Optimized Speech Balloon Placement for Automatic Comics Generation

Wei-Ta Chu and Chia-Hsiang Yu

National Chung Cheng University, Taiwan

wtchu@cs.ccu.edu.tw, xneonvisionx@hotmail.com

ABSTRACT

Comic presentation for videos has attracted more and more attention in recent years. This work deeply discusses one important component, i.e., speech balloon placement, that was depreciated and was done by heuristic approaches before. According to number of words and emotion embedded in subtitles, and audio energy corresponding to the targeted frame, speech balloons of various sizes and appropriate shapes are generated. How to locate speech balloons in panels at the same comic page is then formulated as an optimization problem. The objective function integrates an intra-panel cost with an inter-panel cost, where the former is designed to avoid occluding important regions of frames and to direct viewer's gaze, and the latter is designed to build the reading tempo. The experimental results show that the proposed method facilitates higher readability, higher content coverage, and better speech balloon placement.

Categories and Subject Descriptors

H.5.1. [**Information Interfaces and Representation**]: Multimedia Information Systems – *Animations, evaluation/methodology, video*. I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding – *Representations, data structures, and transforms, video analysis.*

Keywords

Speech balloon placement, comic design principles, particle swarm optimization.

1. INTRODUCTION

Efficient presentation methods are demanded to compactly display videos and large amounts of image collections. This issue becomes increasingly urgent as more and more people browse multimedia content on limited-sized portable devices. Although video summarization and highlight detection techniques have been studied for years, novel presentation styles are still demanded to respond to new requests from mobile users. Recently, comics-based presentation for movies and animation videos [2][3][4][6] has emerged as a new type of presentation that carries important information, and provides entertaining and highly interactive browsing experience at the same time. Comic-style presentation is well known to users in all regions with distinct

culture backgrounds, and thus it would be a good choice to present visual content on a limited-sized device, in an organized and systematic way.

Envisioning the potential of comic presentation, several studies have been proposed. The system in [6] automatically transforms movies into comics. Script-face mapping, descriptive frame extraction, and panel stylization constitute three main components. To facilitate frame selection and frame trimming, Sawada et al. [4] utilized eye tracking information and image features to construct importance maps. Focusing on layout design, Chu and Wang proposed an optimization framework [2] to determine numbers of frames putting at the same comic pages, and represented frame importance and layout as vector forms to determine how frames are allocated in different panels. Cao et al. [3] proposed a data-driven approach to construct parametric style models from examples. Rich layout structure and panel shapes can be generated according to user-specified semantics. Matsui et al. [5] especially work on retargeting an image so that important information can be faithfully displayed in a limited-size panel. Chun et al. [9] focused on automatic speech balloon placement in images. Assuming that the correspondence between balloons and their reference areas (often actors' faces) is known in advance, they developed a heuristic algorithm associated with a few rules to place word balloons so that a reasonable reading order can be made. In existing studies, important regions are manually determined [3][5][9], automatically detected [2][6], or a combination of two approaches [4].

The aforementioned studies provide good starts to generate comics from videos. However, how to allocate and place speech balloons in panels, which is a significant component ensuring the readability of a comic, has drawn relatively little attention. The works [2] and [3] only focus on layout design. In [6], speech balloons are just put next to the character's faces. Based on importance maps generated by eye tracking, speech balloons in [4] are placed next to the speaker without occluding important regions. In [9], speech balloon placement is viewed as an optimization problem, and a heuristic algorithm is designed to ensure the placed balloons maintain a reasonable reading order. However, the reasonability of reading order is only guaranteed in single images. In a comic page, there are often many panels containing many images, and the reading order between panels and within panels should be jointly considered. Moreover, a systematic framework should be built to incorporate various comic design rules, which cannot be well addressed by the heuristic procedure proposed in [9].

In this paper, we view panels in the same comic page as a whole, and design an optimization framework associated with comic design principles to simultaneously find best locations of speech balloons in various panels. According to design principles, the virtual line linking balloons in consecutive panels would direct

viewer's gaze and should better pass important regions in images. Therefore, we argue that good balloon placement is not just to avoid occluding important regions, but also demands the consideration of layout and the browsing trajectory. Contributions of this paper are summarized as follows.

- We view panels in the same comic page as a whole, and formulate speech balloon placement (in all panels) as an optimization problem, which is systematically solved by the particle swarm optimization algorithm. We introduce comic design principles and employ them to elaborately place speech balloons.

- We incorporate the speech balloon information into the framework proposed in [2]. We improve this system by not only inserting speech balloons, but also re-designing image importance measurement for layout selection.

The rest of this paper is organized as follows. Section 2 describes important components of an automatic video-to-comics system. Section 3 describes an improved image importance measure that facilitates better layout selection. Section 4 gives details of the proposed optimized speech balloon location module. Section 5 presents evaluation results and Section 6 concludes this work.

2. BACKGROUND

The proposed speech balloon placement module is implemented based on the framework proposed in [2], because it provides a total solution for transforming videos into comics, and the module design paradigm enables us to add/modify components to make improvement. Figure 1 shows the system framework. We first briefly review this framework, and then describe comic design principles related to speech balloons.
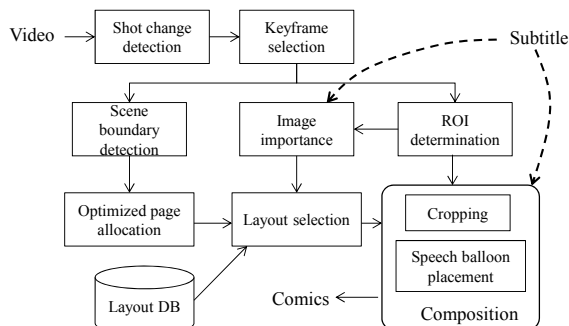

Figure 1. Framework of the automatic video-to-comics system [2].

2.1 System Overview

Given a video, shot boundaries are first detected, and keyframes for each shot are extracted based on color and motion information. Based on keyframes, coherence between shots is evaluated to detect scene boundaries [10]. In [2], the number of comic pages, and which shots are arranged in which pages, are formulated as an optimization problem, which is solved by the genetic algorithm. From each keyframe, the region of interest (ROI) is extracted based on color contrast [11]. Importance of each keyframe is then evaluated based on the ratio of the ROI's area to the whole image. By representing predefined layouts and importance of images in vector forms, best layouts that contain varied numbers of panels are selected to display varied numbers of keyframes. At the composition stage, the region that conforms to the aspect ratio of the targeted panel and covers the most important information (ROI) is extracted from each keyframe, and is then composited into a comic page with a simple resizing operation.

We enhance this system by two points, which are shown by dash lines in Figure 1. First, if there is some subtitle corresponding to a keyframe, this keyframe tends be allocated with a larger panel. Subtitle information is thus considered into the evaluation of image importance. Second, we view panels in the same comic page as a whole, and formulate speech balloon placement in multiple panels as an optimization problem, which is solved by the particle swarm optimization method. Comic design principles are elaborately considered in the design of the objective function.

2.2 Comic Design Principles

Comparing with videos, comics present information with fewer dimensions, i.e., lack of movement and sound. The missing elements can be conveyed to the reader by speech balloons, with the variations of expressive fonts, styles of balloons, and sound effect lettering. In the following, we survey comic design principles related to speech balloons, mainly based on [1].

Dead space. Allocating sufficient space for balloons is incorporated into panel design when an artist decides to arrange a page layout. The space leaving for putting letters is called *dead space*. Without this consideration, putting too many letters in a small panel would occlude vital parts of the artwork, unbalance the composition, and force an unnatural reading order.

Balloon placement and reading order. Placement of balloons is one of the ways to direct the reader's gaze through the page. Western readers often take an initial glace at the entire page, and then scan speech balloons and characters from left to right, and top to bottom[1]. Viewer's attention usually goes from a balloon to a character in a panel, and then goes to the balloon in the next panel. Positions of balloons should be appropriately determined to make the reading order consistent with the reading convention.

Speech balloon shapes. Ellipse balloons with tails pointing to speakers are commonly used. But it is not unusual that artists use a variety of balloon shapes to express emotions and even inner thoughts. For example, jagged edged balloons can suggest exclamations, shouting, and screaming.

Considering the design principles described above, we devise an optimized speech placement module that jointly determines the position and the area of dead space. According to the emotion embedded in subtitles, appropriate balloon shapes are determined to facilitate lively comic presentation.

3. IMAGE IMPORTANCE

Based on the framework in Figure 1, assume that the keyframes $\{f_1, ..., f_n\}$ are determined to put at the same page. The next step is to evaluate the importance value of each keyframe, with which the best layout can be decided. The basic idea is that more important frames are allocated more space. In [2], the importance value of a frame is calculated as the ratio of the area of the ROI to the area of the whole frame. In our work, we advocate that subtitle information provides important clues for importance evaluation. For the keyframes where more words are spoken, larger panels should be allocated for displaying. This is straightforward also because balloons containing more words need larger space. Therefore, the importance value of a keyframe f_i is defined as

---

[1] In Japanese comics, the reading order is right to left, and then top to bottom. We describe western styles by default.

$$M_i = w_1 \times \frac{a_i}{a_N} + w_2 \times \frac{s_i}{s_N} + w_3 \times mCHD_i. \tag{1}$$

The first term is the ratio of the area of ROI in f_i, i.e., a_i, to the total area of ROIs in $\{f_1, ..., f_n\}$, i.e., $a_N = a_1 + \cdots + a_n$. This term measures relative importance between ROIs in frames, rather than the ratios for single frames used in [2]. Note that if there are many ROIs in a frame, the minimum bounding box covering all ROIs is found to calculate this term.

The second term of eqn. (1) is the ratio of the number of subtitle words in f_i, i.e., s_i, to the total number of subtitle words in $\{f_1, ..., f_n\}$, i.e., $s_N = s_1 + \cdots + s_n$.

The third term $mCHD_i$ is the minimum color histogram distance from f_i to other frames in $\{f_1, ..., f_n\}$. That is, $mCHD_i = \min_j \|H_i - H_j\|$, for $j \neq i$ and $j = 1, ..., n$. The notation $\|H_i - H_j\|$ means the normalized Euclidean distance between the RGB histogram H_i of f_i and the RGB histogram H_j of f_j.

Overall, the second term emphasizes the importance of subtitle, while the third term indicates that a frame more visually similar to others is more important. The combination weights w_1, w_2, and w_3 are empirically set as 0.3, 0.3, and 0.4, respectively, because visual coherence is viewed as the most important factor to evaluate image importance. Users can specify different weighting settings to generate preferred comic-style presentation.

Comparing with [2], subtitle information and relationship between frames are further considered by the second and the third terms. Importance values of f_1 to f_n are concatenated as a vector, and the vector form method mentioned in [2] is used to select the best layout.

4. SPEECH BALLOON GENERATION AND PLACEMENT

After layout selection, which frame should be placed into which panel is determined. For a frame f_i, the largest region in f_i that covers the ROI and conforms to the aspect ratio of its targeted panel is extracted. The targeted panel is then pasted with this region with appropriate resizing [2]. Speech balloons are then generated and placed in panels to improve readability of comics.

4.1 Balloon Generation

The factors considered to generate balloons include the number of subtitle words, targeted panel size, sound, and emotion. The first three factors determine balloon size, and the last factor determines balloon shape. It is obvious that the balloon would be larger if it contains more words. In addition, words in different balloons may be with different sizes, because words conveying exclamation or excitement can be enlarged to more faithfully to represent emotion. Emotion can be further exhibited by shape of balloon [1]. Ellipses are most commonly used to show speech in general emotion, while for speech containing excitation, exclamation, and violence words, jagged edged balloons can be used.

Assume that the default size of each character is $w_c \times h_c$, which means every character can be bounded by a box of the width w_c and the height h_c. For English, a character corresponds to an alphabet, and for Chinese, a character corresponds to a Chinese word. In the following, we describe how the height of the bounding box is adjusted, and the width of this box is determined in the same way. Suppose that frame f_i is put into the panel P_i,

and there are some subtitle corresponding to f_i. The height of a character in P_i is calculated as

$$\hat{h}_c = h_c + e(P_i) \times \Delta_h, \tag{2}$$

where $e(P_i)$ denotes the sound energy corresponding to f_i relative to the average sound energy corresponding to all frames put in the same page as P_i. The value $e(P_i)$ is set as follows.

$$e(P_i) = \begin{cases} 0, & e(f_i) < \bar{e} + std(e), \\ 1, & \bar{e} + std(e) \leq e(f_i) < \bar{e} + 2 \times std(e), \\ 2, & \bar{e} + 2 \times std(e) \leq e(f_i). \end{cases} \tag{3}$$

The value $\bar{e}$ and $std(e)$ are the mean and standard deviation of sound energy corresponding to all frames. When the sound energy corresponding to f_i is lower than the average energy plus one standard deviation, no emphasis is applied, and the font size is simply determined by the first term of eqn. (2). When the energy is relatively larger than the average energy plus one or two standard deviations, font size is enlarged by one or two Δ_h, which is set as one percent of the height of a page (in terms of pixels).

Given a character's weight and height, the weight and height of a rectangle able to include all words corresponding to f_i are determined. Based on this rectangle, ellipse or other shapes of balloons are generated. Taking Chinese speech balloons as the example, where Chinese words are shown in a column-major manner, i.e., reading from top to down, and then right to left, the height of the rectangle is determined as

$$h_R = \begin{cases} 0.8h(P_i) & \text{if } \hat{h}_c \times n_c > 0.8h(P_i), \\ \hat{h}_c \times n_c & \text{else,} \end{cases} \tag{4}$$

where n_c is the number of words corresponding to the frame f_i, and $h(P_i)$ is height of the panel P_i. The width of the rectangle is calculated as

$$w_R = \lceil \frac{\hat{h}_c \times n_c}{h_R} \rceil \times \hat{w}_c, \tag{5}$$

where $\hat{w}_c$ is the width of a word, which is basically the same as $\hat{h}_c$. Settings of heights and widths for rectangles showing English words (row-major, reading from left to right, and then top to down) would be reversed.

Based on the $h_R \times w_R$ rectangle, a speech balloon with appropriate shape is generated. We construct a list in advance to store words commonly used for representing violence and excitation. If at least one of the spoken words in a balloon matches with the words in the list, we categorize this balloon as having violent emotion, and general emotion otherwise. Note that the predefined word list should vary for different animations and different languages. Generally, most balloons are categorized as general emotion, and are displayed as the minimum ellipses that cover the rectangles defined by eqn. (4) and eqn. (5). For subtitle containing excitation, exclamation, and violence words, a predefined jagged edged balloon is resized to include the rectangle to show emphasis.

4.2 Balloon Placement

After generating balloons for all panels in a page, the next step is to determine positions of these balloons. The most important guideline for balloon placement is not to occlude important regions in frames [1]. It is quite natural and has been widely adopted to place balloons [4][6][9]. However, balloons positions in different panels are determined separately. In comic design principles, how balloons are placed in different panels lead the

reader's reading order. Therefore, positions of balloons in the same page should be jointly determined.

We treat balloons in the same page as a whole, and formulate balloon placement as an optimization problem. Assume that n balloons $\{B_1, ..., B_n\}$ are to be respectively placed into n panels of a page. The best positions of these balloons, represented by centroids of balloons $\{c_1, ..., c_n\}$, are jointly determined by minimizing the following objective function:

$$E = \alpha E_{intra} + \beta E_{inter}, \qquad (6)$$

where E_{intra} and E_{inter} denote the energy function defined based on information in single panels and between panels, respectively. The weights α and β are set as 0.4 and 0.6, respectively.

The intra term is designed so that (1) the balloon and the ROI have the least overlap, (2) the balloon is placed as close as the ROI, and (3) the balloon is placed as close as the left-top corner of the panel. The third factor is to make the reading order of balloon-object-balloon consistent with left to right, up to down. The energy function E_{intra} is thus defined as

$$E_{intra} = w_1 \cdot \sum_{i=1}^{n} \frac{O_i \bigcap B_i}{O_i} + w_2 \cdot \frac{\sum_{i=1}^{n} d(b_i, e_i)}{\sum_{i=1}^{n} d(b_i, o_i)} + w_3 \cdot \frac{b_i^x \cdot b_i^y}{z_i}. \quad (7)$$

The notation $O_i \bigcap B_i$ means, in the ith panel, the area of the intersection between the ROI O_i and the balloon B_i. In the second term, the value $d(b_i, o_i)$ is the spatial distance from the balloon centroid b_i to the ROI centroid o_i. The value $d(b_i, e_i)$ is the spatial distance from the balloon centroid b_i to the closet point on the boundary of the ROI O_i. Recall that if there are multiple ROIs in a frame, the minimum bounding box covering all ROIs is found to facilitate calculation mentioned above. In the third term, b_i^x and b_i^y respectively mean the x and y coordinates of the balloon centroid b_i, and the normalization term z_i is the maximum value between width and height of the ith panel. Three weights w_1, w_2, and w_3 are empirically set as 0.6, 0.2, and 0.2, respectively. We still put the most emphasis on not to overlap balloons with ROIs.

The energy function E_{inter} is defined as

$$E_{inter} = \sum_{i=1}^{n} |\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle|, \qquad (8)$$

where $\boldsymbol{u}_i$ is the vector from the balloon centroid b_i to the ROI centroid o_i, and $\boldsymbol{v}_i$ is the vector from the ROI centroid o_i to the balloon centroid b_{i+1} in the next panel, if there is a balloon, or the ROI centroid o_{i+1} in the next panel, if there is no balloon. The notation $\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle$ denotes the Euclidean inner product between two vectors. The inter term is designed so that reading tempo is rich. According to [7], the reading tempo in comics is presented by turns of the reading trajectory caused by reading from balloon to object and to next balloon, etc. The design in eqn. (8) prefers that the vector from the balloon to the ROI is orthogonal to that from the ROI to the next balloon.

Finding the best placement can be viewed as finding the best set of positions that causes the minimum objective function in the parameter space constituted by all possible positions. We target on finding a set of positions that simultaneously causes the best placement both for each panel and for the whole page. This problem can be intuitively mapped to the one able to be efficiently solved by the particle swarm optimization algorithm (PSO) [8]. The PSO algorithm is an iterative randomized search algorithm that updates a population (set) of candidate solutions in each iteration. If there are n balloons to be placed in n panels,

initially we randomize one particle for each panel, and thus generate a swarm consisting of n particles. In each iteration, we evaluate the objective function value of the current swarm by eqn. (6), and keep track of the best-so-far positions encountered by the entire population, i.e., *global best*. In each iteration, we also evaluate the objective function value of each particle by eqn. (7) (intra term only), and keep track of the best-so-far position encountered by a specific particle, i.e., *personal best*. In the next iteration, particles move towards associated directions with associated velocities, which are dynamically updated according to their personal bests as well as the global best. The same procedure repeats until the predefined stop condition meets, and the solution that finally gives the global best is found to determine positions of multiple balloons. In this work, the iterative process stops when 300 iterations have completed.

5. EVALUATION

Table 1 shows information of the evaluation dataset, which includes totally ten animation clips associated with Chinese subtitles. We invited 14 subjects (students familiar with animation and comics) in the evaluation. Each was shown the comic pages (by our method and [2]) or storyboards (by the KMP player[2]) generated from three randomly selected animation clips. After watching the comic/storyboard presentation, each subject was asked to give a score from 1 to 5 (larger means better) from the four perspectives of *comprehensibility* (the degree about how a subject understands the presentation), *interestingness* (the degree of fun when reading the presentation), *coverage* (the degree of how the presentation covers important information of the original animation), and *layout* (the degree about whether the presentation layout is appropriate). Figure 2 shows the mean opinion scores in terms the four factors. We clearly see the superiority of our method in comprehensibility and coverage, which mainly accredits to appropriate speech balloons presentation and placement. In terms of interestingness, the superiority of our method and [2] confirms that comics provide more funny presentation over storyboards. Judging appropriateness of layouts is quite subjective, and performance differences between three methods in terms of layout are relatively smaller.

We next investigate the influence of three different objective functions on results of speech balloon placement. The first objective function only includes the first term of eqn. (7), which conforms to most existing works that only take ROIs into account. The second objective function considers all intra terms defined in eqn. (7). The third objective function jointly considers intra terms and the inter term, as defined in eqn. (6). We randomly select 280 comic pages respectively generated by these three functions, and ask subjects to give a score for each page from 1 to 5 according to:

- Q1: Do you think the speech balloons are placed like conventional comic pages?

- Q2: Do you think positions of balloons reasonably direct your gaze to help you understand the content?

Figure 3(a) and Figure 3(b) show results of two questions, respectively. The x-axis denotes five discrete scores, and the y-axis denotes the total times a specific score was given by subjects. For example, in Figure 3(a), the ROI only setting obtains 4 points for totally 42 times. This figure clearly shows that the intra plus inter setting obtains higher scores for more times.

---

[2] http://www.kmpmedia.net/

Table 1. Information of the evaluation dataset.

ID	Name	Length	#shots	#subtitle
1	AIR	1'55''	38	31
2		54''	7	17
3		2'59''	11	65
4	Chu-2	2'12''	26	25
5		1'55''	31	28
6		2'21''	10	55
7	Heaven's Memo Pad	2'37''	21	46
8		2'49''	35	50
9		2'07''	20	47
10	5 Centimeters Per Second	3'53''	54	73
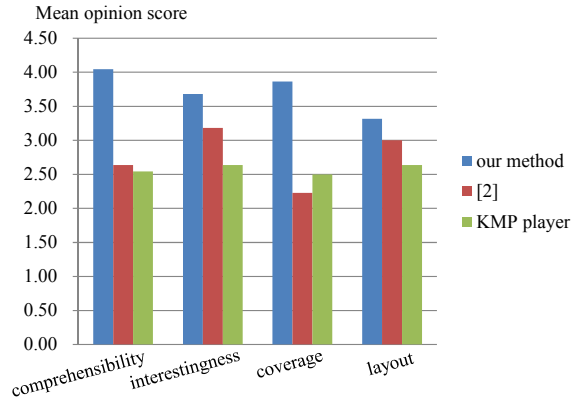
Figure 2. Mean opinion scores in terms of four factors.
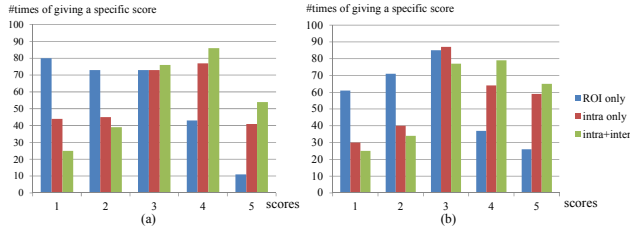

Figure 3. Performance comparison between different objective functions.

Averagely, for the first question, the mean opinion scores for ROI only, intra only, and intra plus inter are 2.4, 3.09, and 3.38, respectively. The mean opinion scores are respectively 2.63, 3.29, and 3.45 for the second question. These results verify that jointly considering information within single panels and between different panels provides lively balloon placement (Q1) and in the meanwhile facilitates reasonable reading order (Q2). Considering ROI only neglects the spatial relationship between panels and often gives boring or unnatural placement.

Figure 4 shows three sample comic pages. Only the ROI term is considered to generate Figure 4(c), and all speech balloons in it are placed at the left-top corners of panels. Although ROIs in its panels are not occluded, balloons placed like this are boring. The objective function defined in eqn. (6) is used to generate Figure 4(a) and Figure 4(b), where more lively balloons are generated and placed. Yellow arrows are superimposed in these samples to show how spatial relationships between balloons and objects, within or between panels, direct viewer's gaze.

The proposed method is essentially working on speech balloon placement for image sequences in specific temporal order. It is thus worth to note that the proposed method is not limited to generate comic pages based on keyframes extracted from animations. Given a photo album consisting of photos with timestamps and annotations, we can employ the same process to generate comic-style presentation, with the associated annotations displayed in speech balloons. Figure 5 shows some sample comic pages generated from a photo album.

Comparing with the most related previous work, i.e., [9], which also formulated balloon placement as an optimization problem, the differences between our work with theirs are at least twofold. First, our optimization framework simultaneously considers balloon placement for *multiple* panels, while the work in [9] only considers balloons in a *single* panel. Second, the solution to the optimization problem in [9] mainly comes from *heuristics*, while our work solves this problem in a more *systematic* way, i.e., particle swarm optimization.
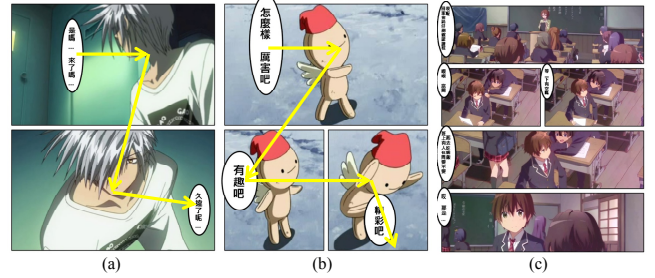

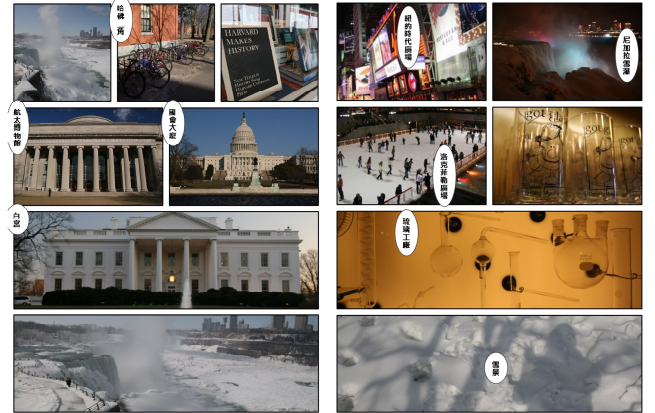**Figure 4. Sample comic pages generated from animations.**

Figure 5. Sample comic pages generated from a photo album.

6. CONCLUSION

This paper presents an optimized speech balloon placement module that improves readability of comic presentation generated by an automatic video-to-comic system. We add newly designed speech balloon generation and placement modules into an existing framework, with the consideration of comic design principles. The main uniqueness of the proposed work is that we view panels in the same comic page as a whole, and develop an optimization framework to systematically solve the speech balloon placement problem. Comparing with existing studies, the proposed optimization approach is systematic, and enables jointly considering information within and between panels. Experimental results verify that the generated comic pages provide higher comprehensibility and coverage.

In the future, we will investigate the optimization issue when multiple balloons are needed to be put into a panel. More comic design principles and comic elements may be integrated in the proposed framework. Currently we just exploit an existing ROI determination method to find important regions in animation keyframes. An ROI determination method specially designed for animations may be a future research direction.

7. REFERENCES

[1] G.S. Millidge. *Comic Book Design: The Essential Guide to Designing Great Comics and Graphic Novels*. Watson-Guptill, 2009.

[2] W.-T. Chu and H.-H. Wang. Enabling portable animation browsing by transforming animations into comics. *Proc. of ACM International Workshop on Interactive Multimedia on Mobile and Portable Devices*, pp. 3-8, 2012.

[3] Y. Cao, A.B. Chan, and R. Lau. Automatic stylistic manga layout. *ACM Transactions on Graphics* (*Proc. of SIGGRAPH Asia 2012*), vol. 31, no. 6, Article No. 141, 2012.

[4] T. Sawada, M. Toyoura, and X. Mao. Film comic generation with eye tracking. *Proc. of International Conference on MultiMedia Modeling*, pp. 467-478, 2013.

[5] Y. Matsui, T. Yamasaki, and K. Aizawa. Interactive manga retargeting. *Proc. of ACM SIGGRAPH*, Article No. 35, 2011.

[6] M. Wang, R. Hong, X.-T. Yuan, S. Yan, and T.-S. Chua. Movie2Comics: Towards a lively video content presentation. *IEEE Trans. on Multimedia*, vol. 14, no. 3, pp. 858-870, 2012.

[7] H. Tobita. Comic engine: Interactive system for creating and browsing comic books with attention cuing. *Proc. of the International Conference on Advanced Visual Interfaces*, pp. 281-288, 2010.

[8] E.K.P. Chong and S.H. Zak. *An Introduction to Optimization*. Wiley-Interscience, 2011.

[9] B.-K. Chun, D.-S. Ryu, W.-I. Hwang, and H.-G. Cho. An automated procedure for word balloon placement in cinema comics, *Proc. of International Conference on Advances in Visual Computing*, vol. 2, pp. 576-585, 2006.

[10] Z. Rasheed and M. Shah. Scene detection in Hollywood movies and tv shows. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 343-348, 2003.

[11] M.-M. Cheng, G.-X. Zhang, N.J. Mitra, X. Huang, S.-M. Hu. Global contrast based salient region detection. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 409-416, 2011.