adopted to group feature vectors into clusters. The rightmost subfigure of Figure 1 shows several sample clusters. Feature vectors corresponding to the same cluster belong to the same screentone primitive. With the clustering configuration, we can quantize each screentone patch into one of the primitives. The feature s_2 of a manga page is then represented as the bag of screentone primitives derived from screentone patches in this page.

3.2. Panel Features

How several images are placed inside a page also presents artistic styles. Generally, more important images are allocated more space to present, and several images arranged in a grid manner is commonly seen. However, to enhance visual attractiveness or manifest different storytelling, artists usually design borderless/unclosed/irregular panels or multiple panels with joint elements.

Here we focus on panel feature design and adopt the method in [4] to extract panels. After panel extraction, we bound each panel with a minimum bounding box, and eleven panel features are extracted: 1) p_1: average panel height (derived from bounding boxes); 2) p_2: average panel width; 3) p_3: standard deviation of p_1; 4) p_4: standard deviation of p_2; 5) p_5: the ratio of total panel area to the whole page; 6) p_6: average panel area; 7) p_7: standard deviation of p_6; 8) p_8: average slope of vertical panel boundaries; 9) p_9: average slope of horizontal panel boundaries; 10) p_{10}: standard deviation of p_8; 11) p_{11}: standard deviation of p_9.

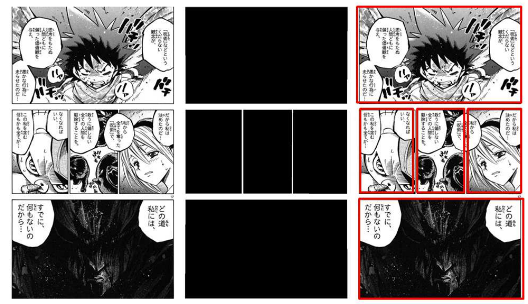


Fig. 2: Sample result of panel extraction. Left: original page; middle: the extracted panels (in black); right: result of panel segmentation (with red border).

Overall, the proposed screentone features (s_1 and s_2) and panel features (p_1–p_{11}) are concatenated with the line features proposed in [5] to form a feature vector describing a manga page. Note that line features proposed in [5] are used to describe a character's face. In this work, we detect the largest and frontal face in a manga page to extract line features.

4. STYLE MODEL

Most prior manga-related works focus on feature design, e.g., [5] or object detection, e.g., [1], and put few efforts on computational style analysis. Although Cao et al. [3] designed a generative model to describe panel partitioning, a computational model to describe styles from various aspects including panel partitioning and line drawing is missing. In this work, we develop a style model based on Latent Dirichlet Allocation (LDA) to discover style elements.

Each manga page is described by a feature vector described in Sec. 3. We extract feature vectors from the evaluation dataset (details to be described in Sec. 5), adopt principal component analysis to reduce dimensionality into 20, and employ the K-means clustering algorithm to construct the visual vocabulary. Each manga page, therefore, can be represented as a visual word through quantizing the corresponding feature vector.

Considering the analogy between manga pages and text document, we can view several manga pages of the same artist as a document, view the visual word representing a manga page as a word, and view the discovered latent topics as style elements. Particularly, a document d_i is represented as a bag of N_i visual words, denoted by $d_i = \{v_1, v_2, ..., v_{N_i}\}$, where v_j is the jth visual word. Documents are assumed to be characterized by K style elements. The latent style model assumes the following generative process for a corpus consisting of M documents each of lengths (number of words) N_i.

1. Choose $\theta_i \sim Dir(\alpha)$, where $i = 1, ..., M$, and α is the parameter of the Dirichlet prior on the per-document style (topic) distributions. θ_i is the style distribution for the document d_i.
2. Choose $\phi_k \sim Dir(\beta)$, where $k = 1, ..., K$, and β is the parameter of the Dirichlet prior on the per-style word distributions. ϕ_k is the word distribution for the kth style.
3. For each visual word $v_{i,j}$, $i = 1, ..., M$, $j = 1, ..., N_i$:
   (a) Choose a style $z_{i,j} \sim Multinomial(\theta_i)$. $z_{i,j}$ is the topic assignment for the jth visual word in the ith document.

   (b) Choose a visual word $v_{i,j} \sim Multinomial(\phi_{z_{i,j}})$.

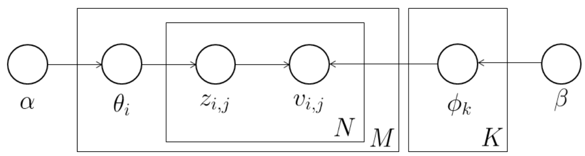


Fig. 3: Graphical model representation of the style model.

Mathematically the generative process corresponds to the joint distribution of the hidden variables (ϕ, θ, z) and observed variables v:

$$p(\phi, \theta, z, v) = \prod_{k=1}^{K} p(\phi_k) \prod_{i=1}^{M} p(\theta_i) \Big(\prod_{j=1}^{N} p(z_{i,j}|\theta_i) p(v_{i,j}|\phi_k, z_{i,j}) \Big). \tag{1}$$

Given a new document with the observed visual words $\{v_j\}$, with the learnt model we can efficiently estimate the marginal distribution $p(v|\alpha, \beta)$ by the Gibbs sampling algorithm [11]. Style probabilities of a document can be estimated, which enable us to represent a document as a distribution of style elements.

Note that the definition of a document in the LDA framework is changeable and application-dependent. For example, to characterize an artist, we can view a set of manga pages produced by an artist as a document. To characterize an artist's artworks produced in a specific period, we can view a set of manga pages in the same period as a document. Details of document definition for different applications will be described in the following section.

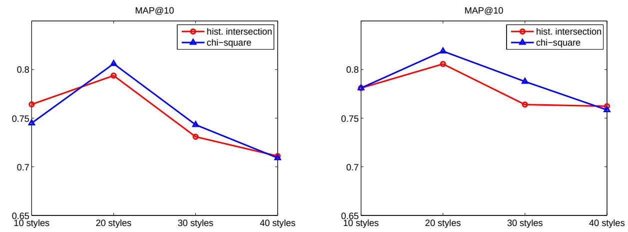
5. APPLICATIONS

Based on the latent style model, style-based access can be achieved, which is beyond conventional manga studies working at feature level. To develop these applications and conduct evaluation, we collect 100 manga pages from eight different artists and obtain an evaluation dataset consisting of totally 800 manga pages.

5.1. Style-Based Artist Retrieval

By viewing manga pages produced by the same artist as documents, we can construct a latent style model to describe artist's styles. Practically, from the manga page collection of the same artist, we randomly group them into subsets each consists of 20 manga pages, and each subset is viewed as a document. Note that an episode of an artist's works usually contains a dozen or more (less than 40) manga pages. An artist's style can be more appropriately described when we consider multiple manga pages coming from an episode, and thus we view 20 manga pages as a document. Documents from all artists thus form a database. With $N = 100$ (100 visual words), $M = 40$ (40 documents), a latent style model $\mathcal{M}_a = \{\alpha_a, \beta_a\}$ can be learnt.

We propose an application of style-based artist retrieval. Given an artist's manga document, we would like to retrieve other documents produced by the same artist. In the evaluation, we randomly select a document as the query from the database. Based on the latent style model, each document (query document and database documents) can be represented as a distribution of style elements. The similarity between the query document and a database document is then measured by histogram intersection or chi-square distance. Most similar database documents are retrieved and ranked first in the returned list.



(a) (b)

Fig. 4: Left: MAP@10 of style-based retrieval for manga pages produced by the same artist. Right: MAP@10 of style-based retrieval for manga pages produced by the same art movement.

Figure 4a shows mean average precisions (MAP@10) of retrieval when we vary numbers of style elements (K), based on histogram intersection and chi-square distance, respectively. Two observations can be made. First, the best MAPs can be obtained when the number of style elements is set as 20 for both distance metric. Second, measuring similarity by chi-square distance yields slightly better performance, i.e., MAP@10 = 0.806.

The top row of Figure 5 shows sample manga pages produced by three different artists, and the bottom row show the corresponding distributions of 20 style elements. We can clearly see that different artists have different habits to produce their artworks, which can be effectively described by the proposed latent style model.

To further verify the effectiveness of describing manga documents by style elements, we compare the proposed style-based description with a baseline description. As mentioned in the beginning of Sec. 4, each manga page is described by the concatenation of line features, screentone features, and panel features. The dimensionality of this feature vector is then reduced to 20 by PCA. For a manga document consisting of twenty manga pages, we thus have twenty 20-dimensional feature vectors $\{\boldsymbol{f}_1, ..., \boldsymbol{f}_{20}\}$, where $\boldsymbol{f}_i = (f_1, ..., f_{20})$ denotes the ith page in the document. To describe the manga document, dimension-wise average is calculated over the twenty feature vectors to form a 20-dimensional description, i.e., $\bar{\boldsymbol{f}} = (\bar{f}_1, ..., \bar{f}_{20})$, where $\bar{f}_j = \frac{1}{20} \sum_{j=1}^{20} f_j$. This description is viewed as the baseline description for a manga document. Same as the retrieval process mentioned above, given an artist's manga document (represented by the baseline description), we would like to retrieve other documents produced by the same artist. Overall, the MAP@10 value obtained based on the baseline description is 0.238. Comparing with the performance shown in Figure 4a, we clearly see the superiority of the proposed style-based description.

5.2. Style-Based Art Movement Retrieval

We can extend the artist-based style description to a broader range. An art movement is a tendency or style in art with a specific common philosophy or goal, followed by a group of artists. There has been a solid manga mentoring system in Japan. Many famous artists apprenticed in a master's studio to learn manga production when they were young. Styles of mangas produced by artists coming from the same studio are thus correlated, and we can say these artists form an art movement. In our evaluation dataset, 8 artists can be divided into three different movements. Relationships between these artists are shown in Figure 6.

Following the experimental settings described in the previous section, we can view 20 manga pages produced by the artists of the same art movement as a document. A latent style model $\mathcal{M}_m =$

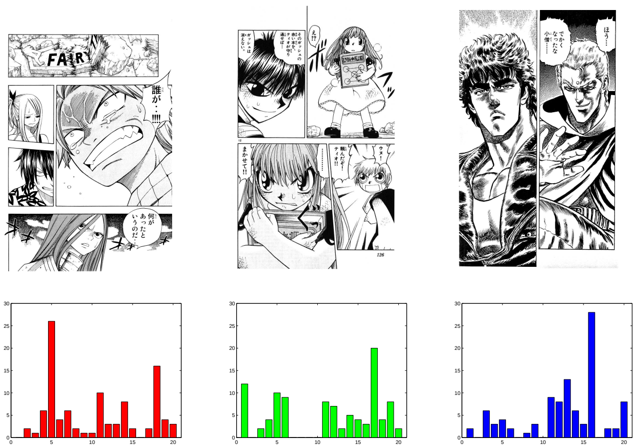


Fig. 5: Top: sample manga pages from three different artists. Bottom: style distributions corresponding to these artists.

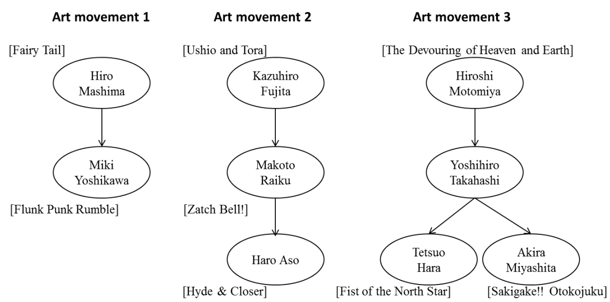


Fig. 6: Three art movements in our dataset, including 2, 3, and 3 artists, respectively. Text in ovals and brackets are names of artists and names of representative manga production, respectively. The arrows heading from the master to the apprentice. Characters in Yoshihiro Takahashi's production are not human-like, and thus are not considered in our work.

$\{\alpha_m, \beta_m\}$ can thus be built. Given a query, we would like to retrieve manga documents produced by artists of the same movement. Figure 4b shows mean average precisions (MAP@10) of retrieval when we vary numbers of style elements, based on histogram intersection and chi-square distance, respectively. Trends similar to that from artist retrieval can be observed.

Similarly to verify the effectiveness of the proposed style-based description, we construct the baseline description for each document based on the same process mentioned in Sec. 5.1. The MAP@10 value obtained by the baseline approach is 0.506, which is far below the performance obtained based on the style-based description.

Success of applications mentioned in Sec. 5.1 and Sec. 5.2 enables that, when a user likes the style embedded in manga pages just read, he/she can retrieve similar artworks without knowing artist's name or the manga title. This could be a novel and very user-friendly scenario to access mangas.

5.3. Style-Based Artwork Period Retrieval

In Figure 5 we can easily perceive style difference between different artists, which is confirmed by the promising performance mentioned in Sec. 5.1. Here we also want to evaluate whether the proposed style model works well at a finer level, i.e., style evolution. Some popular mangas had been published for more than ten years. For

Fig. 7: Left to right: sample manga pages from *JoJo's Bizarre Adventure* Part 1 (1987–1988), Part 3 (1989–1992), and Part 8 (2011–).

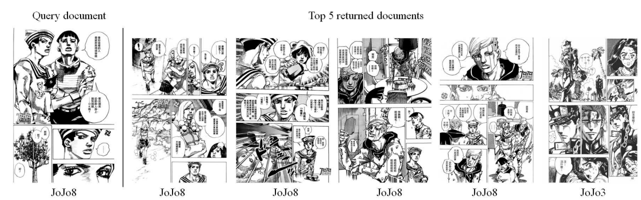


Fig. 8: Sample results of the query and top returned documents. Note that only the fifth returned document (JoJo3) is from the period different from the query (JoJo8).

example, since the first volume published in 1987, *JoJo's Bizarre Adventure* by Hirohiko Araki has been published for three decades. There are totally eight parts consisting of more than 110 volumes. In different parts, styles of this manga evolve gradually. Figure 7 shows samples produced in three different periods (Part 1, Part 3, and Part 8, denoted by JoJo1, JoJo3, and JoJo8 in the following).

We advocate that *artwork period retrieval* can also be achieved based on the latent style model. Given a query manga document from JoJo8, for example, we would like to retrieve documents from the database that are also from JoJo8. From the database we randomly select query documents, and examine whether the top-ranked retrieved documents come from the same period as the query. Figure 8 is an example showing a query and the top returned documents. Overall, the value of MAP@10 is 0.73, which shows the proposed model is feasible to be adopted in finer style analysis.

6. CONCLUSION

We have presented screentone features and panel features especially designed for describing manga styles. Based LDA, implicit style elements are discovered in a probabilistic framework. With the latent style elements, novel applications including artist retrieval, art movement retrieval, and artwork period analysis can be achieved at the style level rather than the feature level. This work is believed to be one of the earliest attempts to systematically study manga styles, and may inspire many potential style-based analysis. In the future, larger-scale experiments will be performed, and more innovative ways to access or manage manga documents based on styles will be proposed.

Acknowledgement. The work was partially supported by the Ministry of Science and Technology in Taiwan under the grant MOST103-2221-E-194-027-MY3 and MOST104-2221-E-194-014.

7. REFERENCES

[1] Weihan Sun, Jean-Christophe Burie, Jean-Marc Ogier, and Koichi Kise, "Specific comic character detection using local feature matching," in *Proceedings of International Conference on Document Analysis and Recognition*, 2013, pp. 275–279.

[2] Kohei Takayama, Henry Johan, and Tomoyuki Nishita, "Face detection and face recognition of cartoon characters using feature extraction," in *Proceedings of the IIEEJ Image Electronics and Visual Computing Workshop*, 2012.

[3] Ying Cao, Antoni B. Chan, and Rynson W.H. Lau, "Automatic stylistic manga layout," *ACM Transactions on Graphics*, vol. 31, no. 6, 2012.

[4] Xufang Pang, Ying Cao, Rynson Lau, and Antoni Chan, "A robust panel extraction method for manga," in *Proceedings of ACM International Conference on Multimedia*, 2014, pp. 1125–1128.

[5] Wei-Ta Chu and Ying-Chieh Chao, "Line-based drawing style description for manga classification," in *Proceedings of ACM International Conference on Multimedia*, 2014, pp. 781–784.

[6] Yueting Zhuang, Weiming Lu, and Jiangqin Wu, "Latent style model: Discovering writing styles for calligraphy works," *Journal of Visual Communications and Image Representation*, vol. 20, pp. 84–96, 2009.

[7] Jiachuan Sheng and Jianmin Jiang, "Recognition of chinese artists via windowed and entropy balanced fusion in classification of their authored ink and wash paintings (iwps)," *Pattern Recognition*, vol. 47, pp. 612–622, 2014.

[8] Shintami C. Hidayati, Wen-Huang Cheng, and Kai-Lung Hua, "Clothing genre classification by exploiting the style elements," in *Proceedings of ACM Multimedia*, 2012, pp. 1137–1140.

[9] Wei Di, Catherine Wah, Anurag Bhardwaj, Robinson Piramuthu, and Neel Sundaresan, "Style finder: Fine-grained clothing style recognition and retrieval," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 8–13.

[10] Brendan J. Frey and Delbert Dueck, "Clustering by passing messages between data point," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.

[11] Mark Steyver and Tom Griffiths, "Probabilistic topic models," in *Latent Semantic Analysis: A Road to Meaning*. 2006, Lawrence Erlbaum.