2014 IEEE Pacific Visualization Symposium

Image Flows Visualization for Inter-Media Comparison

Masahiko Itoh*
The University of Tokyo
Masashi Toyoda†
The University of Tokyo
Cai-Zhi Zhu‡
National Institute of Informatics
Shin'ichi Satoh§
National Institute of Informatics
Masaru Kitsuregawa¶
National Institute of Informatics, The University of Tokyo

Abstract

To understand recent societal behavior, it is important to compare how multiple media are affected by real-world events and how each medium affects other media. This paper proposes a novel framework for inter-media comparison through visualizing images extracted from different types of media. We extract blog image clusters from our six-year blog archive and search for similar TV shots in each cluster from a broadcast news video archive by using image similarities. We then visualize such image flows on a timeline in 3D space to visually and interactively explore time sequential changes in influences among media resources and differences and/or similarities between them such as topics that become popular on only blogs or that become popular on blogs earlier than on TV.

Index Terms: H.5.2 [User Interfaces]: Graphical user interfaces (GUI), Interaction styles— [I.3.3]: Picture/Image Generation—Viewing algorithms H.3.1 [Content Analysis and Indexing]: Abstracting methods—

1 Introduction

The first photo of a plane crash landing during the "Miracle on the Hudson" on January 15, 2009 appeared and spread on Twitter and was then used in TV news. During the "Chelyabinsk Meteor" incident on February 15, 2013, many people reported videos of the incident on YouTube then mass media reused them on TV programs. Our use of media has changed dynamically in the last decade, and this affects our societal behavior; mass and social media affect each other.

Both mass and social media provide image flows that are useful for understanding real-world events. Although it is sometimes difficult to represent the reality of events only with text, even a single image can provide reality and impressions of complex events. Images play a role as effective proxies for content to visually tell stories of our interests and experiences [4]. Popular events in the real world usually have representative images/photos such as accident scenes, pictures of new products, and architecture that reflect societal interests. Each medium often uses such images in different ways and at different times.

Visualization of image flows in multiple media resources, such as blogs and TV news, helps us to understand the difference in exposure time of topics between media by checking the occurrence frequencies of topical images from each medium, the effect of media on each other by examining the difference in burst timing, or which medium first provided the information by tracking the origins of these images.

*e-mail: imash@tkl.iis.u-tokyo.ac.jp
†e-mail: toyoda@tkl.iis.u-tokyo.ac.jp
‡e-mail: cai-zhizhu@nii.ac.jp
§e-mail: satoh@nii.ac.jp
¶e-mail: kitsure@tkl.iis.u-tokyo.ac.jp

Visual analytics for extracting events and reading stories from time sequential data-sets are important research domains [10, 15]. There has been much research on visualizing and analyzing temporal changes in topic on various types of media using textural information [27, 16, 8, 7, 9] or images [12, 20, 11, 21]. Luo et al. visualized images from broadcast video news to explore topic relations extracted from video [23]. They however did not use multiple information sources to explore the effect of media on others. Adar et al. studied the impact between different media [1, 28]. However, no study has visualized and explored transitions of image trends related to multiple topics among different types of media resources in 3D space.

This paper proposes a framework for inter-media analysis through image flows extracted from blogs and TV to understand societal behaviors. Both blogs and TV generate a huge amount of image flows every day. Comparing such large image flows from blogs and TV is one of our biggest challenges to capture activities over both media. Therefore, we built two media archives. One is a blog archive that includes two million blog feeds and one billion posts over seven years. The other one is a broadcast news video archive that includes news videos on six TV channels over nineteen months. To compare the occurrence frequencies of these images in both blogs and news videos, we use a scalable shot retrieval index that can search similar shots in news videos from blog images. Images extracted from blogs and TV are visualized in 3D space. For each topic and medium, images are piled up like a time series histogram and are arranged so that the user can easily compare differences in exposure and timing. We also provide a dynamic query function that helps us to explore images with interesting characteristics from visualized images in 3D space.

The two major contributions of this work are:

- Extraction of similar images on blogs as on TV from huge media archives, which enables us to observe the following typical scenarios:
  - What types of topics usually become popular in which medium?
  - For a given topic, which medium first provided the information?
  - If images captured from TV spread on the web, when and how many times were these images broadcasted on TV?
  - If images first appeared on the Web are used on TV programs, when did these images appear and how many users shared them?
- Visualization environment for overviewing long-term trends in image flows from blogs and TV regarding multiple topics and exploration function for observing characteristic images such as those spread from blogs to TV.

We first give an overview of our framework for visual media comparison in Section 2. We next introduce a function for extracting image clusters for specified topics from blog archives in Section 3. We then describe an outline of searching images from news

978-1-4799-2873-6/14 $31.00 © 2014 IEEE
DOI 10.1109/PacificVis.2014.49

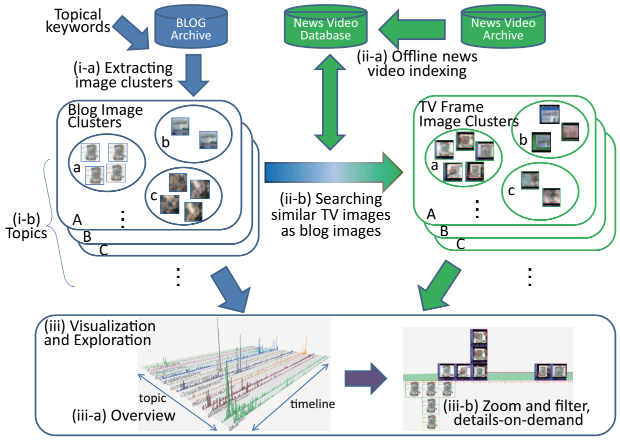


Figure 1: Overview of framework for visual inter-media comparison

video archive in Section 4. Section 5 provides the details of functions for visualization and exploration of image flows. We then show an application and case studies using our framework for analyzing societal behavior in Section 6.

2 Framework for Visual Media Comparison

Our framework aims for comparing occurrence of topical images in social and mass media and tracking the origins of these images. In the first step toward this purpose, we use two media archives. One is a blog archive collected by the University of Tokyo that includes two million blog feeds and one billion posts over seven years. The other one is a broadcast news video archive collected by National Institute of Informatics in Japan that includes news videos on six channels for nineteen months.

We then built a system that enables users to compare exposure of topical images in both media and to detect which medium preceded the other on the topic. Given a set of queries related to the user's interest, our system retrieves relevant blog articles from the blog archive, from which images and surrounding texts are extracted. The extracted blog images are first clustered into sets of near duplicate images based on visual similarity (Figure 1 (i-a)). Each image cluster represents a fine-grained topic of the user's interest. To capture broader topics, these image clusters are clustered again based on textual similarity (Figure 1 (i-b)). We consider these sets of image clusters as topics.

For tracking the origins of blog images, we use the blog image clusters (sets of near duplicated images) as queries for retrieving corresponding shots in the broadcast news video archive. To retrieve shots from large-scale news videos, we build a scalable shot retrieval index (Figure 1 (ii-a)) that enables us to retrieve similar shots from a given image cluster (Figure1 (ii-b)). For a given topic (a set of blog image clusters such as A in Figure 1), we retrieve similar news video shots for each image cluster (e.g. a, b, c in Figure 1). Then retrieved shots are also gathered into a corresponding topic in news videos.

The corresponding topics in blogs and news videos are arranged in 3D space (Figure 1 (iii)) based on the timestamps at which they were posted or broadcasted. We use the first axis in 3D space for time, the second axis for topics, and the third axis for image frequency. To show the frequency of image occurrences over time, images are piled up on each time (e.g. day, week, or month). An overview of all the visualized images enables us to recognize bursting topics, timings, differences and similarities between topics, and those between blogs and TV (Figure 1 (iii-a)). Our framework provides functions that satisfy Shneiderman's mantra [25] for exploring societal trends from visualized images. A function for dynamic query enables us to dynamically access interesting image clusters such as images that became popular on blogs earlier than on TV (Figure 1 (iii-b)).

3 Extracting Blog Image Clusters

We have been collecting two million blog feeds including one billion articles since Feb. 2006. Our blog archive is focused on blogs written in Japanese. About 60% of feeds are in Japanese and the rest are in other languages. We built an inverted index that can retrieve articles from given text queries. For a set of given queries, our system retrieves relevant articles from the blog archive then extracts images and surrounding text included in the articles.

Since web images have diverse qualities, using each single image for retrieving video shots gives us diverse search results. The strategy we adopted is to cluster near duplicate images then merge their visual features before retrieving video shots. We can obtain robust results with this strategy, as shown in Section 4.

We adopt local descriptors that are relatively robust to scaling, rotation, and affine transformations as visual features for clustering images. Typical images in blogs are photos taken by bloggers, and they often include the same object taken from different angles. Other typical images are slightly modified copies from different media. Local descriptors are suitable for determining the similarity of such images. We use Lowe's implementation of SIFT (Scale Invariant Feature Transform) features [22]. The similarity of a pair of images is calculated by the number of matching feature points divided by the average number of feature points in both images.

We use term frequency vectors as textual features for extracting topics (sets of image clusters). We extract several lines of text surrounding each image. For each image cluster, we merge the surrounding text of images in the cluster then build a term vector for the merged text. The similarity of image clusters is calculated by the cosine of the term frequency vectors weighted by term frequency-inverse document frequency (tf-idf).

We can use any standard clustering algorithm such as k-means and hierarchical clustering. In our system, we adopt a modified version of the star clustering algorithm [17] for supporting online clustering by changing threshold parameters. Then each cluster is labeled with sentences chosen from the surrounding text that include frequently occurring terms in the cluster.

4 Matching TV Images with Blog Images

We collected news programs archived over the past nineteen months (starting from Mar. 1st, 2011 until Sep. 30th, 2012) and on six channels (TBS, NHK, TV Tokyo, NET, FUJI and NTV) in Japan. We collected a total of 12,498 news videos with more than 6,000 hours.

This section gives a technical overview for retrieving similar images from the news program archive with query images extracted from the blog archive.

4.1 Offline Indexing

We first offline index those news videos in the archive with a bag-of-words (BoW) framework. The entire flowchart is shown in Figure 2. Since each TV news video may consist of multiple shots, we first detect the shot boundary and segment each news video into shots. Then for each shot, multiple frames are sampled at a rate of 1 fps then for each frame, hessian affine Root SIFT [3] features are extracted. We obtain more than three million shots, twenty-one million frames, and nineteen billion features. After that, one hundred million features are sampled to train a vocabulary made up of one million visual words then the vocabulary is taken for quantizing each frame into a tf-idf BoW vector. Finally, we build an inverted index for all normalized BoW vectors in the database for later online search. The database would be rather large if we take frames as a basic unit. In our experiment, we used average pooling to aggregate BoW vectors of all contained frames of each shot. In this way, each shot could be represented by only one BoW vector in

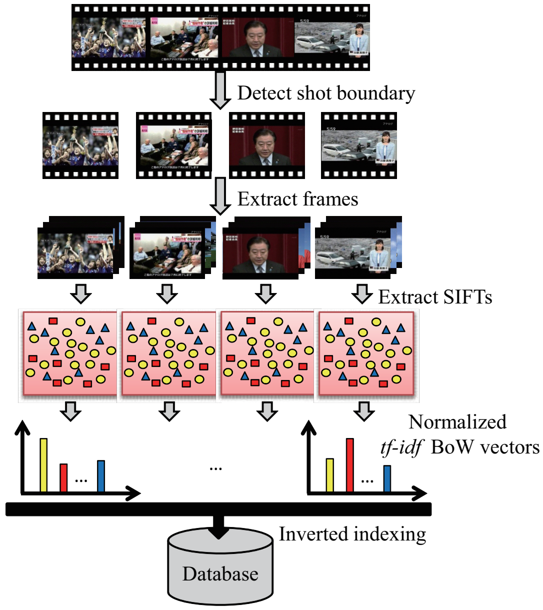


Figure 2: Offline news video indexing

the database. The above average pooling method was first proposed by Zhu and Satoh [29] with which they won the TrecVid Instance Search 2011 Challenge.

4.2 Online Searching

Given a set of query images, a similar process to offline indexing, which includes Root SIFT extraction and BoW quantization, is applied to obtain a normalized BoW representation. The search process is greatly sped up by the inverted index: the initial ranked results are ready in milliseconds. Then the top-ranked 200 videos in the initial ranked list are refined by a local optimized RANSAC [5] based spatial re-ranking, which evaluates geometric consistencies of corresponding multiple feature points between the query image and keyframes of each retrieved video. The process is implemented to run in parallel on 20 cores to achieve reasonable processing time. In total, online searching is less than a second. Thanks to accurate spatial verification, the system is able to filter out most false alarms and return relevant videos with high confidence together with the associated broadcasting information: air time, channel, etc.

5 Visual Exploration Environment

Our visualization system first visualizes all extracted images from blogs and TV for creating an overview of trends. It then enables us to interactively filter out unnecessary image clusters, zoom in to interesting images, and access detailed information about selected images or clusters on demand.

We use a display wall as the frontend of the system. The results of the analyses fill too large spaces to explore using space-limited desktop screens. Moreover, each image on the histograms is too small to recognize detailed content on a normal desktop screen. The display wall allows us to visualize an extremely large amount of images at once while maintaining readable image sizes on the screen. Figure 3 shows a demonstration on the display wall to clearly visualize complex results.

The following are several requirements to visually analyze trends from image flows on various topics.

1. Visualizing the time series of image flows is required to find out the beginning of the topic, changes in topic trends, bursting points, lifetime of trends, and to recognize diversity of images on the topic.

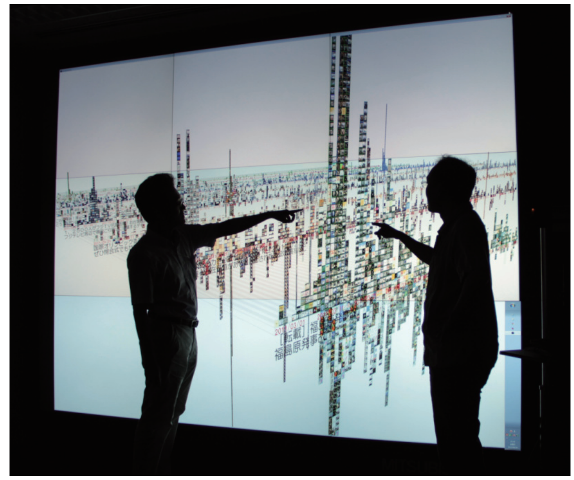


Figure 3: Demonstrating on display wall (9 panels with 4k x 3k pixels)

2. Comparing multiple image flows on different topics is necessary to observe different situations between different topics, trend sequences, and events in the same time sequence on different topics.
3. Comparing similarities and differences between different media resources is required to find out interesting events, which is not extracted using only one media resource, such as events that become hot topics on only blogs or events that become popular on blogs earlier than on TV.
4. Supporting dynamical and easy exploration of image clusters having interesting characteristics is required to find out which medium affects the other and what kind of events become popular on only blogs/TV or both blogs and TV.
5. Accessing original content on both blogs and news videos about selected images on a timeline is required to confirm the context of image flows.

The following Sections 5.1 - 5.5 describe how to materialize each requirement in detail.

5.1 Visualizing Image Flows as Histogram of Images

We adopt a histogram of images for visualizing a time series of images. There are several standard methods for representing a timeline for recognizing bursting points and trends on a topic, such as a polyline chart and histogram. To show not only the frequency of images but also a variety of images and changes in them, we stack images on a timeline as an image histogram.

The image histogram can also show the image diversity on each topic. The images on a topic are diverse when there are various information sources such as news papers, TV channels, or private photos. This is a good indicator for recognizing the activeness of topics and/or media. For instances, we can easily see that the topic shown in Figure 8 consists of various image clusters, whereas the topic shown in Figure 11 (b) consists of very few types of images. Such diversity can be recognized by colors even when the camera is far from the images.

An image histogram uses an the x-axis as the timeline. Images on a topic are aggregated per a specified time window, such as one month, week, or day, and are stacked on the timeline (described as a histogram of images). Examples given in this paper use one day as the time window. We use the y-axis for stacking images on the topic with a specified time window (see Figure 4). Images are sorted on the y-axis by timestamps from top to bottom to easily observe image sequences in detail or by image clusters to easily follow time sequential changes in each image cluster.

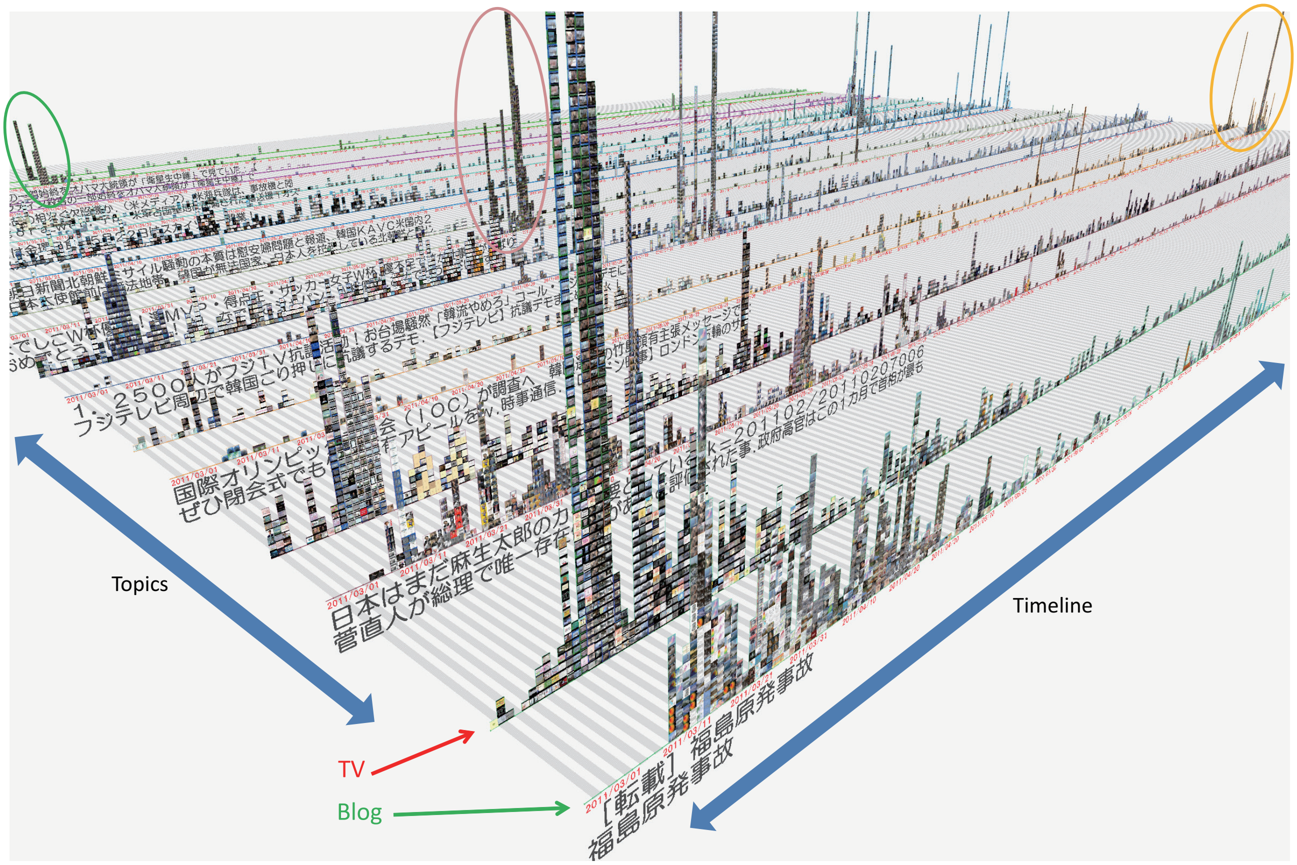


Figure 4: Overview of image clusters from blogs and TV

Labels that annotate topics can be visualized under the stacks of images. We can show summaries of the topic extracted from textual data as annotations. We can display an arbitrary number of labels, but we normally visualize one or two labels from the point of view of readability.

5.2 Multiple Topics Arrangement in 3D space

We arrange multiple histograms of images in 3D space for comparing multiple topics. Topics represented by the histograms of images are arranged along the z-axis in 3D space (see Figure 4).

We can consider several methods to compare multiple image flows for different topics such as arranging multiple histograms or polylines in 2D space or using a heatmap. However, polyline charts and heatmaps are limited in representing the diversity of images. It is also difficult to show an overview of all histograms of images belonging to multiple topics in 2D space. Examples shown in Figure 4 visualize over 16,000 images at the same time. If we arrange these topics in 2D space in parallel, the result uses too much space to display on a screen space.

Multiple image histograms arranged in 3D space allow us to explore the difference in bursting times for every topic, chronological order, and events in the same time on different topics. Figure 5 visualizes topics related to the Great East Japan Earthquake, tsunami, nuclear power plant, and radioactivity. Topics in this scene have time sequential relationships, and we can read stories from the visualized results (details are given in Section 6).

Users can manually or automatically define the order of topics on the z-axis by ranking topics such as the number of images or the frequency of keywords specified by users in the surrounding text. Each topic is colored differently. A frame of each image uses the same color for each topic. Timelines for topics are also colored with the topics' colors. These colored lines represent the topics' active periods.

Perspective foreshortening makes it difficult to compare the heights of histograms from different points from the camera or recognize images taken far from the camera. Our system has an orthogonal projection mode to avoid the problem in which the bars in the different histograms that have the same height look completely the same. We supplementary provide the function for overview + detail, which zooms the selected region to help to see the far-off images.

Histograms of images in 3D space often cause occlusion. Users can zoom, rotate, and pan the 3D space to interactively change a region being focused on and to avoid occlusion.

5.3 Comparison of Images from Different Media

To compare images from different media, we provide two methods for arranging images from blogs and TV; (i) front and back (Figure 6 (i)) or (ii) top and bottom (Figure 6 (ii)). Front and back arrangement enables us to easily find out topics that burst on blogs and TV at the same time even if there is an extremely large amount of images on the timeline (such as circles in Figure 4). However, it is sometimes difficult to compare occurrence times between two types of data resources (Figure 6 (i)). Top and bottom arrangement enables us to easily compare trends on blogs and TV (as shown in Figure 6 (ii)); however, it is difficult for us to obtain an overview if there are too many images on the timeline because the image histograms of the bottom and upper parts overlap. We interactively change the arrangement mode depending on the situation of exploration.

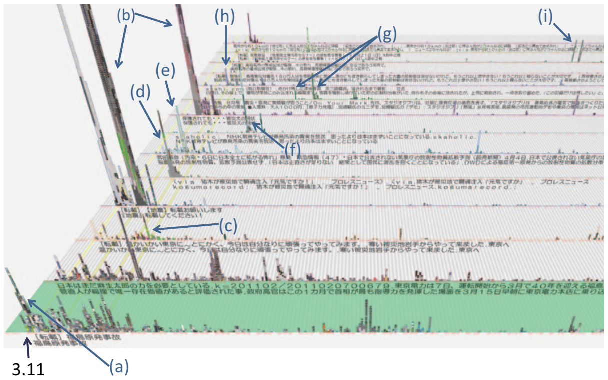
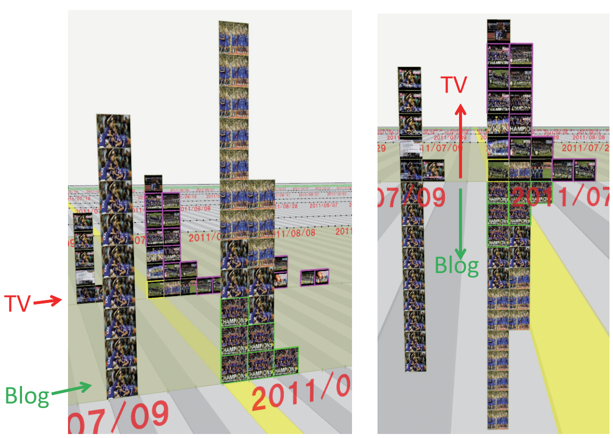


Figure 5: Visualizing topics on blogs related to Great East Japan Earthquake, tsunami, nuclear power plant, and radioactivity: (a) Fukushima nuclear power plant incident, (b) photos of disaster area, (c) photo messages from Fukushima, (d) support for disaster area, (e) radioactive pollution, (f) documentary program on nuclear power pollution, (g) message from Studio Ghibli concerning anti-nuclear power and support for victims, (h) photos of tsunami, and (i) marches protesting nuclear power.



(i) front and back

(ii) top and bottom

Figure 6: Two ways for arranging images from blogs and TV

Such arrangement of image histograms from different media resources enables us to compare image diversity on each topic in different media. For instance, in Figure 11 (a), images from blogs are more diverse than those from TV, which suggest the activeness of the blog.

5.4 Exploring Image Clusters

To explore the similarity and correlation among two data resources, we extract feature values between each pair of image clusters from blogs and TV. Each image cluster can be considered as a time series of frequencies.

To determine similarity, we calculate the cosine similarity between each corresponding time series belonging to the same image cluster. In this case, we use a modified time series using simple moving average, in which each value is calculated as the mean of values in time $t-1$, t, and $t+1$ to absorb small differences such as delay of one day.

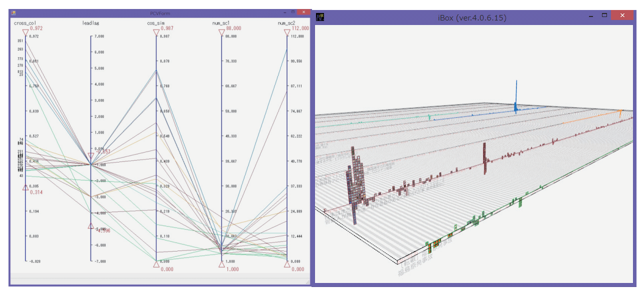
To determine lead/lag among time series extracted from blogs and TV, we calculate the cross-correlation between each corresponding time-series belonging to the same image cluster. We extract the maximum cross-correlation r and delay d at this point, which is the best fitting delay point between the two time series. We use the following cross-correlation function in which b and v are time series of blog images and TV (news video) shots respectively; $\bar{b}$ and $\bar{v}$ are the means of b and v:

$$r(d) = \frac{\sum(b(i)-\bar{b})*(v(i-d)-\bar{v})}{\sqrt{\sum(b(i)-\bar{b})^2}\sqrt{\sum(v(i-d)-\bar{v})^2}}$$

The extracted values are used for sorting clusters or for dynamic query.

We normally merge image clusters into one topic. The system can expand such merged image clusters on the selected topic and arrange them along the z-axis. Image clusters are sorted by cosine similarity, cross-correlation value, or number of images included in each cluster.

To find out interesting image clusters, we provide a function for dynamic query [2]. We can filter out image clusters according to feature values such as v, d, cosine similarity, and number of images in image clusters for blogs and TV. When users set the maximum and minimum threshold values for each feature on the dialog using parallel coordinate view, as shown in Figure 7 (a), the results of visualized images are dynamically updated, as shown in Figure 7 (b).



(a) Parallel Coordinate View (PCV) for dynamic query

(b) Visualizing image clusters only when TV leads blogs

Figure 7: Exploring interesting image clusters by using parallel coordinate view for dynamic query dialog

Users can also interactively pick images to zoom in and see them in detail. The selected images are highlighted and the floor panels that have the same time windows and topics with the selected images are also highlighted to support comparing multiple topics.

Each topic includes multiple image clusters. To recognize a scatter of images belonging to the same cluster in one topic on the timeline, images in a corresponding cluster can be highlighted. To compare the similarity and differences of a scattering of images between blogs and TV on the same topic, images from blogs and TV are highlighted with green and red, respectively (Figures 8 and 11). To focus on the selected image clusters, we filter out image clusters that are not selected. We can also select a topic to focus on and filter out unnecessary topics.

5.5 Accessing Detailed Information

To understand the context of trends in image clusters in detail, our system supports interaction to explore detailed information about a selected image. Each image includes a URL for the blog entry or time of broadcasting and channel name for news video. Users can access the original information source including the selected image such as the blog article or the news video (Figure 9). We also provide a summary page for each image cluster that summarizes the text in the blog articles related to the images in the selected image clusters (Figure 10).

6 Case Studies

In this section, we discuss application case studies using the proposed framework.

Figures 4 and 5 give overviews of image clusters. Figure 4 visualizes the top nine topics sorted by the number of images on each topic. We extract 9,673 blog image clusters from about 4 million images and 654 topics from our blog archive by using the method mentioned in Section 3. Extracted image clusters are related to about 90 topical keywords such as "Great East Japan Earthquake", "nuclear power plant incident" following the earthquake, "women's national soccer team", "London Olympics", and "issues with North Korean rocket", which occurred from Mar. 2011 to Sep. 2012. We then extract TV frame clusters by using the method mentioned in Section 4.

Figure 5 visualizes the top 14 topics on blogs related to keywords such as "Great East Japan Earthquake", "tsunami", "nuclear power plant", and "radioactivity" sorted by the occurrence frequency of keywords in the surrounding text (we display only image clusters from blogs). We can observe time sequential shifts of popular topics by following the bursting points from Figure 5, e.g., just after the earthquake on March 11th, 2011, photos related to the Fukushima nuclear power plant incident became popular. Photos of the disaster area were then diffused. After that, activities and messages for supporting the disaster area and victims spread then messages and marches protesting nuclear power greatly increased.

Section 6.1 gives exploration examples started from these overviews by using the method mentioned in Section 5.4, e.g., image clusters in Figures 8 and 9 are explored from the first topic in Figures 4 and 5.

6.1 Exploration Examples

The examples given in Figures 8 and 9 visualize images on a topic related to the Fukushima nuclear power plant incident after the Great East Japan Earthquake.

Figure 8 shows image clusters related to the outside views of the nuclear power plant. In this case, we select representative images such as damaged buildings, explosions, and aerial photographs of the power plant and filter out other image clusters. Figure 8 suggests that most of these images first appeared on TV then spread throughout the Web. By accessing original news video, we could confirm that TV used the same photos (not videos) as blogs because information sources were very limited.

Figure 9 visualizes an image cluster related to the construction of a nuclear power plant, in which images from blogs were posted 2 days before images from TV were broadcast. We found these images by using the dialog for dynamic query (shown in Figure 7) in which the minimum value for cross-correlation was set to 0.54 and the minimum and maximum values for number of days blogs posted images before TV broadcasted them was set to 1 and 3, respectively. By accessing original content, we can understand the explanation of the structure and situation of the Fukushima nuclear power plant by a researcher at MIT was spread throughout the Web. Television programs then began to use the same image to explain the Fukushima incident.

An example shown in Figure 10 visualizes image clusters related to the topic about the Japan women's national soccer team called Nadeshiko[1]. Figure 10 shows that there are two bursting times. The first peak was on July 14, 2011. It was the day Nadeshiko won the semi-final of the Women's World Cup 2011. The second peak was on July 18, 2011 when Nadeshiko won the final. The image flows on blogs and TV burst at the same timing at both peaks. We found these image clusters by filtering out image clusters with small similarity between blogs and TV. From the visualized image, we can assume that people who depend on mass and social media are equally interested in these topics.

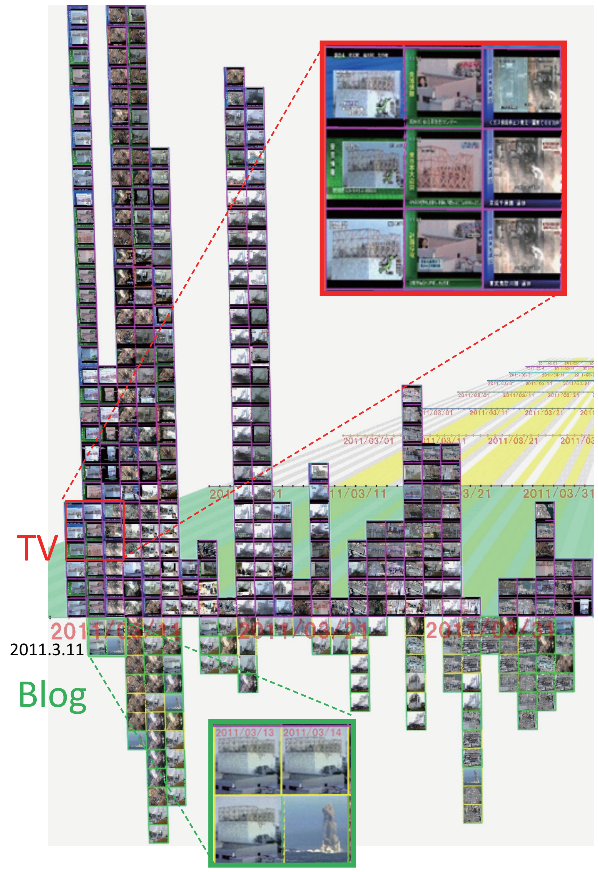


Figure 8: Image clusters related to Fukushima nuclear power plant incident in which images from TV lead images from blogs

Figure 11 shows image clusters related to topics on neighboring countries of Japan, e.g., Korea and North Korea. Figure 11 (a) shows images of an issue at the London Olympics, in which a Korean soccer player showed a message board related to the territorial problem between Japan and Korea. Figure 11 (a) shows a situation in which blogs posted images of the incident before TV (some noise images are included in images from TV). Many bloggers in Japan criticized the player. Afterwards, TV programs started to treat this subject as a serious problem. Figure 11 (b) shows the photos of a North Korean rocket (Japanese believed this rocket was as missile). Image flows burst after the time they announced the schedule. We could find out that this topic mainly became popular on mass media. Examples in Figure 11 may describe the differences in behavior among media for sensitive and political issues.

Figure 12 shows an image cluster related to a topic on Tokyo Skytree[2]. The visualized images in this example are Ukiyo-e[3] painted by Kuniyoshi Utagawa[4] around 1831. We wondered why there were such old pictures on the topic related to this latest style of architecture. From original content on blogs and TV, we found that Kuniyoshi wrote about a similar style of tower. This topic was first discussed on a TV program then spread throughout the Web,

[1]The name comes from a type of flower and used for representing the ideal Japanese woman

[2]Tokyo Skytree is the tallest tower in the world completed on 29 February 2012.

[3]A genre of Japanese woodblock prints and paintings produced between the 17th and 20th centuries.

[4]One of the great masters of the Japanese Ukiyo-e style painting during the Edo era.

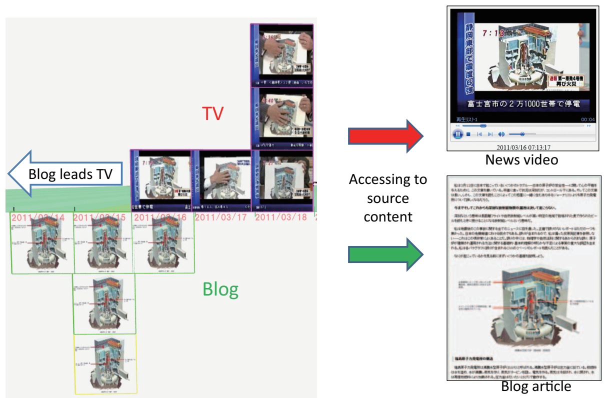


Figure 9: Image cluster related to construction of nuclear power plant in which images from blogs were posted 2 days before TV

as shown in Figure 12. Some blog articles said that bloggers became interested in the topic because the TV program discussed such Ukiyo-e. The images repeatedly appeared on the timeline equally for blogs and TV over one year. Many of the images from blogs appeared just after images appeared on TV. We found such an image flow by using the dynamic query dialog in which the minimum value for cross-correlation was set to 0.19 and minimum and maximum values for leads of days for TV over blog was set to 1 and 5, respectively.

7 Related Work

Recently, there has been many studies on visualizing time-varying data to explore trends.

For time-varying visualization for textual data, researchers have developed a number of approaches based on stacked line chart techniques to visualize changes in topic trends. ThemeRiver [14] provides methods of visualizing changes in the values of multiple attributes on a timeline. TIARA [27] simultaneously combines TheméRiver with tag-clouds to visualize changes in topic keywords. LeadLine [9] provides a flow-like metaphor and arranges flows in parallel to represent topical themes over time.

For visualizing image flows, Gomi et al. [12] visualized images categorized by time, location, and people in life log data. Flake's Pivot [11] provides visualization of magazine cover photos from a particular facet. This method uses a histogram displaying images in each year from the selected facet. Image Depot [20] visualizes image flows from captured data packets from every IP address to check inappropriate Internet use. Dynamic Timelines [21] provides a dynamic 3D visualization framework for interactive presentation of the history of photography. Compared with the above visualization research on temporal image flows, our system enables us to simultaneously compare histograms of images that include a large amount of images, related to multiple aspects.

There have been studies on extracting image clusters using visual features [26, 18, 19], Crandall et al. [6] proposed a method for predicting locations where people take photos from visual, textual, and temporal features. These studies however did not take into account temporal changes in clusters. Our system provides a method for extracting image clusters based on visual features from both images and videos. It also chronologically visualizes extracted clusters to explore changes in societal trends.

There have also been studies on pattern extraction and/or visualization for time series data using cross-correlation and/or time series similarity between different types of media or data-sets [1, 13, 24, 28]. However, there have not been any studies on determining differences and similarities between multiple media based on images. One of the contributions of our framework is that it enables users to interactively explore flows of images interesting to users

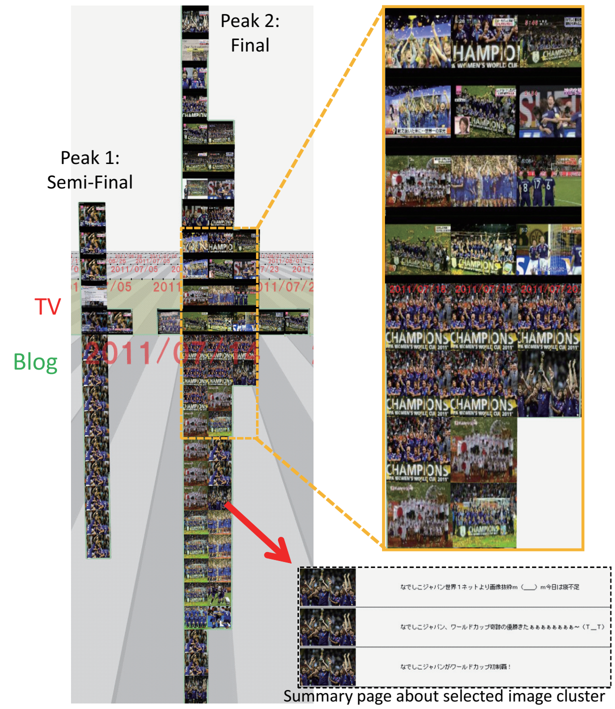


Figure 10: Image flows from blogs and TV having two peaks for Women's World Cup 2011 at same time

that have characteristic feature values such as time-lag and cross-correlation values.

8 Conclusion

In this work, we have proposed a novel interactive visualization and exploration framework for inter-media comparison through image flows. It will be useful for people in the fields of marketing, politics, and sociology to explore image flows from various types of media resources. Investigating the effect of product introduction, events, or sightseeing spots in one medium over other media is important for marketing. The example given in Figure 12 suggests the situation in which people became interested in Kuniyoshi's work and went to art exhibitions after seeing his art on TV. Investigating original visual materials is an important theme for media sociology. Figure 9 shows an example of tracing the original material that we can discover for the first time by using image matching.

Future work will be trying an opposite flow for searching images in which we generate query images from TV and then search for similar images from blogs. This framework can be applied to other media, such as images on Twitter or videos on YouTube, to compare the effects between different types of media resources.

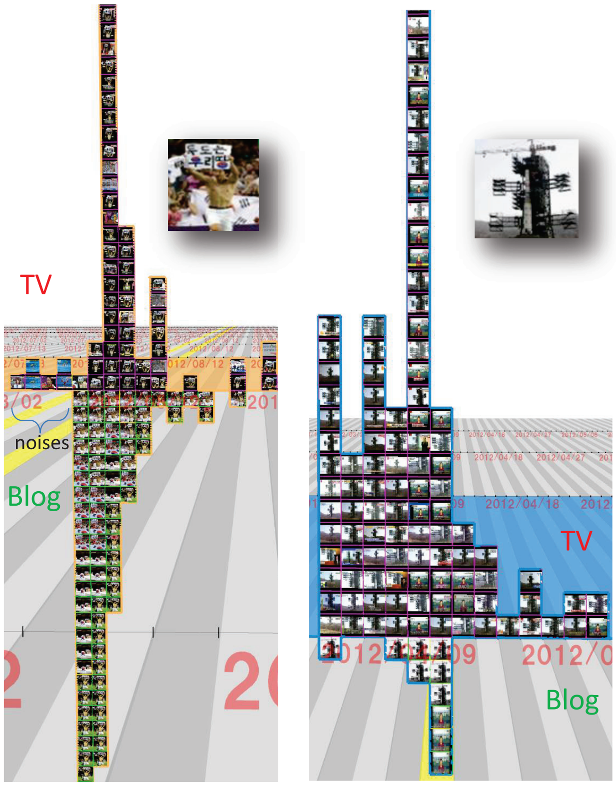
Acknowledgements

We thank Assoc. Prof. Naoki Yoshinaga for constructing the real-time summarization system for image clusters described in Section 5.5. This work was supported by the Multimedia Web Analysis Framework towards Development of Social Analysis Software program of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] E. Adar, D. S. Weld, B. N. Bershad, and S. D. Gribble. Why We Search: Visualizing and Predicting User Behavior. In *Proc. WWW '07*, pages 161–170, 2007.

[2] C. Ahlberg, C. Williamson, and B. Shneiderman. Dynamic Queries for Information Exploration: An Implementation and Evaluation. In *Proc. CHI '92*, pages 619–626, 1992.



(a) Topic on Korea (b) Topic on North Korea

Figure 11: Difference among topics in neighboring countries

Figure 12: Images related to Ukiyo-e by Kuniyoshi Utagawa around 1831

[3] R. Arandjelovic and A. Zisserman. Three Things Everyone Should Know to Improve Object Retrieval. In *Proc. CVPR '12*, pages 2911–2918, 2012.

[4] A. Campbell, C. Wienberg, and A. S. Gordon. Collecting Relevance Feedback on Titles and Photographs in Weblog Posts. In *Proc. of IUI '12*, pages 139–148, 2012.

[5] O. Chum, J. Matas, and J. Kittler. Locally Optimized RANSAC. In *Proc. DAGM Symposium 2003*, pages 236–243, 2003.

[6] D. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the World's Photos. In *Proc. of WWW '09*, pages 761–770, 2009.

[7] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. TextFlow: Towards Better Understanding of Evolving Topics in Text. *IEEE Trans Vis Comput Graph*, 17:2412–2421, 2011.

[8] W. Cui, Y. Wu, S. Liu, F. Wei, M. Zhou, and H. Qu. Context-Preserving, Dynamic Word Cloud Visualization. *IEEE Computer Graphics and Applications*, 30:42–53, 2010.

[9] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. Zhou. LeadLine: Interactive Visual Analysis of Text Data through Event Identification and Exploration. In *Proc. VAST '12*, pages 93–102, 2012.

[10] R. Eccles, T. Kapler, R. Harper, and W. Wright. Stories in GeoTime. In *Proc. VAST '07*, pages 19–26, 2007.

[11] G. Flake. Is Pivot a Turning Point for Web Exploration? In *http://www.ted.com/talks/gary_flake_is_pivot_a_turning_point_for_web_exploration.html (TED.com video - Feb/Mar 2010)*.

[12] A. Gomi and T. Itoh. A Personal Photograph Browser for Life Log Analysis based on Location, Time, and Person. In *Proc. of SAC '11*, pages 1245–1251, 2011.

[13] D. Gruhl, R. Guha, R. Kumar, J. Novak, and A. Tomkins. The Predictive Power of Online Chatter. In *Proc. KDD '05*, pages 78–87, 2005.

[14] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. ThemeRiver: Visualizing Thematic Changes in Large Document Collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):9–20, 2002.

[15] J. Hullman, N. Diakopoulos, and E. Adar. Contextifier: Automatic Generation of Annotated Stock Visualizations. In *Proc. CHI '13*, pages 2707–2716, 2013.

[16] M. Itoh, N. Yoshinaga, M. Toyoda, and M. Kitsuregawa. Analysis and Visualization of Temporal Changes in Bloggers' Activities and Interests. In *Proc. of PVis '12*, pages 57–64, 2012.

[17] K. P. Javed Aslam and D. Rus. The Star Clustering Algorithm. *Journal of Graph Algorithms and Applications*, 8(1):95–129, 2004.

[18] Y. Jing and S. Baluja. Pagerank for Product Image Search. In *Proc. WWW '08*, pages 307–316, 2008.

[19] L. Kennedy and M. Naaman. Generating Diverse and Representative Image Search Results for Landmarks. In *Proc. of WWW '08*, pages 297–306, 2008.

[20] K. Kubota, H. Koike, and M. Yasumura. Image Depot: Research on Image Gathering and Displaying by Packet Capturing. In *Proc. of WISS '09*, pages 167–168, 2009. (in Japanese).

[21] R. L. Kullberg. Dynamic Timelines: Visualizing the History of Photography. In *Proc. CHI '96*, pages 386–387, 1996.

[22] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[23] H. Luo, J. Fan, J. Yang, W. Ribarsky, and S. Satoh. Analyzing Large-Scale News Video Databases to Support Knowledge Visualization and Intuitive Retrieval. In *Proc. VAST '07*, pages 107–114, 2007.

[24] A. Malik, R. Maciejewski, N. Elmqvist, Y. Jang, D. S. Ebert, and W. Huang. A Correlative Analysis Process in a Visual Analytics Environment. In *Proc. VAST '12*, pages 33–42, 2012.

[25] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proc. VL '96*, pages 336–343, 1996.

[26] R. H. van Leuken, L. Garcia, X. Olivares, and R. van Zwol. Visual Diversification of Image Search Results. In *Proc. of WWW '09*, pages 341–350, 2009.

[27] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang. TIARA: A Visual Exploratory Text Analytic System. In *Proc. KDD '10*, pages 153–162, 2010.

[28] J. Yang and J. Leskovec. Patterns of Temporal Variation in Online Media. In *Proc. WSDM '11*, pages 177–186, 2011.

[29] C.-Z. Zhu and S. Satoh. Large Vocabulary Quantization for Searching Instances from Videos. In *Proc. ICMR '12*, pages 52:1–52:8, 2012.