

Continuous Archiving of Group Digital Photograph Collections with a MPEG-7 Based Crowd Sourcing Annotation Methodology

Pei-Jeng Kuo, Department of Computer Science, National Chengchi University, Taipei, Taiwan;

Po-Yu Cheng, Wei-Chen Su, Master's Program of Digital Content and Technologies, National Chengchi University, Taipei, Taiwan;

Sheng-Chih Chen, Master's Program of Digital Content and Technologies, College of Communication, National Chengchi University, Taipei, Taiwan

Abstract

Following the increased convenience and availability of personal recording devices, the number of photographs and video clips one possesses can be multiplied at an accelerated speed. As a result of owning such a huge amount of digital data, one may encounter the difficulty of locating a specific photograph or video. Reformatting of the computer system or damaging on the hard disk also cause a potential risk of losing valuable personal data.

Kuo, Aoki and Yasuda proposed an experiment personal archiving and retrieving image system called PARIS (Personal Archiving and Retrieving Image System) [2]. In PARIS, A MPEG-7 based multimedia description schema with extended spatial and temporal attributes is proposed to annotate personal multimedia data. While PARIS, the proposed experiment system, is designed and implemented specially towards the trend of continuous capture and storage for personal experience; it did not manage to utilize annotations that might come from recent emerging social networking enabled services.

As a continuing project of PARIS, we initiated an experiment platform with thirty smart phone users. At the first stage, which last around six months, we encourage users to capture life events with smart phone cameras and to accumulate related photographs and video clips into our experiment database. In addition, we plan to extend this experiment with another 60 participants, which equipped with personal recording devices such as consumer digital cameras or digital video cameras in the near future.

In this experiment database, we encourage users to upload their media files and provide related annotations. While we provide annotation options with previous proposed MPEG-7 based multimedia description schema, the new system generates annotation suggestions semi-automatically according to various on-line photograph-hosting services.

Known as social-tagging, our proposed system allows users and visitors to create tags based on our previous proposed MPEG-7 structure and utilize pre-defined spatial and temporal ontology as well as resources provided via various on-line crowd generated resources. Users are allowed to create free personal tags and utilize our system generated suggestions in order to create relevant annotations which can lead to increased retrieval precision.

Keywords: Social-Tagging, Information Retrieval, MPEG-7, Crowdsourcing

1. Introduction

With its continuing advancement, the mobile phone has the ability to perform more than its original function of calling, such

as surfing the internet, obtaining GPRS, and navigation. These functions have transformed mobile phones into personal digital assistants (PDAs). Along with keys, wallets and money, the multifunctional mobile phone has become an essential item of one's daily life.

An survey conducted by International Data Corporation (IDC) in 2005 stated that as personal digital image and video recording devices become more common, the number of photographs and videos a person has also increased by a large extent. The constant price reduction for such devices and the improvement in image resolution have also boosted the amount of storage space occupied by high-resolution images and videos.

However, this increase poses challenges for digital archiving and content management. The most common search method is to label images and videos with keywords or tags, which allows content providers and professionals to categorize and create indexes easily. Such method however requires much human effort and resources and hence, detailed categorization is usually left for professionals to execute.

Searching on the internet still mainly depends on text/keyword input on search engines. With various Web 2.0 enabling interactivity between users and administrators, uploading and sharing of internet content is no longer restricted to website administrators and moderators. Administrators, moderators and users now play equivalent roles in producing and uploading massive but diverse internet content, among which includes images and audio-visual materials.

With uploading of images and audio-visual content on the internet becoming more popular, the traditional method of text input on search engines seems to be inadequate when the search is performed on digital image and audio-visual sharing websites. In attempt to provide alternative methods for searching image and audio-visual content, the Moving Picture Expert Group (MPEG) created MPEG-7 (Multimedia Content Description Interface) in 1998, which later became an internationally-accepted multimedia content description standard in 2001. MPEG-7 uses external information about the multimedia content to represent the multimedia content itself (also commonly known as "the bits about the bits"), in order to be available for searches using text.

The increased interactivity on Web 2.0 services also led to the popularization of social networking applications. With reference to the theory "Six Degrees of Separation" and the Small World Model, the social networking website Facebook has become a worldwide phenomenon which integrates real-life and internet-based people networks. Containing more than 600 million active users, Facebook is a multifunctional internet platform which provides its users image/video storage space, blogging capability and instant updating function through "The Wall" where users can

view instant updates on their friends, therefore increasing interactivity between one another.

Facebook also allows users to share images and videos easily with their friends on the same network, especially with the development of mobile technology and constant access to the internet on mobile devices. However, with the increased flexibility and convenience in uploading and sharing of images and videos, new ways of locating desired images and videos have yet to be developed. Crowd-sourcing in tagging of images and videos hence became an increasingly popular concept when there is an open 'community' or 'network' involved.

Aiming to increase effectiveness and efficiency of locating images, Our research aims to increase effectiveness and efficiency for locating images by exploring the concept of crowd-sourcing in tagging of images through the utilization of image-capture functions and sharing on mobile phones and the MPEG-7 standard to document various metadata about the images.

2. Literature review

2.1 Life-Logging

The commonality of digital recording devices has allowed one to record his/her life anywhere and anytime. Daragh Byrne (2007) used SenseCam previously to capture moments in one's daily life. SenseCam is capable of capturing more than 3000 images each day and 20,000 images in a week, which may mean that in a lifetime, Sensecam is capable of capturing 50 million images. The extensive number of images also prompted the need for organizing and managing so as to not become a hindrance when searching for specific images. To accelerate the speed and efficiency of locating images, Bryne used GPS and Bluetooth information to sort out images according to the sequence of events that took place while he was using SenseCam.

Microsoft researcher Dr. Gorden Bell carried out a research experiment in 2002, named MyLifeBits, where books, papers and any paper documents he owned were scanned into digital copies. Besides reducing physical space in his office, Dr. Gorden Bell also pointed out that the human biological memory is unreliable in remembering events and human memory is only recountable by humans themselves through the form of storytelling. He proposed that recounting of events and incidents can be conducted through the screening of images and video clips. An advocate of visual storytelling and digital archiving, he even made PDF copies of webpages he visited and used SenseCam to make records of his life. He proposed that all content in this "E-Memory" should not be deleted but instead, a comprehensive database could be established to aid the locating of information as and when he needed.

However, he also acknowledged that simply containing his digital documents in a harddisk storage space was insufficient. Specific information in the harddisk had to be able to be located quickly and accurately when the need arose. Hence he suggested that a detailed file name containing also appropriate metadata would be more effective and efficient in locating and retrieving specific files. Bell also predicted that by year 2020, digital documentation of a person's life would be fully available and searchable online, further prompting the need to better locate and retrieve specific information.

2.2 Tagging System

To aid the locating and retrieving of information, it became a common action to "tag" digital files. Tagging allows the user to label digital files with customized text or keywords, which in essential described the content of the files or digital documents. To locate images, past search methods only allowed users to enter into a specific category in order to view corresponding images. However, through tagging, it became possible to locate relevant images despite being contained in different categories. For example, online image archive Flickr uses tags to increase users' convenience in finding specific images across multiple categories.

The advantages of customized and precise tagging for locating of files has also prompted websites such as *Del.icio.us*, *Yahoo! MyWeb2.0*, *CiteULike*, *YouTube*, *ESP Game*, *Last.fm*, *Yahoo! Podcasts*, *Odeo*, *Technorati*, *LiveJournal* and *Upcoming* to make use of "communal tagging" (tagging of digital files by a community of users) to increase searching efficiency and accuracy.

2.3 Past Research of MPEG-7

Research on such customized and precise tagging methods have increased in the recent years. For example, Kuo (2004) used the MPEG-7 standard and established a Personal Archiving and Retrieving Image System (PARIS). The PARIS system extended the Structured Annotation Basic Tool of MPEG-7 Multimedia Description Schemes (MDS) and proposed a semantic description tool for multimedia content. The proposed content description tool enabled geo-referenced multimedia data annotation with twelve main attributes regarding its semantic representation. The twelve attributes include answers of who, what,when, where, why and how (5W1H) the digital content was produced, as well as the respective direction, distance and duration (3D) information. Kuo defined digital multimedia content including image, video and music embedded with the proposed semantic attributes as Dozen Dimensional Digital Content (DDDC).

Although the PARIS experiment provided an alternative solution to storing and categorizing images captured by users in their daily lives, "communal tagging" by social networks (or community of users) was not explored in view of the increasing popularity of online communities and social networking sites. The technology advancement in smart phones' sensor functions and the increased accessibility to mobile internet as well as mobile versions of social networking sites increased the possibility of gaining more accurate external and spatial information of captured images.

2.4 CrowdSourcing

Raymond (2010) once said that when "given enough eyeballs, all bugs are shallow", which was the basis for "crowdsourcing", where bugs in a software program will reduce and cease to exist when being reviewed consistently by other users. The term "crowdsourcing" hence originated in the spirit of open source software where the structure of the Linux operating system was in fact constructed by users on the internet network, despite its kernel system being first developed by Liuns Torvalds.

Howe and Robinson of *Wired* magazine further advocated the concept of crowdsourcing in an article ("*The rise of crowdsourcing*") that was published in the magazine in 2006. They proposed to solve problems by utilizing communities of internet users and rewarding them with monetary rewards. Crowdsourcing in real life was applicable in various industries such as retail,

finance and news. For example, *Amazon.com*'s Mechanical Turk was a platform where it helps third-party companies outsource jobs requiring human intelligence to internet users hoping to earn some easy cash, *Wikipedia.org* utilizes collaborative writing to produce its massive encyclopedia, *Threadless* encourages the public to design its t-shirts while *iStockphoto.com* allows amateur photographers to upload and share their images, which are often of similar quality and standards as those taken by professional photographers on commercial online image archives.

3. Conceptual Framework

The conceptual framework for our research can be examined in the following diagrams:

3.1 The Need for Precise Tagging

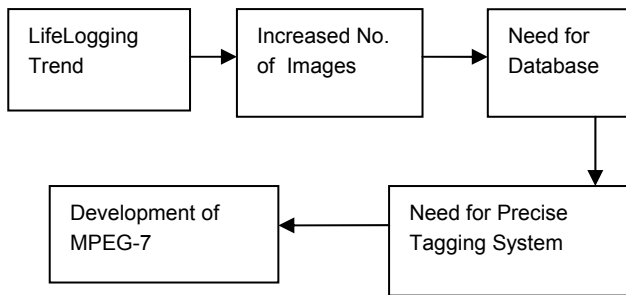


Fig 1. Concept Framework describing the need for precise tagging

The increased availability of camera functions on mobile phones, the development of mobile technology and accessibility of mobile internet, coupled with the increasing popularity of social networking sites, has allowed social networking sites to become real-time sharing platforms for images.

This paper explores the enhancing of search capability for images uploaded on real-time sharing platforms. It is proposed that an information retrieval system can be established through analyzing images uploaded by users on social networking sites using the MPEG-7 standard and examining the process of tagging between users and their social networks.

3.2 Crowdsourcing

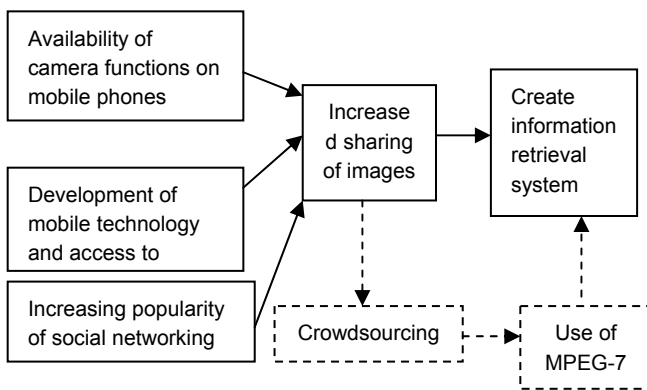


Fig 2. Concept framework describing the trend of crowdsourcing

The easy availability and accessibility to digital recording devices have allowed people to create their own *Lifelogs*. However,

this also resulted in a increased number of images captured which over the years, required external storage space or database to contain and manage them. While Flickr offers such a storage space online, the need for tagging to locate and retrieve images efficiently and accurately also arised, which the MPEG-7 standard of labelling multimedia content with metadata is capable of fulfilling such need.

4. Research Hypothesis

Photo-taking is a form of creating memory of specific events that allows humans to reminisce at a later time as biological memory may become unreliable as one ages. An advocate and practitioner of the personal digital archive, Bell (1995) encourages that one should use the camera to record his/her life anywhere and anytime, and that the content should be stored on an online storage space or database where searching for specific content is enabled. He also pointed out that the most appropriate device to record one's memory digitally was the smart phone.

Although Bell suggested the use of tags to increase searching efficiency and accuracy on online databases, an earlier survey conducted the author showed that in normal circumstances, users do not take initiative in tagging images, hence posing a challenge on how to encourage users to tag images voluntarily on such databases. Hence we hypothesize that through the use of MPEG-7 standard and the collaborative effort of a user's social network, a semi-automatic tagging system can be created to increase efficiency and accuracy of searching for images on online sharing platforms or databases.

5. Methodology

This study uses a daily log application and a survey to examine image-taking behavior and experience of mobile phone users. The daily log application is installed in the phone as part of another related study (the *X-Mind* research project) and is most appropriate for recording data such as time and GPS location, while the survey is the most appropriate tool for understanding users' behavior before and after image-taking and the frequency of accessing, uploading and tagging images on social networking sites through mobile phones.

In the *X-Mind* project, the length of the data recording was one year. However, to guarantee the validity of the result, the data collected at the very beginning of the project is not included. Instead, the result is based on the data collected from three months of highest stability in data transmission and supplemental survey focusing on smartphone's users' behavior and experience.

5.1 The *X-Mind* research project

The daily log application was developed by National Chengchi University in Taiwan for the *X-Mind* research project, which explores the use of smartphone as a device for interpersonal communication which the "camera" function on mobile phones has increased the image-taking behavior among mobile phone users. This study attempts to gain insights into mobile phone users' photo-taking behavior by using information gathered from the daily log.

In the *X-Mind* project, participants were given phones of various models such as HTC Desire, Motorola XT701, Huawei IDEOS, HTC HD2 and Samsung i908, which were equipped with the Android operating system and unlimited 3G internet access.

Each phone was installed with the log application that recorded information on applications that were used.

Participants in the project consisted of students from various disciplines such as radio and television studies, digital content and computer science. A total of 28 participants were selected for the project. At the current stage of the project, participants are required to activate GPS locating function while using the camera function. Data would then be transmitted back to a server so as to gather the time and locations where participants performed image-taking action.

5.2 Survey

Data collected through the daily log application was unfortunately insufficient for providing insight into participants' behavior before and after image-taking. Hence, a separate survey was conducted to supplement the deficiency. The frequency of accessing social networking sites and uploading and ways of tagging them are also noted to provide deeper insights into creating the proposed semi-automatic tagging system. The survey was conducted among 28 participants in the *X-Mind* Project which 17 valid survey responses were collected for further analysis.

6. Findings And Discussion

Data collected from the daily log reveals that the highest frequency for image-taking via mobile phones is during the time zones of 1400 to 1600. There is also a significant increase in the frequency of image-taking after 1800 hours while the lowest frequency happens between 0200 and 0800 hours.

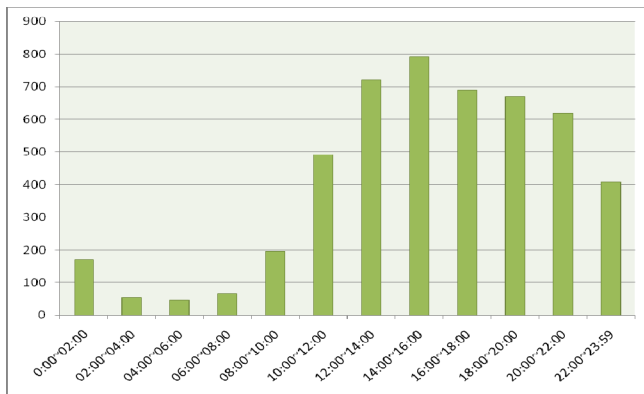


Fig 3. Time distribution of image-taking activity

The survey shows that the most common actions/activities performed by participants prior to the capturing of images are (in the order of popularity): shopping(71%), having meals(50%) and surfing the internet(36%). Actions taken immediately after taking the images include uploading them directly onto Facebook(71%), previewing the images (57%) and showcasing them to surrounding friends (43%).

Table. 1 Activity conducted prior to taking of image

Surfing the internet	36%
Playing games	14%
Doing homework	0%
Reading	14%
Watching TV	0%

Listening to music	7%
Having meals	50%
Shopping	71%
In class	14%
Other	14%

Table.2 Activity conducted after taking of image

Uploading onto Facebook	71%
Adding effect(s) to the image	29%
Previewing the picture myself	57%
Showing image to surrounding friends	43%
Storing only	21%
Other	29%

Through the survey, one has discovered that participants generally use social networking applications in sharing images once a week. When asked about their image-tagging behavior, more than 50% of the participants admitted that they would usually tag images (“agree” and “strongly agree”) while more than 40% expressed their appreciation and willingness to tag images if provided with suitable tagging options. Participants who “disagree” and “strongly disagree” only consist of 21% while 36% chose to remain neutral. When asked about their preferred type of images taken or the types of images they usually capture, most participants enjoy capturing images of locations and attractions, group gatherings, food and travel.

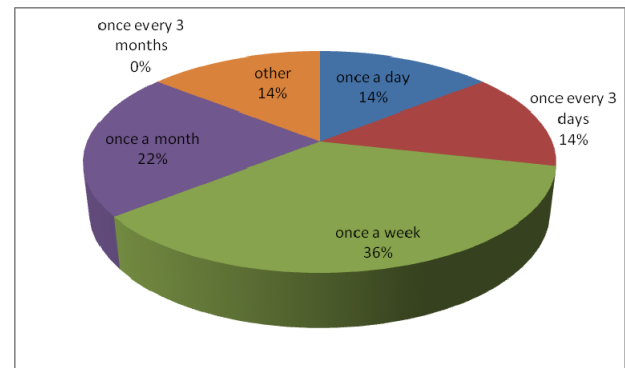


Fig 4. Frequency of using social networking applications for image sharing

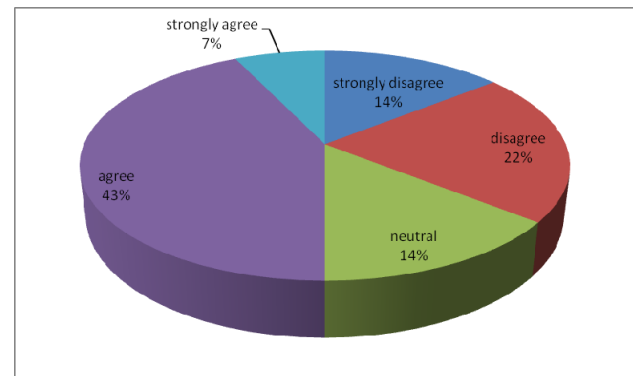


Fig 5. "I would tag photos that I upload"

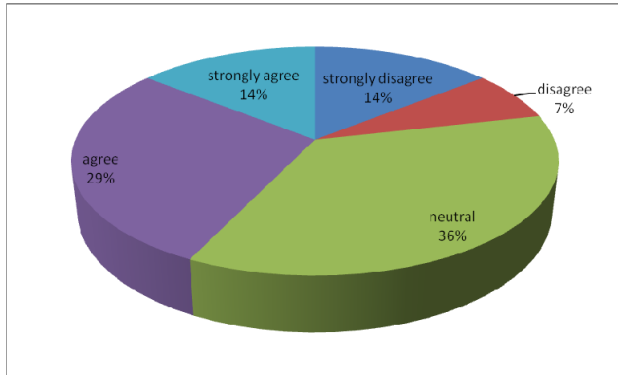


Fig 6. "If the system suggests relevant tagging options, I am willing to tag photos as appropriate"

Table. 3 Preferred type of image(s) taken

Images of self	43%
Group Gatherings	64%
Travel Photos	50%
Photos during KTV sessions	14%
Food	57%
Works of Art	21%
Locations and Attractions	71%
Architecture	29%
Natural Scenery and Animals	43%
Comical and Funny Photos	50%

The above-mentioned statistics paints a clear scenario for image-taking. Image-taking is typically conducted in the afternoon during leisure activities of travelling, eating, having gatherings at locations or attractions, and shopping. After taking images using mobile phones, users' top priority is to share them with others via social networking applications, in which they feel a sense of pleasure from the act of sharing. Users also typically prefer that tagging options are automatically offered when they wish to tag images.

7. Conclusion and future work

This study has proven that a large majority of users are most likely to upload images directly onto social networking applications such as Facebook, yet the frequency of such action is once a week. The 3G internet capability of the smartphone and the mobile versions of social networking applications no doubt prompt users to share images they have taken immediately with others. Users also expressed the wish for a system that offers automatic suggestions while tagging images.

Through an understanding of the preferred type of images taken and the before-after image-capturing behavior of the users on their mobile phones, it is possible to create a tagging system consisting of categories or indexes relevant to the users. Hence, the following functions are proposed for future research:

- 1) Creating a database capable of searching images among various social networking sites, since these sites do not offer such function currently;

- 2) Documenting spatial information of images such as the perspectives and angles they were taken to enhance searching accuracy;
- 3) Designing a new camera application for mobile phones, which is capable of both capturing and searching for desired images.

This research attempts to gain a better understanding of smartphone's user experience and behavior through the data collected from the daily log application and the survey. It also analyzes the image-taking behavior and offers a new search method based on image-taking and tagging via mobile phones within the MPEC-7 structure. Further experiments will be conducted to evaluate the application and suggest possible improvement.

References

- [1] Ames, M. & Naaman, M. (2007). *Why Do People Tag-Motivations for Photo Tagging*.
- [2] Bell, G & Gemmell, J. (2009). *Total Recall: How the E-Memory Revolution Will Change Everything*. USA, New York: Penguin Group Inc.
- [3] Byrne D, Lavelle B, Doherty A, Jones G and Smeaton A.F. (2007). *Using Bluetooth and GPS Metadata to Measure Event Similarity in SenseCam Images*. Accepted for presentation at IMAI'07 - 5th International Conference on Intelligent Multimedia and Ambient Intelligence. Salt Lake City, Utah.
- [4] Gemmell, J., Bell, G., Lueder, R., Drucker, S. & Wong, C. (2002). *MyLifeBits: Fulfilling the Memex Vision*. ACM Multimedia 2002 and Juan-les-Pins, France, pp. 235-238
- [5] Howe, J. (2009). *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. USA, New York: Random House, Inc.
- [6] Marlow, C., Naaman, M., Boyd, D. & Davis, M. (2006). *HT06, tagging paper, taxonomy, Flickr, academic article, to read*.
- [7] P. J. Kuo, T. Aoki and H Yasuda. (2004). *MPEG-7 Based Dozen Dimensional Digital Content Architecture for Semantic Image Retrieval Services*. Proceeding of 2004 IEEE Int. Conf. on e-Technology, e-Commerce and e-Service (EEE 2004).
- [8] P. J. Kuo, T. Aoki and H Yasuda. (2004). *Building Personal Digital Photograph Libraries: An Approach with Ontology-Based MPEG-7 Dozen Dimensional Digital Content Architecture*. Proceeding of IEEE Computer Graphics International 2004 (CGI 2004).
- [9] P. J. Kuo, T. Aoki and H Yasuda. (2005). *Continuous Archiving of Personal Digital Photograph Collections with a MPEG-7 Based Dozen Dimensional Digital Content Architecture*. In Proc. Archiving 2005.
- [10] Stvilia B. & Jørgensen, C. (2009). *User-generated collection-level metadata in an online photo-sharing system*. Library & Information Science Research, 31, 54-65

Author Biography

Pei-Jeng Kuo received her B.S. in Physics from National Taiwan University, Taiwan; M.E. from Massachusetts Institute of Technology, Cambridge, USA; and PhD from The University of Tokyo, Japan. She is currently an Assistant Professor in the Department of Computer Science at National Chengchi University, Taipei, Taiwan. Her research topics include the indexing, archiving, delivering, and retrieving of multimedia contents with MPEG-7 technology.

Po-Yu Cheng and Wei-Chen Su are currently graduate students of the Master's Program of Digital Content and Technologies in National

Chengchi University, Taipei, Taiwan. Their research interests include mobile and web application development.

Dr. Sheng-Chih Chen is currently an assistant professor at the National Cheng Chi University and a research fellow of its Center for Creativity and Innovation Studies. Dr. Chen received his Ph.D. in Computer-Aided Design from the National ChiaoTung University Department of Civil Engineering. His area of concentration includes computer media and cognitive science, computer-aided design, interactive design, virtual reality and communication technology, and multi-media. His primary research interests focus on social network and media, digital media and games, smartphone, and Ipad and so on.