# Pomics: A Computer-aided Storytelling System with Automatic Picture-to-Comics Composition

Ming-Hui Wen[1], Ruck Thawonmas[2], and Kuan-Ta Chen[3]
[1] Department of Digital Multimedia Design, China University of Technology
[2] Department of Human and Computer Intelligence, Ritsumeikan University
[3]Institute of Information Science, Academia Sinica

***Abstract*—People now use photo browsing, photo and video slideshow, and illustrated text to share stories about their lives in pictures; however, these popular mediums are far from perfect. Some are not expressive enough for sophisticated storytelling, while others inevitably have a high usage threshold and involve a great deal of efforts.**

**In this paper, we propose a framework for comic-based computer-aided storytelling systems to help users become comic storytellers. Such systems take users' photos as the input and output comic strips that tell the story behind the photos. We see the system as a vehicle for media fusion, with the art of comic-making as the basis and inspiration. We also discuss the research challenges involved in improving such systems, and present our proof-of-concept implementation, Pomics (available online at `http://www.pomics.net`).**

## I. Introduction

People now use photo browsing, photo and video slideshow, and illustrated text to share stories about their lives in pictures. Consider the example of Mary, a student who had just finished a self-help trip to India. During the week-long trip, she had many new experiences, including taking a rickshaw for the first time and unexpectedly finding herself in a small, friendly town because she took a wrong train. Upon arriving home, she could not wait to share her exciting experiences with her family and friends. She had taken about 800 photos on her digital camera, so she uploaded them to a web album service. Now, she has to deal with a big problem: How can she use the photos to share the story of her trip with friends? There are four popular options:

- *Photo browsing*: Photos are listed chronologically, usually as thumbnails. Viewers can browse and view individual photos at their will.
- *Photo slideshow*: Photos are shown one by one with each taking around 3–5 seconds. Viewers are allowed to manually fast forward, rewind, or pause automatic transitions.
- *Video slideshow*: Photos are presented sequentially into a video clip, with a voice over or background music, and feature transitions and pan-and-zoom effects. Although VCR-like controls are available, such clips are usually watched thoroughly like music videos and films.
- *Illustrated text*: Each photo in a set is given a short textual description, and presented in the form of a blog article.

Unfortunately, each of these popular strategies have certain limitations:

- *Expressiveness*: Illustrated text is obviously the best choice. Video slideshow is also good for presentation if the photos are properly paced and well annotated with text, voiceover, or music. In contrast, photo browsing and photo slideshow only deliver pictures but not narrate the stories behind them.
- *Production threshold*: A medium's production threshold seems to be positively correlated with its expressiveness. Illustrated text requires writers to master words, phrases, and narration techniques, while video slideshows require the producers to master the timing of photo transitions and the mixing the visual and audio elements (e.g., sound effect and music).
- *Viewers' control*: Viewers normally like the browse stories at their own pace and/or focus on certain events. However, photo and video slideshows give viewers less control on the pace and target of picture browsing.
- *Ubiquitousness*: Except for illustrated text, all the other media forms require an electrical device to function, as the presentation cannot be printed on a paper for reading anytime, anywhere.

Table I summarizes the strengths and weaknesses of the popular media used for photo-based storytelling. Clearly, no medium can satisfy all requirements. Illustrated text is the most effective in terms of expressiveness, readers' control, and ubiquitousness, but it has a high production threshold and viewers must be literate and willing to read text. Therefore, we investigate whether there is another medium that possesses the advantages of current media without their weaknesses.

### *Comic: An Alternative Medium*

A comic can be regarded as an advanced collocation of visual material, with balloons, onomatopœias, and a volatile two-dimensional layout. Comics have a reputation of being shallow and oriented towards younger age groups; however, we believe that they provide an ideal medium for

Table I
COMPARISON OF PHOTO-BASED STORYTELLING MEDIA

| Medium | Production Threshold | Viewers' Requirement | Viewers' Control | Expressiveness | Ubiquitousness |
|---|---|---|---|---|---|
| Photo browsing | Low | Low | High | Low | Low |
| Photo slideshow | Low | Low | Moderate | Low | Low |
| Video slideshow | Moderate | Low | Low | Moderate | Low |
| Illustrated text | High | High | High | High | High |
| Comic | High | Low | High | High | High |

visual storytelling because of the following characteristics: *rich expressivity*, *medium portability*, *reading flexibility*, and *readability*. The rich expressivity is due to the fact that relative importance of photographs and the story's progression shown by the photographs can be expressed with various frame sizes and shapes (e.g., rectangles, quasi-squares, and trapezoids). A picture depicting a critical moment can be given larger space, and an action sequence may be framed in matching trapezoids. Comics often use zig-zag reading lines to guide the reader. This style is more interesting and more efficient than slideshows presented in straight-lines slideshows and line-less galleries, while retaining the planar browseability of the latter. Thus, comics *can be conveyed by any display medium* without information loss. Moreover, readers have *full control over their reading pace and target* as no temporal restrictions are imposed. In terms of the reader's literacy level, available time, and patience, comics require *less effort* than illustrated text because a significant part of the information is conveyed in pictorial form.

Nevertheless, comic creation is not an easy task, especially the storyboarding and layout planning phases. Storyboarding requires creators to select the most informative and representative photos; layout planning involves arranging the photos (i.e., frames) on a fixed rectangular page, where a frame's size and shape are related to its contribution (in terms of storytelling) and graphical content. In addition, editing a comic's two-dimensional layout is much more challenging than editing one-dimensional content, such as a slideshow. Specifically, inserting or removing frames in the middle of a storyboard means that the layout of all subsequent frames must be modified.

Because of the above reasons, comic creation is seldom considered by ordinary computer users, even with the help of comic authoring software such as ComicLife. It seems that only professional comic writers and amateurs, who are skilled in image processing packages, are capable to create comics. When faced with the same dilemma as Mary, i.e., using photos to share personal stories, most users can only resort to photo browsing, slideshow, or illustrated text. *The comic format, as a potential pictorial storytelling medium, is seldom used because of its high production threshold.*

*Simplifying the Comic Creation Process*

To address the challenges involved in comic creation, we propose a framework for *comic-based computer-aided storytelling systems* that would simplify the comic storytelling process for users. We envisage that such systems would have the following capabilities:

1) Take a sequence of digital pictures as input;
2) Identify the events and the storyline in the pictures, and quantify such semantic information;
3) Accept input from creators, including the desired number of pages, the markup style, picture attributes, captions, and conversations;
4) Convert pictures' attributes to visual vocabulary and generate a comic;
5) Allow the creator to fine tune the presentation of the generated comic and re-iterate the process from Step 2 with his/her feedback.

The objective of such systems is not only the semi-automatic generation of comics from photographs of trips, social events, or humorous incidents in life. The systems should be able to accept and deal with *any form of visual media*, such as game screenshots, scanned documents, home videos, and demonstrations/tutorials. In this regard, *we see the software as a vehicle for* ***media fusion****, with the art of comic making as a basis and an inspiration.*

There has been comparatively few researches in this area, which can be classified into two types. The first type focus on methods for authoring comics, e.g., Comic Life [9] and Manga Studio [11], while the second type focus on automatic summarization of text conversation [7], video [5], and interactive 3D games [2, 10] in comics. While the researches in the second category is similar to this work, they are either based on representation-level information (e.g., color features [5]) or application-specific logs [7, 10]). On the contrary, this work focuses on helping users illustrate their own stories using pictures with the assistance of image analysis and understanding techniques.

In the remainder of this paper, we discuss the challenges involved in developing comic-based computer-aided storytelling systems in Section II. We then present our proof-of-concept implementation in Section III and summarize our conclusions in Section V.

## II. RESEARCH CHALLENGES

Designing a comic-based computer-aided storytelling system involves a number of fundamental challenges which can be classified into two categories: image understanding and automatic comic creation.

### A. Image Understanding

To facilitate automatic storytelling, a system needs to be capable of identifying the storyline from a set of time-ordered photographs. While complete understanding of images may not be possible at the moment (even humans cannot always succeed), some clues about the time and place a photo was taken as well as clues about objects and people (e.g., their emotions and behavior) in photos would enhance the automatic narration process significantly. Next, we consider the challenges involved and explain why these challenges are worth taken:

- *Human recognition*: Since most stories involve humans, it is important to identify (automatically) the people (if any) in a photo [3]. Even if we do not know who is the "lead" character in a story, photos containing humans normally have more storytelling elements than scenic phtographs.
- *Emotion recognition*: The emotions of people in a photo, revealed by facial expressions, gestures, and postures [4], could help decide the importance of pictures. For example, photos of a trip with people smiling are normally worth remembering.
- *Behavior recognition*: How people in a photo behave and interact also provides a great deal of information. For example, a group of people giving a victory sign may indicate that they were at a celebration or party. Interactions like shaking hands, shouting, fighting, or raising wine glasses also yield clues about the plot or scenario associated with a picture.
- *Object recognition*: The objects in a photo may indicate the context of an event [8]. For example, sun umbrellas and surfboards may imply a surfing trip; a cake and color balloons may imply a birthday party; and a large number of vehicles and traffic signs may indicate a city intersection during rush hour.
- *Location identification*: Even with a global positioning system, in many cases, image-based location identification is helpful or even necessary to determine the location of an event, especially for indoor environments. For example, if a photo shows pots, pans, a stove, and a microwave, it was probably taken in a kitchen.
- *Natural language processing (NLP)*: If a picture is accompanied by a sound recording, NLP techniques could be used to translate the verbal dialogue into text. This would help us recognize the events associated with the photos and could be used later to annotate comic frames with word balloons and onomatopœias.

### B. Automatic Comic Creation

The second major task of a computer-aided storytelling system involves extracting appropriate pictures automatically from the input photo stream and narrating the story in a comic format. We discuss the steps of the process in the following sub-sections.

*1) Significant photo selection:* Normally, a comic page has between 4 and 16 frames, each of which contains a photo with layered word balloons (e.g., captions and dialogue) and onomatopœias. A creator usually has a lot more photos than he/she would use in comic storytelling, e.g., a five-page comic may be created from 200 photos. For this reason, the first step of comic creation is to automatically select $m$ representative photos from $n$ available photos, where the set $m$ can reasonably describe the story covered by the $n$ photos [6]. Identifying which photos are more representative than others is complex and depends on 1) the story the creator wishes to tell, and 2) the context and semantic details captured in each picture.

In some events, certain moments are typically more memorable or impressive than others; for example, the moment everyone raises their wine glasses at a celebration or the moment a runner crosses the finish line in a race. However, such judgments can be rather subjective. Some people may consider that the silhouette of a couple hand-in-hand couple is romantic and the picture should be part of a comic narrative, while others may think the picture is unsuitable because the couple's faces are hidden.

*2) Pagination and page layout:* Pagination groups the extracted pictures $m$ into $k$ (user-specified) pages. This is a critical step because panels of various shapes cannot be inserted or deleted as easily as paragraphs in a text document without affecting the layout. Thus, pagination must be considered in conjunction with the frame layout on each page. The page layout step arranges comic frames on the $k$ pages so that the frame order is chronological and each frame's size is approximately proportional to its relevance (in terms of narration capability). In addition, a frame's shape must be decided based on the associated picture's content and layout. For example, a horizontal frame would be better for a picture of a car, while a vertical frame would be more suitable for a picture of a high-rise office building.

*3) Narrative design:* Finally, we have to consider the temporal control and annotation of the narrative elements. Temporal control decides the pace of storytelling. Usually, the pace of a story is not even or steady. For example, assuming a creator wishes to describe five interesting events during a trip, for the best narration, the most memorable event would probably span three pages and the other events would be covered by two pages. The concept of time control also applies to specific aspects of an event. In other words, more frames can be used to narrate memorable moments, and less important moments can be dealt with briefly or even omitted.

The annotation of narrative elements, e.g., word balloons, monologues, conversations, and onomatopœias, also plays an important role in storytelling. Sometimes a user-provided picture may contain too much information (e.g., the information may not be relevant to the subject), so a degree of picture clipping may be required. For example, when

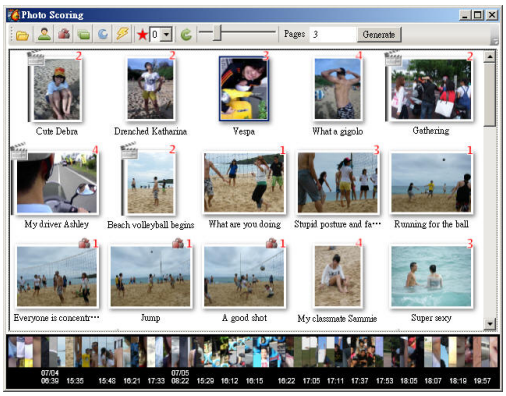


Figure 1. Scoring interface

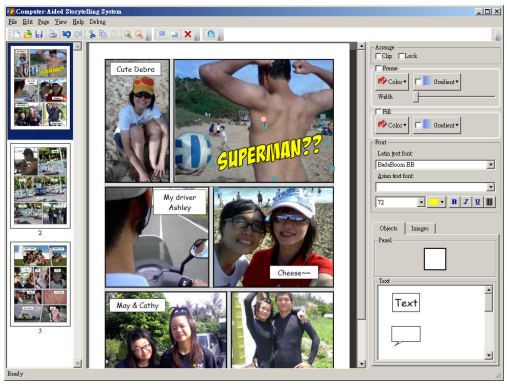


Figure 2. Editing interface

describing a birthday party, a frame that only shows the birthday cake (cropped from a wide-angle picture) would be especially meaningful. Word balloons and onomatopœias can convey the spirit of an event or special occasion and make a comic vivid as though readers were actually attending the event. However, the graphical elements should not cover up meaningful areas of the frame, such as people's faces and important objects. Understanding the meaning of each pixel in a picture is also quite challenging [1].

## III. Proof-of-Concept Implementation

To realize the proposed comic-based computer-aided storytelling system, we have developed a proof-of-concept implementation called Pomics, which is available online at `http://www.pomics.net`. In the following, we explain how we address the challenges in Pomics.

Pomics implements a two-phase comic creation process: photo scoring and comic editing. Figure 1 shows the photo-scoring interface, which consists of a toolbar, a gallery of photographs with their assigned scores, and a scrolling index of thumbnails. When a set of photos is loaded, the system automatically assigns a score to each photo according to the following rules. A photo is deemed representative if it 1) contains people, 2) contains more than one person, 3) is one of a series of shots, 4) shows a new location, and 5) the exposure is acceptable. To detect the presence of humans and human faces, we use OpenCV and its modules. Successive shots and location changes are determined from the time and exposure information in EXIF records; and a color histogram of the pixels is used to judge whether the exposure setting is reasonable. The user interface (Figure 1) displays the computed score (1–9) of each photo so that the user can tune the results and give appropriate descriptions to the pictures.

When the user is satisfied with the score and descriptions, he/she selects the desired number of comic pages, $k$, and presses the "Generate" button to generate a comic. The effectiveness of computer-aided storytelling now becomes evident. No matter how many photos the user provides, the system always creates $k$ comic pages using the most representative pictures. For instance, the user can create a 20-page comic for himself and a 5-page version to share with friends *without any extra effort*. Specifically, the system calculates the number of photos required to create the desired comic (with a certain degree of randomness) and implements an algorithm similar to Huffman coding to subgroup the photos into rows and pages. At the same time, it ensures that the frame size of a picture is approximately proportional to its significance score. When overlaying word balloons on a picture, the system decides the location and size of balloons based on a salience map [6] of the picture so as not to cover informative areas, such as people's faces. The generated comics are then displayed on the comic editing interface, as shown in Figure 2. At this point, the photographs are in place, with the user-given titles and descriptions converted to balloons and text boxes. The user only needs to refine the format by adding onomatopœias, revising the dialogues and narratives, altering panel borders, rotating pictures, or replacing pictures if the result is not satisfactory.

As an example, in Figure 3, we show two sample pages that were generated automatically from a set of 60 travel photos. The pages were not edited manually, so the quality could certainly be improved by some degree of fine-tuning.

## IV. Related Work

Video Manga [5] similarly saw the feature of comics and adopted the medium for video summarization. A video is automatically analyzed and represented with different-sized key frames packed in a visually pleasing form reminiscent of a comic book, allowing users to get a quick overview of a video's contents at a glance without watching the video from beginning to end. Video Manga's algorithm is based on the color features of each frame, which clusters them according to their similarities. The authors introduced an importance score to rank the segments, where a segment is considered to be important if it is long and rare. The key frames are extracted from highly ranked segments and sized according to their scores so that more important key frames


Figure 3. Sample auto-generated comics

are presented as bigger frames. As for authoring support, the generated comics can be enhanced by adding captions to its frames. Text data for captions may be retrieved from transcripts of the video or embedded closed-caption data.

The authors of [10] succeeded in creating sequences of comic-like images summarizing the main events in a virtual world environment and presenting them in a coherent, concise, and visually pleasing manner. Their system can extract important events from a continuous temporal story line using image processing techniques and convert the events into a graphical representation automatically. The system is based on principles of comic theory and can produce different comic sequences on the basis of user-provided semantic parameters like viewpoint and granularity.

On the other hand, Comic Life [9] and Manga Studio [11] are commerical comics authoring software respectively targeted to amateur and professional users. The former boasts its easy-to-use interface—picture drag-and-drop, fancy balloons and onomatopœias, layout templates, and styling filters—while the latter gives very much the appearance of Adobe Photoshop, with more functionality and less restrictions than Comic Life, and imaginably a heavier creation workload.

## V. Conclusion

In this paper, we have proposed a framework for computer-aided storytelling in comics. The system takes users' photos as input and outputs nearly-complete comic strips for further refinement or direct publishing. Although using the comic format for completely automated storytelling may be not practical at present, the output of the developed proof-of-concept implementation, Pomics, satisfied most field testers. The prototype saved the testers a significant amount of time and effort in creating comics (especially selecting photos, paginating, and arranging the page layout). We have demonstrated that a comic-based computer-aided storytelling system can make it easier for people to share personal stories by using their own photos presented in comic format. In our future work, we will continue to develop and improve Pomics[1] in the hope that end users can easily use comics to keep note and share their own life stories with the world.

## VI. Acknowledgement

The authors would like to thank Chien-Hung "Kevin" Lu and Wei-Ju "KK" Chen for their efforts in developing Pomics. We also appreciate Hwai-Jung Hsu, De-Yu Chen, and Yen-Chen "Erin" Tu for their technical and administrative support. This work was supported in part by the National Science Council under the grant NSC101-2221-E-001-012-MY3.

## References

[1] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychological review*, vol. 94, no. 2, pp. 115–147, 1987.

[2] C.-J. Chan, R. Thawonmas, and K.-T. Chen, "Automatic storytelling in comics: A case study on World of Warcraft," in *Proceedings of ACM CHI 2009 (Works-in-Progress Program)*, 2009.

[3] R. Chellappa, C. Wilson, S. Sirohey *et al.*, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[4] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor, "Emotion recognition in human-computer interaction," *IEEE Signal Processing*, vol. 18, no. 1, pp. 32–80, 2001.

[5] Video manga. FX Palo Alto Laboratory, Inc. Http://www.fxpal.com/?p=manga.

[6] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[7] D. Kurlander, T. Skelly, and D. Salesin, "Comic chat," in *Proceedings of ACM SIGGRAPH'96*, 1996, p. 236.

[8] D. Lowe, "Object recognition from local scale-invariant features," in *IEEE International Conference on Computer Vision*, 1999, p. 1150.

[9] Comic life. plasq. Http://plasq.com/comiclife/.

[10] A. Shamir, M. Rubinstein, and T. Levinboim, "Generating comics from 3D interactive computer graphics," *IEEE Computer Graphics and Applications*, pp. 53–61, 2006.

[11] Manga studio. Smith Micro Software, Inc. Http://manga.smithmicro.com/.

[1]Readers please feel free to go online and try out Pomics at `http://www.pomics.net`.