# Automatic Generation of Visual-Textual Presentation Layout

XUYONG YANG, University of Science and Technology of China
TAO MEI, Microsoft Research Asia
YING-QING XU, Tsinghua University
YONG RUI, Microsoft Research Asia
SHIPENG LI, Microsoft Research Asia

Visual-textual presentation layout (e.g., digital magazine cover, poster, Power Point slides, and any other rich media), which combines beautiful image and overlaid readable texts, can result in an eye candy touch to attract users' attention. The designing of visual-textual presentation layout is therefore becoming ubiquitous in both commercially printed publications and online digital magazines. However, handcrafting aesthetically compelling layouts still remains challenging for many small businesses and amateur users. This paper presents a system to automatically generate visual-textual presentation layouts by investigating a set of aesthetic design principles, through which an average user can easily create visually appealing layouts. The system is attributed with a set of topic-dependent layout templates, a computational framework integrating high-level aesthetic principles (in a top-down manner) and low-level image features (in a bottom-up manner). The layout templates, designed with prior knowledge from domain experts, define spatial layouts, semantic colors, harmonic color models, and font emotion and size constraints. We formulate the typography as an energy optimization problem by minimizing the cost of text intrusion, the utility of visual space, and the mismatch of information importance in perception and semantics, constrained by the automatically selected template, and further preserves color harmonization. We demonstrate that our designs achieve the best reading experience compared with the re-implementation of parts of existing state-of-the-art designs through a series of user studies.

CCS Concepts: •**Computer systems organization** → **Embedded systems;** *Redundancy;* Robotics; •**Networks** → Network reliability;

Additional Key Words and Phrases: Visual-textual presentation layout, design templates, typography, color harmonization, rich media presentation

**ACM Reference Format:**
Xuyong Yang, Tao Mei, Ying-Qing Xu, Yong Rui, Shipeng Li, 2015. Automatic Generation of Visual-Textual Presentation Layout. *ACM Trans. Multimedia Comput. Commun. Appl.* 9, 4, Article 39 (August 2015), 23 pages.
DOI: 0000001.0000001

## 1. INTRODUCTION

With ubiquitous access to and usage of social media, people are now creating and sharing more and more rich-media content on the Web than ever before, either for experience sharing or product promotion. One of the fundamental challenges in publishing rich media content is how to design a visually compelling layout consisting of heterogeneous media elements (e.g., images and textual description). The visual-textual presentation layout, which combines beautiful image and overlaid readable texts, can

---

This work described here has been completed in Microsoft Research Asia.
Author's addresses: X. Yang, University of Science and Technology of China, Hefei, China (email: yxy8023@live.cn); T. Mei, Y. Rui, and S. Li, Microsoft Research Asia, Beijing, China (email: {tmei, yongrui, spli}@microsoft.com); Y.-Q. Xu, Tsinghua University, Beijing, China (email: yqxu@tsinghua.edu.cn).

DOI: 0000001.0000001


(a)visual-textual presentation layout (b) real magazine cover

Fig. 1. Examples of visual-textual presentation layout: (a) the layout automatically generated from our proposed approach, and (b) the layout of a real magazine cover. We aim to automatically create a professional looking layout from the given image and text in this work. Please note that the layout of the real magazine cover is approved to use by the publisher.

result in an eye candy touch to attract user's attention. The designing of visual-textual presentation layouts is therefore becoming ubiquitous, ranging from existing commercially printed publications, to online digital magazines, to personal media posts. However, crafting aesthetically compelling layouts still remains an impediment for many small businesses and amateur users. In this paper we look at a way to create visually compelling visual-textual presentation layouts to deal with the above challenge, such as with the example shown in Figure 1(a).

The key concepts for properly designing a good visual-textual presentation layout have been studied in [Jahanian et al. 2013], where Jahanian *et.al.* introduce the elements, principles and aesthetics in design. The basic elements of design refer to six basic visual elements perceived by people, including the compositions of color, line, shape, tone, texture, and volume. At a higher level, the designing principles are the main ways to work with and arrange the elements, involving specific rules such as visual balance in symmetric or asymmetric visual structure, typography with regular repetition and alignment points, and color with pleasing harmonization, among others. The aesthetics of design, however, measure not only the form of element combinations, but also the emotion conveyed by the visual-textual layout. The magazine cover is one of the most popular media formats which fully embodies these key concepts in the design of the layout. In the real magazine cover shown in Figure 1(b), the image contains color harmonization with the hue of text in Type "V" according to the dominant hue color [Cheng et al. 2011]. The different sizes and positions of texts build a visual path movement to lead the reader's eye tracks along with the visual importance in perception. The repetition form of texts and the balanced cover lines layout on the non-salient regions depict a well-organized structure, while the title on the left bottom of the image reflects the asymmetrical balance and controllable variation. The most obvious aesthetics in color theory here is the selection of warm color which expresses the enthusiastic and attractive emotion in this magazine cover.

Many engineers and researchers have been inspired to generate visual-textual presentation layout automatically. In industry, with the shifting of media consumption from traditional PCs to various mobile devices which are attributed with limited

screen size and touch screen, the tile- or flipping-based reading experience has attracted significant attention. For example, Flipboard organizes social media in an appealing magazine-style layout [1]. Although no detailed information about the implementation of Flipboard is available, it is likely that after aligning the uniformed image with the entire screen, the left space will be filled with a white block. To keep the information on the screen maximally presented, a piece of abstracted text is overlaid on the image with a translucent mask on the fixed position (usually at the bottom) with fixed scale and color. Such a layout scheme has been widely applied on many websites. However, by neglecting the consideration of image content, the overlaid text may occlude the important parts of the image, leading to a breaking of basic aesthetic principles (e.g., little intrusion), and thus degrading users' reading experience.

There have also been quite a few attempts at automatic generation of visual-textual layout in academia. By analyzing design principles, researchers proposed a lot of computational aesthetic rules such as visual balance, equilibrium, unity, density, and proportion [Bauerly and Liu 2006]. By applying these computational models, the generation of visual-textual layouts is usually treated as a graphical constraint and energy optimization problem. For example, Yin *et al.* propose an optimization-based approach to computing the position, font size and color for overlaying a piece of text when automatically generating a visual-textual social media snippet for mobile browsing [Yin et al. 2013], a similar approach is applied in textual ads on visual image [Mei et al. 2012]. Kuhna *et al.* recommend the position of text automatically by minimizing the important image pixels covered by text blocks [Kuhna et al. 2012]. Some researchers try to involve people's prior knowledge explicitly to supervise the layout quality, including the automated layout generator for webpages by leveraging layout knowledge learned from massive layout [Kröner 1999], the pre-defined spatial constraints of textual elements [Jahanian et al. 2013], and a set of templates to align the layout generation [Jacobs et al. 2003].

Previous work has predominantly focused separately on either using templates as design examples or optimization-based schemes without prior knowledge. As a result, it only achieves limited performance when the resource image content is diverse across different topics (e.g., "fashion," "travel," "entertainment," "food & drink," etc.). It is known that both domain-specific knowledge and content features play important roles in layout generation [Arnheim 1954]. A good design of visual-textual presentation layout should take a set of predefined templates as *prior* (dependent to topic), domain-specific aesthetic principles, and computational content features into account.

Motivated by the above observations, we present a system to automatically generate visual-textual presentation layouts by investigating a set of aesthetic designing principles, through which an average user can easily create visually appealing layouts. The system is attributed with a set of topic-dependent layout templates (as knowledge *prior*), a computational framework integrating high-level aesthetic designing principles (in a top-down manner) and low-level content features (in a bottom-up manner). The layout templates, designed with prior knowledge from domain experts, define spatial layouts, semantic colors, harmonic color models, font emotion and size constraints. The framework formulates the typography as an energy optimization problem by minimizing the cost of text intrusion, the waste of spare visual space, and the mismatch of information importance in perception and semantics, constrained by the automatically selected template, and further preserves color harmonization.

Figure $1(a)$ shows an automatically generated layout by our proposed approach which reflects many design principles in a real magazine cover as shown in Figure $1(b)$. Among various types of visual-textual presentation layout, magazine cover em-

[1] https://flipboard.com/

bodies the most comprehensive design concepts. The aesthetic principles maintain universal. However, the templates including spatial layout and font property are different in various types of publication. For example, in a poster, to make the text blocks fully distinguish and attractive, the range of font size tends to be wider and the repository of font family is more comprehensive. We provide a computational framework with practical components to combine the high-level aesthetic principles (in a top-down manner) and low-level visual features (in a bottom-up manner). To validate the effectiveness of our proposed framework, we summarize a set of templates from the most complicated magazine covers by designers and apply it into our framework to generate the corresponding visual-textual layout. The templates can be revised and easily extended for other publications in our proposed framework.

Although we only show the layout application on a magazine cover, the proposed visual-textual presentation layout can be easily extended to many publishing applications, such as posters, Power Point presentations, and other self-published rich media.

In summary, this paper makes the following contributions:

— We propose a set of topic-dependent templates summarized by dominant experts with a set of aesthetic design principles. The templates guide the design in terms of spatial layout and color harmonization, ensuring satisfying layout performance.
— We design a computational framework to integrate all the key elements of layout design, including the layout templates as *prior*, the summarized high-level aesthetic principles (in a top-down manner) and low-level image features (in a bottom-up manner). We formulate the problem of typography as a template constrained energy minimization problem.

The remaining sections of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the template design. Section 4 presents our proposed system for automatic generation of visual-textual layout, followed by the evaluations in Section 5 and conclusions in Section 6.

## 2. RELATED WORK

Designing a visual-textual layout is a multidisciplinary research topic which can be divided into three stages: 1) compositing the image to match the size standard of the target media, 2) typesetting the texts on the overlaid image, and 3) coloring the textual elements. The following parts involve related work in each stage, and brief comparisons between our proposed approach and previous work.

### 2.1. Image composition

To address the mismatch between the original image and the diverse media size standard, the image needs to be resized to comply with the target layout and preserve important regions such as faces and salient objects. The most straightforward image composition operation is cropping. The image cropping in [Kuhna et al. 2012] and [Yin et al. 2013] is based on maximizing the corresponding area of the self-defined importance map. Another approach is to optimize photo composition by scaling and cropping the image under some basic aesthetic guidelines such as rule of third, diagonal dominance, and visual balance [Liu et al. 2010]. Such composition commonly highlights the salient object and discards the redundant parts. In some cases, the salient objects lay a little far apart in the image. Direct cropping or scaling results in missing parts of salient objects. To address such a problem, many image retargeting techniques have been proposed, to resize an image without losing the important objects within it [Avidan and Shamir 2007] for content-aware image resizing. Segmentation-based approach, patch-based methods and warping-based methods have also been proposed to address the image composition problem, however such approaches usually create

artifacts and distortions in important areas of one image. To prevent human's basic aesthetic from the unnatural effect in image resizing, we just apply image cropping and scaling to keep essence of the image.

### 2.2. Automated layout

How to determine the sizes and positions of textual and graphical elements in the information presentation media such as in magazines and newspapers still remains a challenge, since it significantly affects information perception and aesthetic experience. The primary layout for sequential presentations is applied for some commercial presentation and word-processing system. A survey introduces the automated layout techniques for information presentation in [Lok and Feiner 2001]. In constraint-based automated layout system, the most representative and effective one is based on adaptive grid-based document layouts [Jacobs et al. 2003]. Many popular auto editing tools such as the user control "FlowDocumen" in Microsoft Windows Presentation Foundation support grid-based layout templates and solve template constraints accordingly. The layout constraints are also introduced in [Jahanian et al. 2013], where the title element is constrained at the top of page and the font type is increased as long as the title element fits the page width. The cover lines are constrained by the cover templates and the salient objects. Learning-based automated layout generation algorithm has been explored with the help of large database. Especially Zhou divides the learning space for presentation generation into information learning space, visual learning space and rule learning space.In [Zhou and Ma 1999], Zhou introduces a fully automated graph generation system by leveraging a large database of presenation. In [Yin et al. 2013], the interaction of text and graphics is treated as an optimizing problem, by minimizing the energy, including text position, size, and color. The research reveals the effectiveness to address the automated layout problem as an optimizing problem with aesthetic rules and visual perception principles. It's the way we are going ahead in this paper, moreover, we introduce the template-based constraint to guarantee the optimized performance.

### 2.3. Color design model

Color design is a crucial component in creating an appealing rich media presentation. Text with disharmonious or indistinguishable colors will degrade the reading experience. Much research has focused on providing user-pleasing color. Munsell has introduced the "Munsell color system" which is widely accepted in the design [Munsell 1950]. According to the color system, the color harmonic schemes include eight types of hue distributions and 10 types of tone distributions [Tokumaru et al. 2002], which are developed into a well-known computational model in [Cheng et al. 2011]. In addition to the harmonization in the color space, some effects attempt to build the link between the color and semantic models, so that the color itself reflects the user's emotion and thus can improve the vividness of imagination. The primitive trail is shown in color image scale [Kobayashi and Matsunaga 1991], where Kobayashi *et al.* first define the relationship between three-color combination with human perception to the image such as "clean and clear" or "dynamic and active." One single color could also contribute to the semantic feeling in [Havasi et al. 2010], where the word is tagged with a certain color like "snow" with the "white" color. In practice, there are a number of online color designing services (e.g., Kuler[2]), supporting the retrieval of semantic colors from a large user-provided dataset.

Our proposed approach is inspired by these previous works. For the automatic generation of visual-textual layout, to the best of our knowledge, we are one of the first

[2]https://kuler.adobe.com/

Table I. Comparisons with previous work

| | | System A | System B | This work |
|---|---|---|---|---|
| Layout | image retargeting | image auto-cropping | – | image auto-cropping |
| | pre-define layout | one text block | one basic layout template | topic-dependent layout templates |
| | font-size | fixed | adaptive | adaptive |
| | visual balance | – | left right balance | symmetric and asymmetric visual balance in golden ratio distribution and art of space |
| Style | color model | – | fixed hue & tone templates | topic-dependent hue & tone templates |
| | color selection way | dominant color from all images in the article | dominant color from cover image with fixed harmonic rules | — color palette from cover image<br>— pre-define semantic colors<br>— topic-dependent harmonic models |
| | font family | fixed | fixed | topic-dependent font sets |

to propose the topic-dependent template integrating human's knowledge and aesthetic principles. The topic-dependent templates in a top-down manner combined with traditional low-level image features in a bottom-up manner, can significantly improve the reading experience.

The work in [Kuhna et al. 2012] focuses on automated digital magazine generation including the magazine content page and magazine cover. Its solution of magazine cover generation is popular and applied in many consumer software such as Flipboard. IUI'13 [Jahanian et al. 2013] targets to automatic design of magazine cover. It supports recommendation and creation of design based on color moods that the user wishes. It also recommends designs based on the users style (of color preference). Therefore, it provides the design of a set of user interactions. We compare our proposed approach with the re-implementation of the generation of magazine cover part in these two representative work. Specifically, Table I lists the similarities and difference among the three approaches. System A and System B are our re-implementation of magazine cover generation part of MM'12 [Kuhna et al. 2012] and IUI'13 [Jahanian et al. 2013]. In addition to the topic-dependent templates in our work, there are more comprehensive considerations of image composition, spatial layouts and color harmonization.

## 3. TEMPLATE DESIGN

The topic-dependent template design is motivated by two main observations in the automatic generation of visual-textual layout. One is the difficulty in precisely describing the visual and textual elements in terms of human perception. Even if an algorithm works well in automatically analyzing individual elements, there still remains a gap between objective and subjective definitions. In particular, the generation of visual-textual layout is a complicated interaction between numerous elements. Rigid computational models derived from aesthetic principles may generate unexpected performance which is a sum of the rules rather than a whole set of aesthetics, or achieve limited performance when the source images are quite diverse in terms of topics (e.g., "'travel," "entertainment," "news," etc.) The other finding is that people used to apply high-level experience and impalpable psychology in designing visual-textual media.

Such kind of knowledge, for example, reflected by putting the most important element on the top or bottom, or rarely coloring blue in an image related to "food & drink" topic, is used to guide the designing, yet it is difficult to be formulated as a computational model.

To bridge the gap between domain-specific design knowledge on layout and computational content features, we introduce topic-dependent templates, which serve as constraints of the interactions between visual and textual elements. Inspired by interviews, with seven design experts from universities and corporations, the templates are defined by two aspects, spatial layout and topic-dependent style. The symmetric and asymmetric visual balance in golden ratio distribution and the art of space are both considered in spatial layout, meanwhile the topic dependent font emotion, font size constraints, semantic colors and color harmonic models are included in the style design.

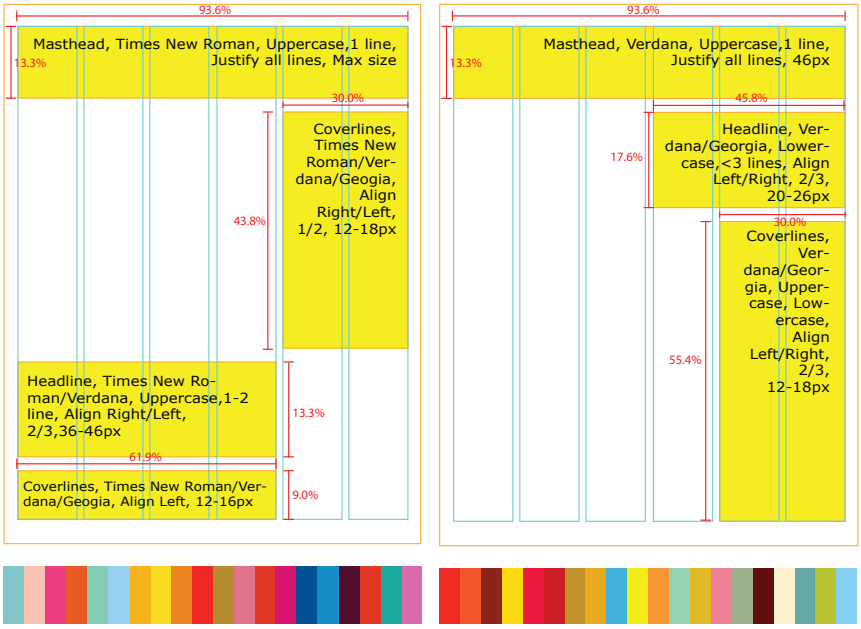
These seven design experts summarized their design style they will follow in designing magazine cover, creating poster, and making Microsoft PowerPoint. The common design style point are listed as follows.

— Textual information completeness. For making a magazine cover visually complete, textual elements should not go beyond the boundary of background image or overlap among themselves.
— Visual information maximization. The image should be resized to target resolution while preserving important visual information (i.e., image regions) such as faces, text, salient objects, human-attended regions, and so on. In addition, the embedded textual elements should not occlude salient regions [Kuhna et al. 2012].
— Spatial layout reasonableness. To be a natural and appealing magazine cover, the position of textual elements should follow aesthetic principles. For example, symmetrical balance are critical rules in human's aesthetic perception, that text should be placed in empty place of background image [Jahanian et al. 2013].
— Perception consistency. The importance of textual elements in visual perceiving and semantic perceiving is consistent. Thus, important text should be displayed in an attractive manner in the non-salient regions, with more distinctive text size, attractive typefaces and higher contrast color. The similar idea is applied in [Yin et al. 2013].
— Color harmonization. From the perspective of vision perception, the color of textual elements should be harmonious and appealing [Cohen-Or et al. 2006].
— Textual information readability. In order to be easily glanced and further understood by readers, textual elements with enough space size are strongly required. Also, color contrast between textual elements and background image should improve the availability of textual information to readers.

It is worth noting that although we mainly focus on the magazine cover style as an instance of visual-textual layout, the proposed system can be easily extended to other visual-textual layouts such as posters, PowerPoint presentations, online self-published rich media, and so on. According to magazine style target media, we define the layout elements as "$Masthead$," "$Headline$," "$CoverLines$," and "$Subtitle$." Figure 2 show two layout templates for the topic of "fashion" and "food & drink."

### 3.1. Aesthetic principles

Some aesthetic principles are essential to construct appealing visual-textual layouts. However, they might be difficult for machines to understand. To bridge this gap, a set of topic-dependent templates with aesthetic principles provides human knowledge and supervises the automatic generation of the visual-textual layouts. Here, a topic is defined as the main category of an image. One image is associated with only one topic




(a) "fashion" template (b) "food & drink" template

Fig. 2. (a) and (b) are two samples in the "fashion" and "food & drink" template, respectively. The template includes the spatial layout, font family and height constraints and semantic colors (better viewed in color). The text regions in the defined spatial layout may vary according to the image importance map.

in our current setting. For example, the input image in Figure 1(a) belongs to the topic of "fashion." In the following parts, we will introduce how to apply the principles in templates including spatial layout and topic-dependent style.

*3.1.1. Spatial layout.* The elements in the layout affect how other elements are perceived. It is therefore important to treat the entire visual-textual layout as a whole rather than the sum of individual visual and textual elements. Many principles of spatial relationships, contrast and similarity, and proportion are applied in the designing of a spatial layout template. We consider the visual weight of each element to keep symmetrical balance, and the golden ratio (i.e. salient objects' positions in an image). We have defined 16 types of common spatial layouts for a magazine cover (see examples in Figure 2). For different topics, the distributions of spatial layouts are different. Before typesetting the textual elements, the spatial layouts are ranked by the topic constraint and the degree of intrusion when the layouts are overlaid on the image. Our spatial layout templates could effectively solve the occlusion problem in Filpboard and [Jahanian et al. 2012] with only one template. In addition, according to the constraints of spatial layout with human knowledge, we can perform better than [Yin et al. 2013] where the textual elements are optimized globally to some unexpected positions.

*3.1.2. Topic-dependent style.* The database "AVA" (a large-scale database for aesthetic visual analysis) supports a large-scale relationship between aesthetic score with the image[Murray et al. 2012]. It is known that different kinds of images correspond to different aesthetic styles. For example, an image in the "fashion" topic is colorful with wide distribution on hue wheel, encouraging the usage of bright colors as well. This is the reason we design the topic-dependent template style. For each topic, the preferred emotion, colors, and even color harmonic models are different. We survey many print-

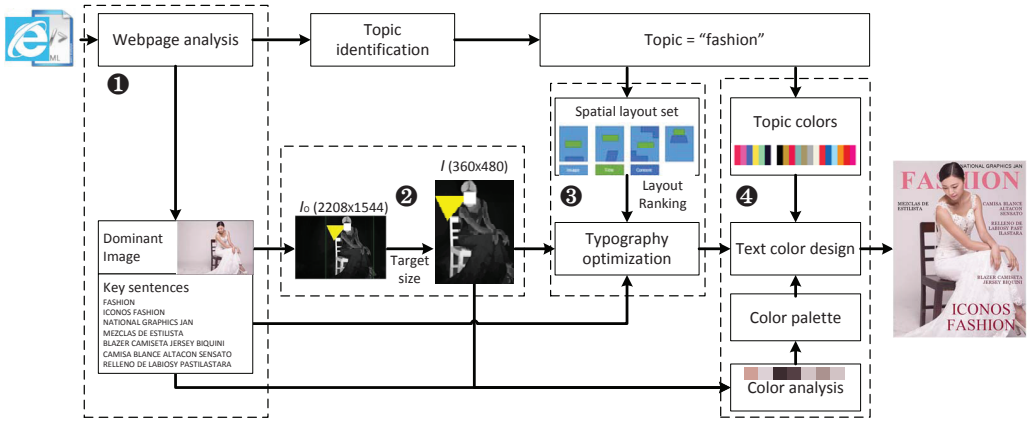



Fig. 3. The framework of automatic generation of visual-textual layout with topic-dependent templates. The framework includes four major modules: 1) the materials generator, where users can directly upload the image and texts or we analyze the webpage to obtain the dominant image and key sentences; 2) image composition, where the original image is automatically cropped and scaled to match the target layout size; 3) typography optimization, where the texts are overlaid on the resized image under the spatial constraints of the selected layout template; and 4) text color design, where the texts are re-colored with consideration of global color harmonization and local readability in a topic-dependent style.

ed magazine covers and posters, as well as popular digital media such as Flipboard, Reuter news, and CNN. We have collected the eight most frequent topics including: "fashion," "economy," "food & drink," "travel," "entertainment," "IT & Tech," "sports," and "politics." For each topic, we have collected around 800 existing rich media posts from popular social media sites and recruited four designers to summarize the designing principles.

The main elements in the style design include font emotions, font size constraints, semantic colors, and color harmonic models. Currently we defined five styles for each topic.

**Font emotion.** We associate font emotion with the font family which determines the external shape of character. The external shape of character acts as a visual element which will stimulate a certain human perception, e.g., serif fonts such as "Times New Roman" with brisk corner bring the happy and elegant feeling while other Sans-serif fonts such as "Segoe" make people feel peaceful and sober. For each topic, we pre-define four to six suitable fonts. By assigning the font family to different layout elements "$Masthead$," "$Headline$," "$CoverLines$," and "$Subtitle$," four font emotion templates are generated. The samples are shown in the Figure 2.

**Font size constraints.** Font size is important in guiding the movement of readers' focus. People are used to perceiving information with descending text size. According to focus flow in our target visual-textual layout [Jahanian et al. 2012], we define four sequent streams "$Masthead$," "$Headline$," "$CoverLines$" and "$Subtitle$," that have a range of reasonable font sizes and keep consistent relative size with each other.

**Semantic colors.** People are very sensitive to color, and color itself reflects semantic information. For example, the eyes are sensitive to red compared with other colors, meaning red is usually used to convey warning or danger. A three-color combination is used to present a certain adjective feeling like "dynamic and active" or "warm" [Kobayashi and Matsunaga 1991]. Some others use five-color combinations, and even one single color which reflects semantic information. We group individual colors into topics. Specifically, given a topic, we provide certain colors that designers regard harmonic and semantically relevant to this topic. For example, the high saturation

"blue" color contributes to the topic of "travel," "health" and "fashion." However, "blue" does not account for "food & drink", since the high-saturation of "blue" causes uncomfortable feelings when put together with food and drink. We support 20 semantic colors for each topic by learning from exist magazine covers on the Internet. Figure 2 provides the semantic colors in the topic "fashion" and topic "food & drink".

**Harmonic color model.** The classical color harmonic models with eight types of hue distribution and 10 types of tone distribution are first introduced in [Tokumaru et al. 2002]. Color models are used to overlay the text with the harmonic color on the layouts, by shifting the text color to the most appropriate harmonic template [Cheng et al. 2011]. The type "V" and "Y" are used most frequently in color harmonic models [Jahanian et al. 2013]. However the models are generic for common images. The text color is designed dependently to local and global image color features. However, as mentioned above, the aesthetic styles change according to different kind of image. The most suitable color harmonic model varies as well. For example, the images of "travel" often contain a natural scene with a wide view and a big portion of natural color. When defining a harmonic text color for travel images, the complementary color to background dominant color is expected, while the images for "fashion" usually present one or two persons. Accordingly, the definition of harmonic text color is analogous to salient dominant color. For each topic, we align one or two color harmonic models from Tokumaru's eight hue harmonic templates according to the topic's attribution [Tokumaru et al. 2002].

### 3.2. Template samples

In total, we cover eight most frequent topics, defining 16 types of common spatial layouts. For each topic, we design topic-dependent styles with 20 semantic colors, four font emotion templates and one or two color harmonic models.

Figure 2 gives two template samples with specific definitions. The magazine cover style layout template is defined as four types of template elements including "$Masthead$," "$Headline$," "$CoverLines$," and "$Subtitle$." For each type of element, we predefine some mask regions shown in the yellow area within the percentage scale of the image. Aesthetic principles of spatial layout are considered in the mask regions. For example, Figure 2(b) is designed with the assumption that when a salient object is located in the left-bottom of the image, the texts should be constraints in the pre-defined areas that guide the textual elements flow from top to bottom. The actual textual area may vary according to the image importance map to avoid occlusion with salient objects. Correspondingly, the range of font size, font family and text alignment are defined as styles in the template.

## 4. GENERATION OF VISUAL-TEXTUAL LAYOUT

In addition to the predefined layout templates, content-based image features such as saliency map should also be considered in the automatic generation of visual-textual layout. By combining high-level template constraints and low-level image features, we define a computational system framework, as shown in 3. The system allows the web analyzer to transfer the web html to the dominant image, key sentences and topic attribution. Since this is not our main focus in this paper, we adopt the extraction approach introduced in [Yin et al. 2013]. The system also allows users to upload a visual background image with a specified topic and some textual sentences. In the second stage, the original image is processed to obtain the visual perception map by combing saliency, face, text, and gaze attention maps. The image is resized to match the target layout size and preserve the important regions according to the visual perception map. The resized image is then used to rank the layout templates in terms of spatial

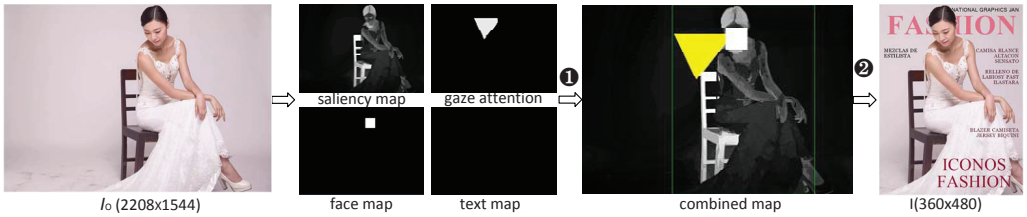



Fig. 4. The complete procedure to align the original image to the target layout. We analyze the original $I_o$ to obtain the saliency, face, text, and gaze attention maps. These maps are combined to measure the importance of different regions. We crop the image to have the same aspect ratio as the target layout by maximizing the importance of the remaining regions and scale it to the resolution of the target layout, so that image $I$ matches the target layout in size and preserves important regions.

distribution. With the resized image, the given sentences, and the spatial layout, the texts are overlaid on the background image by an energy optimization process in stage 3. In the text coloring of state 4, the color palette is first analyzed from the cropped image, while the topic colors are selected through the topic attribute. By applying a certain hue/tone model, color palette, semantic color, and content features, the texts are recolored by keeping the global color harmonization and local readability.

### 4.1. Image composition

There exists a resolution mismatch problem between the original image and the target visual-textual layout. Several image retargeting techniques are proposed to preserve the important regions and minimize distortions [Vaquero et al. 2010]. However, we find that these algorithms usually induce noticeable distortion, especially for images in complex scenes. Considering that image quality is essential for appealing visual-textual layout, we propose a cropping-and-scaling-based image resizing algorithm to address the mismatch problem.

The proposed algorithm crops and scales the original image to the target resolution. It is non-trivial since important regions should be detected and preserved. These regions contain key information like faces, texts, salient objects and human attention. As shown in the Figure 4, we apply saliency detection [Cheng et al. 2011], OCR [Huo and Feng 2003], and face detection [Liang et al. 2008] to the input image. Accordingly, the salience, face, text, and attention maps are computed and the visual perception map is defined as the max operation on all the maps. The image composition from image $I_o$ with resolution $[w_o, h_o]$ to the image $I$ with resolution $[w, h]$ is formulated as maximizing the importance value under the cropping mask with the same aspect ratio to image $I$. The cropped image is then scaled to the resolution $[w, h]$. The cropping and scaling is combined as transformation $T$. Compared with the traditional definition of importance, we instead incorporate gaze attention when detecting the profile or side face, which is quite important in designing visual-textual layout. We get the positions of two eyes on image and the direction of human's head [Liang et al. 2008]. Then the gaze direction can be easily computed, by which we estimate the gaze attention map shown in Figure 4. The importance map is defined as the max operation on saliency map, face and text maps. By applying transformation $T$ on gaze attention map and importance map, we get the gaze attention map $I_a$ and importance map $I_m$ with resolution $[w, h]$, which are useful in following typography process.

### 4.2. Typography

The typography of visual-textual layout is defined as the process of overlaying texts onto the background image. The typography of visual-textual layout is defined as the

process of overlaying several sentences onto the background image. From humans visual perception, the representation of a sentence on an image is usually treated as a text block. The contour of this text block is defined as the bounding box of the corresponding sentence, seeing the red rectangles of image (e) in 5. During displaying these text blocks on visual image, designers follow several basic principles. First, the text blocks should not overlap too much with salient visual objects in original image. The overlapping region is defined as text intrusion. Second, the text blocks should take full use of spare visual space. Third, the text block with important information in semantics should be displayed in an important location in non-salient regions of background image. Accordingly, we formulate the typography as an energy optimization problem that minimizes the cost of text intrusion, the waste of spare visual space, and the mismatch of information importance in perception and semantics, with constraints in the automatically selected templates. We formulate the typography as an energy optimization problem that minimizes the cost of text intrusion, the waste of spare visual space, and the mismatch of information importance in perception and semantics, with constraints in the automatically selected template.

The input of typography consists of the text sentences $\mathcal{S} = \{ S_1 , \ldots, S_N\}$, and the processed image $I$ from image composition. Let $u_i$ denote the weight of semantic importance for the sentence $S_i$. Since the order of sentences indicates the order of importance, we set a descending array $\mathcal{U} = \{u_1, \ldots, u_N\}$, $u_i = N/i$, so that each sentence $S_i$ is assigned with an importance weight $u_i$. Priority of sentences decrease along with the index, as shown by the descending weights $u_1 \geq u_2 \geq \ldots \geq u_N$.

To avoid texts intrusion into salient object in image $I$, the sentence should wrap accordingly. In traditional grid-based methods, image is gridded with fixed width and height, texts are treated as a sequence of individual characters and each character occupies one grid. The visual-textual layout is rigid without aesthetics. We proposed the concept of text block and describe it in an image as $L_i = (p_i, h_i, (x_i, y_i))$. $p_i \in (D_i, U_i, F_i)$ indicates the shape of contour for text block $L_i$. The shape of contour depends on the line wrapping way, alignment way and the text's font family. $D_i$ consists of all possible wrapping ways on sentence $S_i$, a sentence $S_i$ with $m_i$ words has $2^{m_i-1}$ wrapping ways, $U_i = \{$ "$left$," "$center$," "$right$"$\}$ is the alignment ways of text in block, and $F_i = \{$ "$TimesNewRoman$," "$Segoe$," "$Geogia$," "$Verdana$," "$Calibri$"$\}$ contains the aspect ratios of each font family. After the shape of contour is determined, the height of the character $h_i$ in the sentence will scale the shape and control the size of contour for text block $L_i$. $(x_i, y_i)$ is the pixel-wise 2D shift of the text block's left-top point relative to the left-top point of image. $L_i$ describes the representation of sentence $S_i$ on background image $I$. The union of all text block $L_i$ is defined as $L = \{L_i\}$ presenting all the textual contents on visual image, which is showed as red rectangles of image (e) in Figure 5. Further we define the text region as $R(L)$, so we have each pixel covered by the text block belongs to $R(L)$, that is $(x, y) \in R(L)$.

We measure the energy cost from the following three aspects:

$$E(L) = E_s(L) + \mu_u E_u(L) + \mu_m E_m(L), \tag{1}$$

where $E_s$ is the cost of text intrusion into salient visual objects on image $I$, $E_u$ indicates the waste of spare visual space while $E_m$ represents the mismatch between semantic importance $u_i$ and the visual perceived importance $w_i$ of the text blocks.

$$E_s(L) = \sum_{i=1}^{n} a_i J_i, \tag{2}$$

where $J_i = \frac{\sum_{(x,y)\in R(L_i)} I_m(x,y)}{\sum_{(x,y)\in R(L_i)} 255}$, and $a_i \in A$ indicates the weight for each element in a certain template $T$, in our case $A = \{0.1, 0.1, 0.7, 0.1\}$ corresponds to the weight of "$Masthead$," "$Headline$," "$CoverLines$," and "$Subtitle$."

$$E_u = \big(1 - \frac{\sum_{(x,y)\in R(L)} 1}{\sum_{I_m(x,y)\le t} 1}\big), \tag{3}$$

$E_u$ is defined as the waste of spare visual space, meaning that after binarizing the importance map $I_m$ with threshold $t = \max_{(x,y)\in R(L)} I_m(x,y)$, we encourage the full use of regions under threshold.

$$E_m = \sum_{i=1}^{n} a_i \frac{|w_i - ku_i|}{hw}, \tag{4}$$

where $w_i = \sum_{(x,y)\in R(L_i)} I_a(x,y)$, and $k$ is the fitting coefficient which will be adjusted to minimize the $E_m$. $E_m$ is defined to measure the mismatch between semantic importance $u_i$ and the visual perceived importance $w_i$ of the text blocks. The energy aligns important sentences to attractive regions with gaze attention. According to empirical tests, we set $\mu_u = \mu_m = 0.5$, when the face can not be detected in a visual image, the energy $E_m$ will be ignored in the optimizing process so that $\mu_m = 0$.

The solution space of $L = \{L_i\}$, where $L_i = (p_i,\ h_i,\ (x_i, y_i))$ is extremely huge. As defined above, $p_i \in (D_i, U_i, F_i)$ indicates the shape of contour for text block $L_i$. The space of $D_i$ is $2^{m_i-1}$ for the sentence $S_i$ with $m_i$ words. The space of $U_i = \{$ "$left$," "$center$," "$right$"$\}$ is 3. And the space of $F_i = \{$ "$TimesNewRoman$," "$Segoe$," "$Geogia$," "$Verdana$," "$Calibri$"$\}$ is 5. The height of the character $h_i$ is ranged from 1 to $h$. And the shift space of $(x_i, y_i)$ is the whole image $I$. Accordingly, the complexity of each sentence $S_i$ is $O_i = 2^{m_i-1} \times 3 \times 5 \times h \times w \times h$, and the total complexity of solution space is $O = 2^{m-N} \times (3 \times 5 \times h \times w \times h)^N$. Usually the amount of sentences $N$ is $3 \sim 6$, and the total words in the textual content $m$ is $30 \sim 60$. $[w, h]$ is the resolution of image $I$. We cannot find a global optimal solution in an acceptable time. To make the problem solvable, we assume that the element type in the layout works as a series sub-problems of the energy optimization process, so that we can treat each element individually.

*4.2.1. Ranking of Templates.* According to the template design from designers, there are 16 types of spatial layouts with consideration of various visual balances. It is time consuming to try each template for the given image, so we propose a simple yet effective method to rank templates. We define a score for each spatial template as $S_c(L) = 100\big(1 - \sum_{j=1}^{4} a_j Q_j\big)$, where $Q_j = \frac{\sum_{(x,y)\in R(T_j)} I_m(x,y)}{\sum_{(x,y)\in R(T_j)} 255}$, let $T_j$ note the element type $j$ in the layout template, and $R(T_j)$ indicates the mask regions in $j$ type element as the yellow area in Figure 2. The five templates with the highest scores will be filtered to the following process.

*4.2.2. Energy minimization.* To address the complexity problem mentioned previously, we process the typography in each element individually. The whole process is shown in Figure 5. Most previous works binarize the image into salient and non-salient regions by a fixed threshold. In our system, the threshold is adaptive to a sub-optimized problem. We make our search path as follows. We increase the threshold value $t$ from 1 to 256 one by one, the smaller the threshold is, the less error and intrusion we may

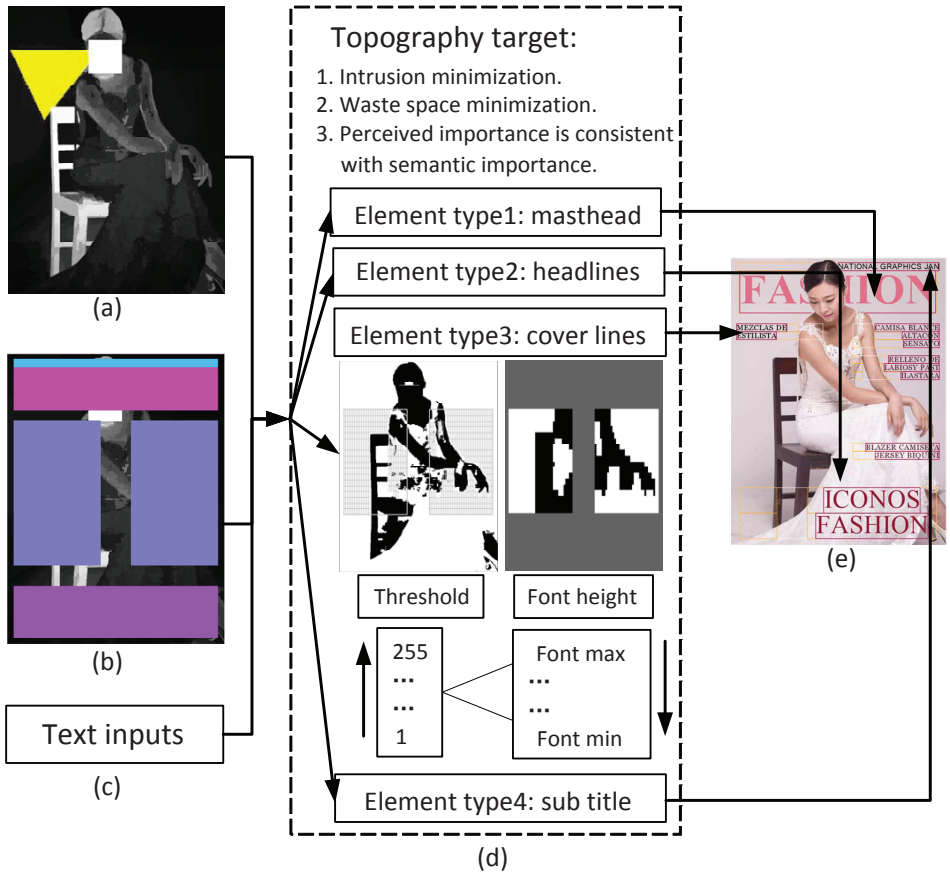



Fig. 5. The typography procedure: (a) the visual importance map (in gray) with gaze attention (in yellow); (b) the selected template from the top-5 ranked templates; (c) the input texts waits; (d) the details of the typography procedure, where the energy defined as $E(L)$ will be minimized in a sub-optimized solution by controlling front height iteratively (e.g., "Coverlines"); (e) the typography result with bottom-up image features and top-down spatial layout constraints.

make but the smaller space could be used to place textual content. Once the threshold $t$ is given, the system try to place the texts in non-salient regions where the importance value $I_m(x, y)$ is lower than current threshold $t$. The threshold will increase gradually until all the texts are inserted into image under constraints. In this way, the adaptive threshold $t$ is determined and the intrusiveness of texts will be minimized. By minimizing the waste of empty space, the font size is selected in a given range. The maximal font size is not only determined by its own constraint, but also bound by other type of template elements with relative size constraints. In most previous work, the textual contents are listed sequentially in the image with the same order in texts. Ignoring the attention distribution on the image, some non-important text may be overlaid on the attractive region. However, in our energy minimization process, the position of text is flexible so that it obtains the consistence of information importance in perception and semantics.


Fig. 6. The illustration of color design procedure for an image in the "fashion" topic. First, the dominant color is extracted from the salient regions in the auto-resized image. Then, one semantic color is selected out with the dominant color in the analogous type harmony model. The masthead of the layout is set as the semantic color. Finally, the "hue" value of other texts are determined by the "i" type hue model [Tokumaru et al. 2002]. To compensate for the contrast with the background, we set the tone of text at the golden ratio between the background tone and the farest possible opposite direction in the tone space in saturation and value coordinates.

### 4.3. Harmonic color design

The color design for textual and graphical elements is always a grand challenge in creating high-quality visual-textual layouts. Since people are very sensitive to color, a harmonic color can generate an eye candy touch to attract users' attention and offer a good experience for long periods of reading. The two requirements for harmonic color design include: 1) keeping text color in global harmonization with the background image, and 2) preserving texts' local readability. In order to satisfy these requirements, we leverage the semantic colors summarized by designers and some well-known color harmonic models [Tokumaru et al. 2002]. We provide a state-of-the-art procedure by adopting the topic-dependent templates in harmonic color design.

As shown in Figure 3, the color palette is extracted from a resized image. The color palette consists of seven colors, in which the first four are from salient objects and the other three are from non-salient objects[Kuhna et al. 2012]. Meanwhile the semantic colors are identified by the image topic, which will be used to supervise the generation of text color. According to the definition of dominant color in the template, the dominant color is selected from the color palette. The semantic colors are iterated to calculate the matching scores with the dominant color in a certain hue harmonic template [Cheng et al. 2011]. The color with maximum response is extracted as a basis color for texts. To meet the first requirements, we apply the type "i" hue harmonic template to control the hue of other texts. After identifying the hue of each text, we apply certain tone models to ensure enough visual contrast against the background.

For different topics, the semantic colors, rules to select dominant colors, and color harmonic models are different accordingly. More details are introduced in Section 3. Figure 6 demonstrates the color design procedure for an image in the "fashion" topic. In "fashion" topic, the dominant color is defined as the most frequent color in the salient region. According to this, the first color in the color palette is selected as the dominant

color which reflects the basis color in visual parts. By applying the analogous hue type in this topic, the basis color for textual elements is assigned to a semantic color which has the maximum matching score with the dominant color in the analogous hue type. Then the harmonic color is selected as the one closest to the dominant color in the hue wheel. In a magazine cover style layout, the "Masthead" with the most salient location and maximum allowable font size, is usually used to determine the basis color of textual elements. Then we set the harmonic semantic color to "Masthead". Based on the color in "Masthead", the texts in other parts are identified through the topic-dependent harmonic models and the local image features. First, the "hue" value of the text is set in "i" type template. To compensate the contrast with the text's local background, we apply an extended tone template. The tone of the text is set at the golden ratio point between the local background tone and the farest possible opposite direction in saturation and value coordinates.

## 5. EXPERIMENTS

In this section, we evaluate the performance of the proposed visual-textual layout system. We build a material database by selecting 104 pieces of news in eight topics, including "fashion," "economy," "food & drink," "travel," "entertainment," "IT & Tech," "sports" and "politics" from three popular websites: CNN , Bing news, and Google news, the distribution of news source is revealed in Figure 7.

We conduct four simple interviews about the reason why a virtual-textual layout is beautiful. Briefly, the background picture, the shape, position and readability of text, the global color harmony of the layout and the overall impression are very essential factors to influence the aesthetic judgement of a visual-textual layout. Apart from the background picture which is not our focus in this paper, the evaluation of our proposed system is based on the following five criteria:

— **Rationality of text position**: whether the text position looks beautiful and reasonable without occluding important objects in the image.
— **Readability of text**: whether the overlaid text is easy to read on the background image.
— **Global harmonization of color**: whether the text color looks in harmony with the whole image.
— **Emotion consistence of the font**: whether the font's family style is consistent with the topic emotion reflected in the image.
— **Overall rating**: the overall impression when seeing the visual-textual layout.

The main goals of the experiments are two-fold: 1) we want to know whether our proposed approach performs well on different criteria, and 2) we want to know whether our proposed approach generalizes layout well on different topics.

### 5.1. Experimental Settings

For each news from web site, we select the dominant image and key sentences according to the approach in [Yin et al. 2013]. To compare the automatic generation algorithm with manual design, the layout are created manually by recruited designers who have never seen our proposed visual-textual layout before. Note that we haven't compare our work with [Yin et al. 2013] because their approach can not process the multiple line text. We also compare our proposed approach with our major relevant work. The visual-textual layouts are generated through re-implementing core and related components (including importance map computing, typography of text, and coloring with harmonic models) in [Kuhna et al. 2012] and [Jahanian et al. 2013], as well as by our proposed system. The following content shows how we actually implemented these parts from [Kuhna et al. 2012] and [Jahanian et al. 2013].

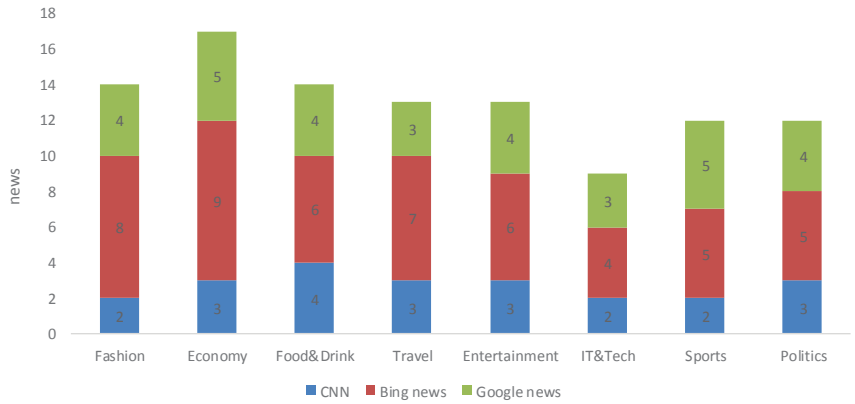



Fig. 7. The distribution of news source.

Huhna et al. provide a semi-automated system so that once a designer has designed a style for a magazine, the processing of articles can be automated based on a set of content-based image features. We re-implemented parts of this work. We calculate the importance map by combining the saliency map, face map, text map, and gaze attention map, and then compute the dominant colors from the background image. Similar to designing a magazine cover style like MM12, the textual contents are placed regularly in a block with translucent background. The position of text block is automatically computed from the importance map as the area with the lowest importance to avoid intrusion. The color of the title is determined by the dominant color. The color of text block is selected from the color with the highest color difference to the area in the underlying image, white, black, translucent white or translucent black. Finally, the color of content text is chose to have highest color difference with text block. The compared work is the re-implementation of part of Kuhnas work [MM12].

Ali et al. [Jahanian et al. 2013] propose a recommendation system for automatic design of magazine covers. We re-implemented parts of this work. The importance map is calculated by combining the saliency map, face map, text map, and gaze attention map, and then the binary mask is obtained by a fixed threshold. The non-salient pixels (black, here) indicate empty regions of background image. To determine which side to use in proposed template, the visual balance is computed. Different weights are assigned to the cells that is a 3x3 grid under each side of the cover image. The whiter the cells, the lighter the weights. By summing the product of each cells weight and the number of salient pixels, the side with a lighter sum of weight will be considered as an empty space to insert coverlines. Since Ali et al. havent shown the specific weight for each cell, we set it all as 1.0. That means the side with less salient pixel will be considered empty. Then, the coverlines are inserted into the non-salient regions line by line and with a space line between two sentences. During the typography the size of text is as large as possible and the form of text indentation follows the boundaries of the binary mask. Finally, the texts are colored by the aesthetic principles. The color of masthead is chose as complementary contrast to the background. For the color of cover lines, the Matsudas hue template and tone template are used to provide color harmonization and legibility. We have added the statement that the results are not generated by the original system, but by our re-implementation according their approaches.

Based on the goals of our user study to verify the effectiveness of our proposed approach on different criteria and different topic, two tasks are designed:

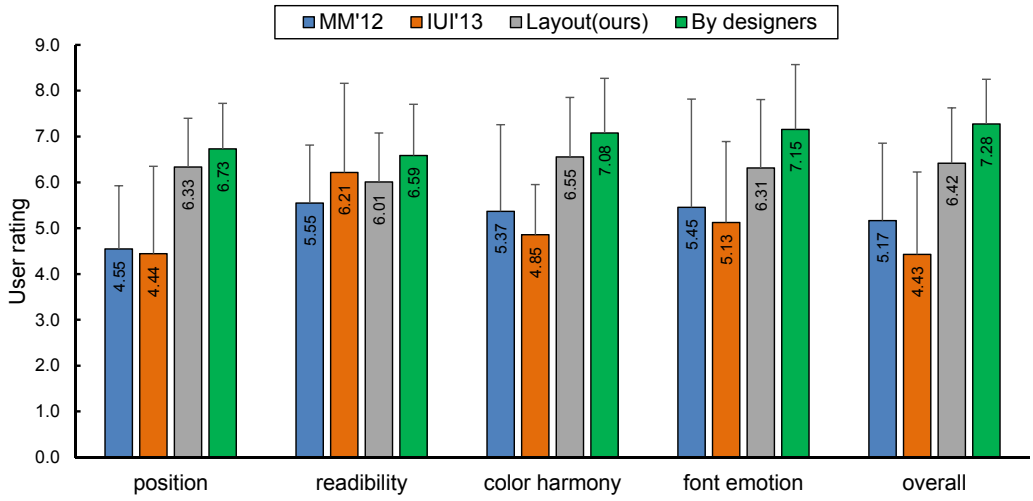



Fig. 8. User ratings of each approach in terms of rationality of text position, readability of text, global harmonization of color, emotion consistence of the font and the overall rating.

— **Task 1**: "Criterion control." For each participant, 10 contrast groups are randomly selected from all the materials. In each contrast group, four visual-textual layouts are displayed to participant. The four layouts are generated by re-implementation of [Kuhna et al. 2012] and [Jahanian et al. 2013], our proposed system and designers in a random order, such as each row in Figure 10. The participants rate the four layouts in a given criterion.

— **Task 2**: "Topic control." In this task, the participants are asked to go through all the layouts in one topic and give an overall rating for the generation approaches including re-implementation of [Kuhna et al. 2012] and [Jahanian et al. 2013], our proposed system and designing manually. The distribution of each topic is revealed in 7.

The rating ranges from 1 to 9 with 5 as the middle point. The higher the rating is, the more satisfied the participants are with the results. We receive questionnaires from 15 subjects aged 21 to 30, among which eight are females, and seven have design-related working experience. Between each two contrastive groups, there is a short break by force to keep participants from tired. Also the visual-textual layouts generated by d-ifferent approaches are displayed in a random order to prevent participants' bias in rating. It takes about one hour to complete the study for each participant.

### 5.2. Evaluations

By conducting the two user study, we collect the rating score to compare our proposed approach with previous works and the ultimate performance by designers. Also the rating score to validate the effectiveness of our topic-dependent template approach are achieved. The following parts are two quantitative analysis and a summary of participants' user feedback. Through the quantitative user study,

*5.2.1. Performance on different criteria.* Figure 8 shows the mean score and standard variance of the four approaches under different criteria.

**Rationality of text position:** our proposed approach makes full use of non-salient space and organizes the text in a reasonable template with aesthetic principles. That's why it performs close to the cover designed by designers. Texts in [Kuhna et al. 2012] are grouped in a text block and placed on the area with minimized intrusion to salient objects. It allows all the textual elements to be perceived as one unit visual block that

results in brevity of designing. However, such strategy leads to occlusion of other visual elements unavoidably. The work in [Jahanian et al. 2013] suffers from the same problem that texts intersect with salient visual objects under a unique basic spatial layout. From the human's bias on error, people cannot tolerate too many occlusions in [Jahanian et al. 2013] where texts are divided in several parts but only one support spatial layout.

**Readability of text:** participants rate highest for [Jahanian et al. 2013] in readability of text, because the color of "masthead" is the complementary color of the image's dominant color, the color of other parts are set in the hue-type "V" with a 45 degree offset in hue. All the colors are in full saturation, black or white, which have the biggest contrasts to the background. That is the reason that it outperforms our proposed approach and [Kuhna et al. 2012]. However there is no significant difference among the three approaches.

**Global harmonization of color:** the text coloring technique adopted in [Jahanian et al. 2013] supports high text readability but it seriously degrades color harmonization. Because pleasing colors are not only related to the harmonic template in the hue wheel, they have a relationship with the tone template as well. People feel depressed by a heavily saturated color. The color used in [Kuhna et al. 2012] is a dominant color, white or black with a translucent mask overlaid on background, which smooths the contrast with the original image. Thus users feel the color is more harmonious than in [Jahanian et al. 2013]. Our proposed approach receives a very high rating due to three reasons. The first is that the basis color is chosen from semantic colors suitable for the topic. Another reason is the definition of dominant color. We divide dominant colors into foreground dominant color and background color. For different topics, different dominant colors and hue models are applied to optimize color harmonization. The last reason is that the tone model results in layering and harmonization.

**Emotion consistence of the font:** : we provide more flexible font styles to match the emotion in the topic, which wins a much higher rating than [Jahanian et al. 2013] and [Kuhna et al. 2012]. Both the latter two methods use fixed font style. It is obvious that the designers' cover outperforms ours. They usually have more choices on font styles, especially some customized font for accurate expression of the emotions.

**Overall rating:** participants give an overall rating with consideration of the above mentioned factors and some others like the text proportion and variability. It is easy to see from Figure 8 that our automatic generation of visual-textual layout achieves close rating with the cover designed manually, which indicates our proposed approach has more practical use compared with previous work.

*5.2.2. Performance on different topics.* One of our main contributions in this paper is that we propose a set of topic-dependent templates with aesthetic principles. To validate the performance of our proposed approach, we conduct the quantitative user study on overall rating of each compared approach on eight popular topics. The statistic result is shown in Figure 9, which reveals the consistent rating in each topic. The re-implementation of parts of MM'12's higher score than the re-implementation of parts of IUI'13' indicates that people prefer a conservative block-based layout than a bad flexible layout. Our proposed approach performs better than the magazine cover generation in re-implementation of parts of MM'12 and IUI'13, because for each topic we have the spatial layout templates and topic-dependent styles which benefit to generate a much pleasing visual-textual layout.

Note that the overall rating for each topic is commonly higher than the overall rating in each contrast group of last study. By collecting the user feedback of our participants, we find the explanation that people usually have impression on the best results and give the overall rating according to the highest rating in a topic.

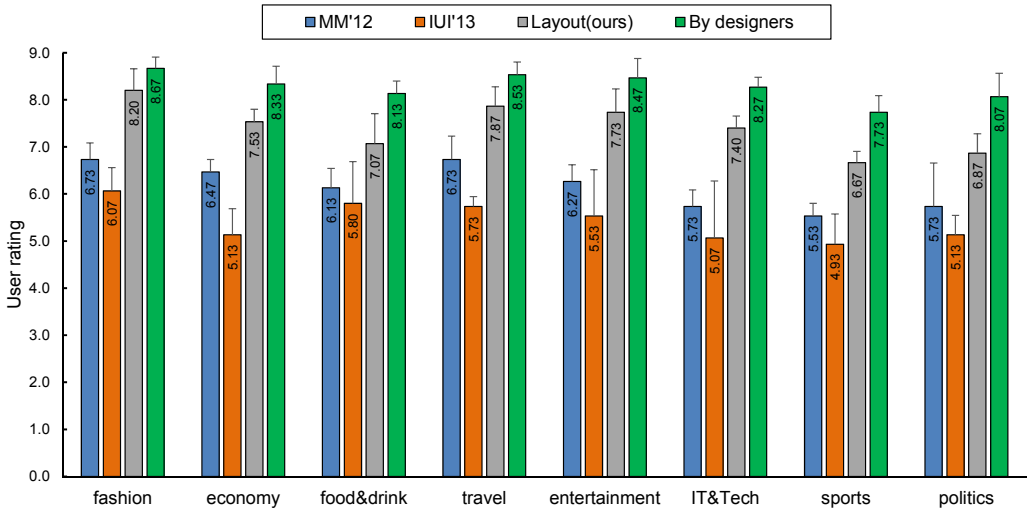



Fig. 9. The overall rating of each approach with eight popular topics.

*5.2.3. User feedback.* We receive a lot feedback from participants. They comment on our generated media that "It is amazing that the automatic generated layout looks so similar to the designer-made one and real magazine cover." "Some results seems to be done by a designer." They think the color of texts in our media is good-looking and harmonious as a whole with the image. "The serif font on fashion topic looks very harmonic." According to such comments, our rating is so close to the magazine cover designed by designers manually.

There are also some suggestions that we should give "Masthead" more flexibility. Moreover they suggest that performance would be better if we could control the font size for each character. In our future work, we will extend our proposed approach to other types of media like poster so as to make the textual elements more spatially flexible.

Some users offer potential applications such as "I will use the results as my e-magazine cover" and "It provides some inspiration for my own design and saves my time if the automatic results look acceptable".

## 6. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a computational framework to generate visually appealing visual-textual presentation layout. It is one of the first attempts towards integrating topic-dependent template with domain specific designing principles to supervise layout creation. The system analyzes low-level image features (in a bottom-up manner) and applies high-level aesthetic designing principles and predefined templates (in a top-down manner) to the given images and texts to automatically suggest the optimal template, text locations and colors. User studies show that the visual-textual layouts generated by our system are able to achieve the best reading experience compared with state-of-the-art automated layouts, while achieving comparable performance with professional magazine covers.

Our work has a lot of extensions as follows. 1) We will enrich both the topic coverage and the topic-dependent templates. 2) We will make the topic identification automatic by leveraging webpage and image analysis techniques. 3) We will extend the layouts from magazine covers to many other rich media such as posters and PowerPoint p-

resentations. 4) We are investigating adding user interaction into our framework to provide personalized layouts.

## ACKNOWLEDGMENTS

This work is supported in part by the National Basic Research Program of China under Grant 2012CB725300, in part by NSFC under Grant 61373072 and in part by the National Natural Science Foundation of China under Grant 61371192. The authors would like to thank Junjie Yu and other designers' help.

## REFERENCES

Rudolf Arnheim. 1954. *Art and visual perception: A psychology of the creative eye*. University of California Press, berkeley and Los Angeles.

Shai Avidan and Ariel Shamir. 2007. Seam Carving for Content-aware Image Resizing. *ACM Transactions on Graphics (TOG)* 26, 3 (2007), 10.

Michael Bauerly and Yili Liu. 2006. Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *International Journal of Human-Computer Studies* 64, 8 (2006), 670–682.

Ming-Ming Cheng, Guo-Xin Zhang, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu. 2011. Global contrast based salient region detection. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 409–416.

Daniel Cohen-Or, Olga Sorkine, Ran Gal, Tommer Leyvand, and Ying-Qing Xu. 2006. Color Harmonization. *ACM Transactions on Graphics* 25, 3 (2006), 624–630.

Catherine Havasi, Robert Speer, and Justin Holmgren. 2010. Automated color selection using semantic knowledge. In *Proceedings of the AAAI Fall Symposium Series*.

Qiang Huo and Zhi-Dan Feng. 2003. Improving Chinese/English OCR performance by using MCE-based character-pair modeling and negative training. In *Proceedings of International Conference on Document Analysis and Recognition*. IEEE, 364–368.

Charles Jacobs, Wilmot Li, Evan Schrier, David Bargeron, and David Salesin. 2003. Adaptive Grid-based Document Layout. *ACM Transactions on Graphics* 22, 3 (2003), 838–847.

Ali Jahanian, Jerry Liu, Qian Lin, Daniel Tretter, Eamonn O'Brien-Strain, Seungyon Claire Lee, Nic Lyons, and Jan Allebach. 2013. Recommendation System for Automatic Design of Magazine Covers. In *Proceedings of International Conference on Intelligent User Interfaces*. ACM, 95–106.

Ali Jahanian, Jerry Liu, Daniel R Tretter, Qian Lin, Niranjan Damera-Venkata, Eamonn O'Brien-Strain, Seungyon Lee, Jian Fan, and Jan P Allebach. 2012. Automatic design of magazine covers. In *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 83020N–83020N.

Shigenobu Kobayashi and Louella Matsunaga. 1991. *Color image scale*. Kodansha international Tokyo.

Alexander Kröner. 1999. The DesignComposer: Context-based automated layout for the internet. In *Proceedings of the AAAI Fall Symposium Series: Using Layout for the Generation, Understanding, or Retrieval of Documents*.

Mikko Kuhna, Ida-Maria Kivelä, and Pirkko Oittinen. 2012. Semi-automated Magazine Layout Using Content-based Image Features. In *Proceedings of the 20th ACM international conference on Multimedia (MM'12)*. ACM, ACM, New York, NY, USA, 379–388.

Lin Liang, Rong Xiao, Fang Wen, and Jian Sun. 2008. Face alignment via component-based discriminative search. In *Computer Vision–ECCV 2008 (Lecture Notes in Computer Science)*, Vol. 5303. Springer Berlin Heidelberg, 72–85.

Ligang Liu, Renjie Chen, Lior Wolf, and Daniel Cohen-Or. 2010. Optimizing photo composition. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 469–478.

Simon Lok and Steven Feiner. 2001. A survey of automated layout techniques for information presentations. *Proceedings of SmartGraphics* 2001 (2001), 61–68.

Tao Mei, Lusong Li, Xian-Sheng Hua, and Shipeng Li. 2012. ImageSense: Towards contextual image advertising. *ACM Transactions on Multimedia Computing, Communications, and Applications* 8, 1 (2012), 6.

Albert Henry Munsell. 1950. *Munsell book of color*. Munsell Color Company.

Naila Murray, Luca Marchesotti, and Florent Perronnin. 2012. AVA: A large-scale database for aesthetic visual analysis. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, IEEE, 2408–2415.


(a) MM'12 (b) IUI'13 (c) Our proposed (d) By designers

Fig. 10. Comparisons with previous work. (a) and (b) are generated by the re-implementation of parts of MM'12 and IUI'13, respectively. Our results are shown in (c), which look natural and professional with balanced spatial layout and harmonic color. (d) are designed manually by recruited designers who have never seen our proposed visual-textual layout. The figure is better viewed in color.

Masataka Tokumaru, Noriaki Muranaka, and Shigeru Imanishi. 2002. Color design support system considering color harmony. In *Fuzzy System, 2002. FUZZ-IEEE'02. Proceedings of the IEEE International Conference on*, Vol. 1. IEEE, IEEE, 378–383.

Daniel Vaquero, Matthew Turk, Kari Pulli, Marius Tico, and Natasha Gelfand. 2010. A survey of image retargeting techniques. In *Proc. SPIE*, Vol. 7798. International Society for Optics and Photonics, 779814–779814–15.

Wenyuan Yin, Tao Mei, and Chang Wen Chen. 2013. Automatic generation of social media snippets for mobile browsing. In *Proceedings of the 21st ACM international conference on Multimedia (MM'13)*. ACM, ACM, New York, NY, USA, 927–936.

Michelle X Zhou and Sheng Ma. 1999. Toward applying machine learning to design rule acquisition for automated graphics generation. In *Proc. 2000 AAAI Spring Symp. on Smart Graphics*. 16–23.

Received November 2014; revised June 2015; accepted June 2015