Contents Enforme: Automatic Deformation of Content for Multi-features without Information Loss

Hiroaki Tobita
Sony Computer Science Laboratory Paris
Paris, France
tobby@csl.sony.fr

Abstract—**We introduce a deformation technique that enlarges the feature areas in an image while retaining the information in the non-feature areas. Our main purpose is to provide an effective thumbnail that is useful for practical use in small display devices (e.g., cellular phones, digital cameras, and game devices). Even though many approaches to achieve our purpose have been developed, they are not useful because they require enough time to calculate features. In contrast, our approach can quickly deform image features based on a rapid segmentation technique that our laboratory has proposed. Moreover, our system calculates each segmented area, so it can treat multi-features as deformation elements. As a result, the feature areas were enlarged and the non-feature areas were reduced at almost processing speed, and the total information contained in the original image was retained. Our smooth deformation technique is useful not only for image deformation, but also for a wide variety of contents such as net-meeting and video contents. In this paper, we describe the concept underlying the image enforme technique and its applications.**

Keywords-Contents Deformation; Information Retrieval; Video Compression; Zooming; Thumbnail; Net-meeting.

I. INTRODUCTION

The continuing improvements in computer hardware have led to improved and more compact devices (e.g., cellular phones, personal digital assistants, and digital cameras), thereby making it easier for people to record and carry a huge amount of content, such as images, video files, and audio files. These devices are characterized by small displays with high resolution, so users can easily obtain information and examine the contents in detail. In addition, many content formats have been developed with a focus on effective file size. However, even though there have been considerable advances in device styles and media formats, the methods for previewing contents have remained almost unchanged. For example, to browse a set of images, a user has to look at a set of simple thumbnails or to look at the images individually.

To solve this viewing problem, considerable effort has been made towards providing effective thumbnails. Cropping [1, 2] is an effective way to create a clear thumbnail. The non-feature areas are cropped, leaving only areas with features. The resulting picture is clearer, and the thumbnails produced are useful for browsing images in small devices. Since the calculation involves mainly identifying features, it is very simple. However, cropping the non-feature areas eliminates some of the image's initial information, so the total amount of information is reduced. This makes it difficult to distinguish between similar images. For example, if the original images contain similar objects but different backgrounds, the cropped images will be almost the same, making it difficult to retrieve a desired image.

Image retargeting is also an effective approach to emphasize an image's feature. Setlur et al. [3] developed an image deformation technique that uses a non-photorealistic algorithm. The objective is to provide effective small images by preserving the ability to reconcile important image features, so the quality of the features is an important element. Foreground objects are first separated from the background, and then a new smaller background image is created. After the background area is reduced, the foreground objects are restored to the image. As a result, the blank areas of the image are removed, and an effective small image that has the same-sized features is created. However, this technique focuses on a whole clear object, so the deformation scale is quite limited. Feng et al. [4] also provided a deformation technique that is based on a fisheye view [5]. In a deformed image, a feature area can be enlarged by liner scaling and non-feature areas can be reduced by non-linear scaling. However, this technique can only treat one feature area, so other features are not deformed and will be reduced if an image has multi-features. Most images contain multi-features (e.g., "two people" or "a person and a building"), so this technique has practical limitations. Moreover, these two approaches require enough time to calculate features, so they are impractical for small display devices that do not have enough calculation power.

We have developed a technique that supports effective content viewing [16]. We think that the following two elements are important for image deformation. First, consideration should be given to the relationship between the feature and non-feature areas, as Feng proposed [4]. The feature area is important for finding an image because it becomes the trigger for browsing. The non-feature area is also important because it becomes an element for recognizing similar images. Second, multi-features should be supported, as Setlur proposed [3]. An image generally has more than one feature, so the technique should deform such features clearly. Based on our approach, the deformed image contains advantages outlined by Setlur [3] and Feng [4].

We also think that calculation speed is important for small devices. Our main purpose is to provide an effective thumbnail for image browsing in small devices, so efficient calculation to deform image features is important because the calculation power of a small device is quite limited. The calculation process is divided into finding regions that users

 ISBN: 978-1-61208-177-9

focus on and deforming these regions. As we have also proposed a rapid image segmentation method [7], we used this method to retrieve features rapidly. The deformation depends on the result of image segmentation, so we treated each part of an object (e.g., only a face area or part of a building) as a feature area. To deform multi-features, we used rubber sheet [10] and constrained fisheye [11] techniques, thus, the deformed image had clear multi-features.

As a result, the deformed image had bigger feature areas and smaller non-feature areas than normal images. Although the gradation of the image changed, the total amount of information contained in the image remained the same. These processes were rapidly and automatically performed, so complex user interaction was unnecessary. A unique thumbnail was provided for use with small devices on the basis of setting deformed images.

The next section describes the image enforme overview. Section 3 describes the implementation of the system in detail. In Section 4, we mention the applications of our concept. In Section 5, we discuss advantages and limitations of our system. Section 6 concludes the paper.

II. IMAGE ENFORME OVERVIEW

Figure 1 shows example deformed images, of which the retargeted image (1) had the highest quality. With image retargeting, however, if the foreground object itself is big, the deformed features are of the same size as the input features. Thus, the deformed image is not effective for information retrieval. The fisheye warp image (2) had a clear contrast between feature and non-feature regions. With fisheye warping, however, if an image has multi-features, the system can only recognize one region as a deformation area. Thus, second or third features are no longer useful.

An image produced using our deformation technique is shown in Fig. 1 (3). Using our approach, even part of an object can be a feature area, for example, if the image shows the face and body of a person and only the face area was deformed. Thus, in this example image users could identify the man's face more clearly than they could with the image shown in (1). Also, our approach treated multi-features as deformation areas, so the multi-features were more deformed and clearer in (3) than in (2). As we described, the non-feature areas were reduced and retained in the deformed image. Thus, users could easily understand what the subject was wearing as compared to the cropping method. Our objective is not to obtain excellent image quality but to produce clear features for effective thumbnails.

III. IMPLEMENTATION

As shown in Fig. 2, there are two processes in the deformation step: extracting the feature areas from an image and deforming the feature areas. An image generally contains characteristic objects (faces, buildings, symbols, etc.); so the system retrieves these kinds of objects and then deforms them.

Figure 1. Example images deformed by image retargeting (1), fisheye warping (2), and image enforme (3).

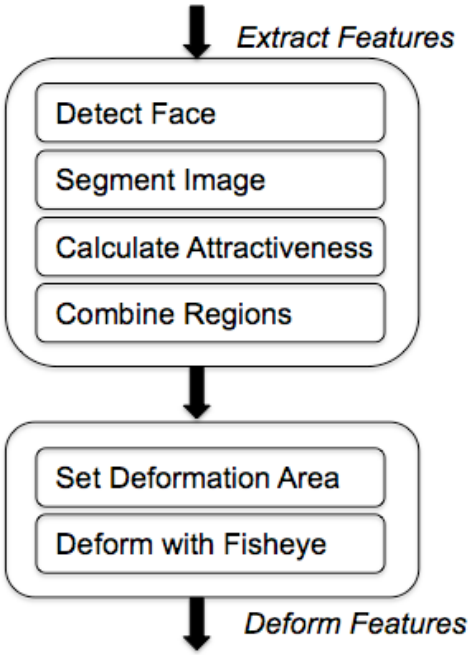

Figure 2. Deformation step is grouped into two processes: extracting feature areas and deforming them.

A. Face Detection

We think that a human face in an image is the most important element, because the image itself would become strange if the system did not recognize a face as a feature, as shown in Figure 1 (2). Thus, a face detection process is first performed on an image. We use the method described by Sabe and Hidai [6] to detect the faces.

B. Image Segmentation

After face detection, the system starts an image segmentation process. For this we use our previously proposed image segmentation method [7], which is a practical and readily available method for segmenting images (Fig. 3 (top)). It segments images into various regions; examples of the segmentation are shown in Fig. 3. After image segmentation, the system deletes the background regions using the method proposed by Tanaka [9]. In this method, the system first calculates the border between a focused region and its neighboring regions, then determines the foreground using the relationship between the length of the border and the length of the contour of the focused region.

 ISBN: 978-1-61208-177-9

Figure 3. Extracting features: System retrieves feature areas through image segmentation (top). If the distance between regions is less than the threshold, regions are merged as a new region (bottom).

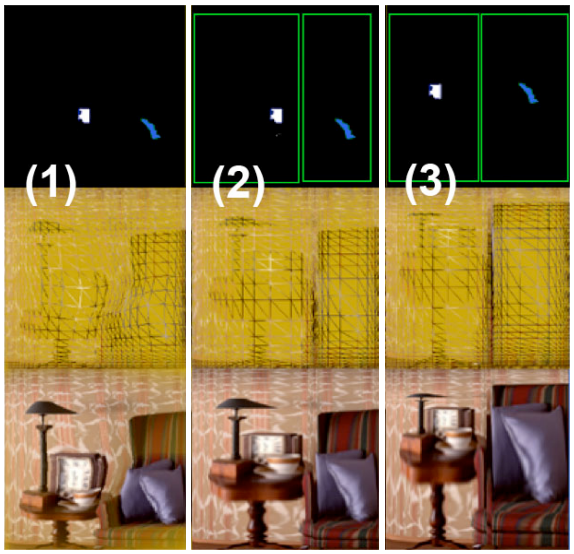

Figure 4. Comparisons of deformations: Images are deformed by a normal fisheye (1) and our multi-fisheye method (2, 3).

C. *Attractiveness*

To identify features, we need to know which region is important. We used Kansei factors [8] to identify highly attractive regions. The Kansei factor for each area was based on simple parameters (e.g., color, size, and position). Based on the calculated Kansei factors, we can determine which regions of a picture will interest viewers. The system calculates this factor for each segmented area and then detects features by comparing each attractiveness value with an attractiveness threshold.

D. *Combining Regions*

If the distance between two regions is shorter than a distance threshold, the regions are combined into one. Figure 3 (bottom) shows an example of combining two regions that exist within the threshold. If the distance between the two regions is longer than the threshold, the regions are treated separately.

E. *Deforming Feature Areas*

We used a multi-fisheye technique to treat multi-features [10, 11]. The image was initially mapped onto a mesh and then deformed by the mesh translation. The system sets multi-viewpoints for zooming on the basis of the desired features. The system calculates the inside of the feature areas linearly and the outside of them non-linearly.

Figure 4 shows simple example of multi-deformation. In Figure 4 (1), there is no borderline to deform multi-features. In Figure 4 (2), the system sets a borderline between two features and deforms each feature. The system also treats features as a kind of node, so it can move the node position freely. In Figure 4 (3), the system first moves features merely by controlling the mesh's UV coordinates (texture mapping parameter) and then deforms them by using the mesh's XY coordinates (position parameter). By the fisheye method, the structure of the mesh is translated based on the relationships between the center of each feature area and a constant force that controls deformation size. We can thus control the size of a feature area by adjusting the force interactively.

With our method, only the feature areas were deformed and were made larger and clearer than in normal images, while the non-feature areas became smaller than in normal images. Thus, feature and non-feature areas can be distinguished more clearly even if an image has multi-features. In addition, the two processes used (identifying features and deforming them) were very simple and were based on almost processing speed for a particular image (256 pixels x 256 pixels) and base mesh (200 polygons) using a normal PC. Since we can adjust the image resolution and mesh flexibly, we can adapt them to the capabilities of the target device.

F. *Results*

Two groups of thumbnails are shown in Fig. 5 (121 images, each 256 x 256 pixels). The original images are shown on the left and the images deformed using the image enforme technique are shown on the right. Although the feature areas in the original images were found using image processing, the characteristic objects (human faces, buildings, symbols, etc.) were successfully identified and deformed. Since most of the original images contained such objects, the subjects on the right are much clearer.

Figure 6 shows some examples of effective and ineffective deformations using the contents enforme technique. When the feature areas were small, such as those shown in (1) and (2), the objects were deformed more effectively. Although the object in (2) was initially difficult to identify due to its small size, the deformation clarified the image. When there were multiple small areas, such as those shown in (3) and (4), the deformation again clarified the objects in the image. Two groups of thumbnails are shown in Fig. 7. Those on the left are 16 normal thumbnails and those on the right are 25 deformed thumbnails. Although the deformed images are smaller, the feature areas are about the same size as in the originals. Two pairs of representative thumbnails are shown in Fig. 8. The objects in the feature areas, a building and two people, are virtually the same size as in the originals. Our technique thus produces images suitable for display on small devices.

 ISBN: 978-1-61208-177-9

Figure 5. Example thumbnails deformed using our technique. Original thumbnails are shown at left and deformed thumbnails are shown at right.

Figure 6. Comparisons of normal images (left) and images deformed using image enforme technique (right).

Figure 7. Comparison of thumbnails of two different sizes. Originals are shown at left and images deformed by using image-enforme technique are shown at right.

Figure 8. Representative thumbnails taken from images in Fig. 8. Although the deformed images are smaller, the feature areas are about the same size as in the originals.

IV. APPLICATIONS

A. *Net-meeting*

We think our deformation technique should be useful for net-meeting systems by making contents more fun and interesting. Net-meeting systems operate using an input image from a camera. The "attendees" are linked by a network and communicate "face to face" (Fig. 9 (top-right)). Net-meeting system users communicate with other users through the camera image and their voice, so users' facial expressions are one of the most important elements. However, conventional net-meeting systems focus on how to achieve an effective network model or how to share information, rather than on interaction techniques. With our deformation technique, net-meeting content could be enhanced. For example, our method recognizes a user's face as a feature area and deforms it automatically to enlarge it (Fig. 9 (bottom, left to right)). After deforming a face area, the user's face becomes slightly comical (big face and small body). Thus, the deformation can make net-meetings more fun and interesting. Moreover, as our system can treat multi-features, more than two face areas are deformed in the same frame. This kind of net-meeting system is suitable for casual use cases, such as communicating with relatives and friends or playing video games.

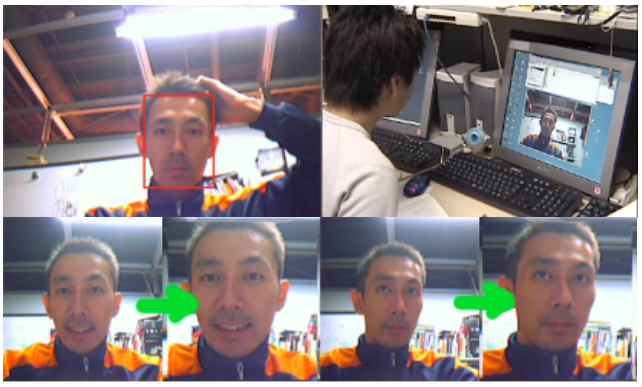

Figure 9. Net-meeting enhancement with contents-enforme technique. In this case, a user's face is deformed and made bigger.

Also, user expressions are enhanced, which should improve communication and clarify who is speaking. In conventional meeting systems, all attendees' faces are the same size, so voice is the only trigger to understand who is speaking. However, by integrating voice information with our deformation, users can clearly understand who the speaker is. Thus, the deformation method is also useful for regular meetings such as business meetings.

B. *Video Contents*

The use of video contents, such as recorded TV programs and home videos, is becoming widespread in everyday life, and many people have to handle large amounts of contents. Thus, it is important for users to understand video contents quickly. Generally, to find out what the video contents are people have to play them one by one manually at high speed. In this case, deformation is an effective way to roughly determine the nature of the contents. There are two different types of deformation for treating video contents: frame deformation and timeline deformation.

As video contents consist of a number of frames set on a time axis, they are based on two elements, the frame data and the time data. Both elements have feature areas, so our technique can be used for video deformation. Since the frame data is deformed in the same way as net-meeting data, we here describe only deformation of the time data. Our

 ISBN: 978-1-61208-177-9

technique can also treat the time data along with the content features and deform the video contents. We focus here on using it to view or compress the contents. The deformation is done by controlling the video play speed on the basis of the frame features and by simple image processing that combines non-linear and linear deformation.

A simple example of timeline deformation is shown in Fig. 10. This example focuses on a soccer game. The actions are reflected by the differences between each frame and the average frame. The histogram at the top shows the differences between each frame and the average frame. The average frame is created first from all the frame data; each frame is then compared with the average one. In a soccer game, one camera catches the overall image of the field and many other cameras closer to the field catch more detailed images. The average frame is the one captured by the main camera. When an exciting moment occurs, such as a free kick or a chance for a goal, other cameras zoom in on the action. The frame data thus dynamically changes and these changes are obtained by comparing the frame caught by each camera with the average frame.

The feature areas are set by controlling the threshold level (Fig. 11 (1)). The play speed is calculated by the fisheye algorithm. Thus, the play speed depends on the distance between an area and a feature area (Fig. 11 (2)). If an area is close to a feature area, the area is played at almost normal speed. On the other hand, if an area is far from a feature area, the area is played at quicker speed. The feature area contents and close areas to the feature are played at normal speed and the non-feature area contents are played quickly. As a result, the total playing time can be as little as one-fourth that of the original video (Fig. 11 (3)).

V. DISCUSSION

A demonstration of our contents enforme technique during a laboratory open house prompted many interesting reactions and comments from the visitors. Although this was not a scientific evaluation, it still provided useful input.

A. Images Enforme and Thumbnails

The visitors could generally identify the deformed images more clearly than the original images. Most visitors could understand the concept of providing partly deformed images for use as thumbnails and the value of retaining reduced non-feature areas. They quickly comprehended the efficiency of doing this by comparing normal and deformed thumbnails (Fig. 5) and different size thumbnails (Fig. 7). People often take several similar pictures with a digital camera, so the zoom capability is useful for browsing among such images. The visitors could easily identify the objects in the deformed image even when it was difficult to identify them in the normal image (Fig. 6). Also, since the features are produced through a combination of non-linear and linear deformation, most of the deformed images, especially the faces, looked quite funny. Thus, these deformed images are useful for comic creation [15].

We think our approach offers two advantages for information retrieval compared to conventional approaches. One is that it supports memory-oriented browsing. Generally, users can remember the most impressive part of the pictures they took with a camera or created using paint tools. The elements often make a strong impression on them and become a trigger for browsing. Although our approach deforms only part of an object, we think it is particularly effective in this regard. The other advantage is that it supports more effective browsing among similar images because the total information is kept the same as for the original image by combining non-liner and liner deformations. Also, if two images are similar, the differences between the deformed images are greater than the differences between the original images. Thus, our approach can enhance differences between similar images.

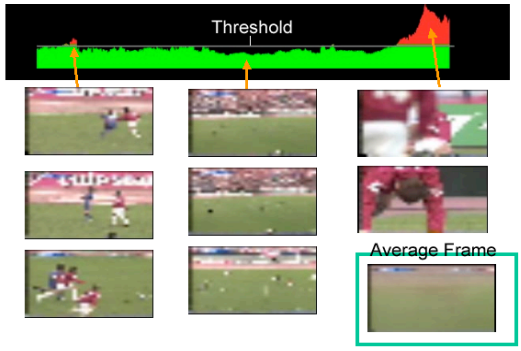

Figure 10. Retrieving feature data. The histogram shows the difference between a frame and the average frame.

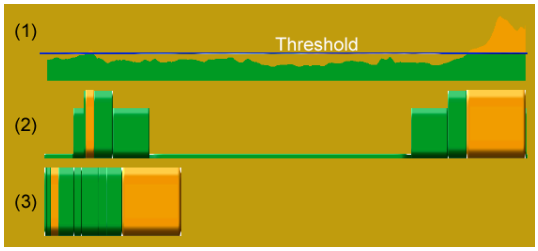

Figure 11. Deformation of video data: the video data is deformed along with the feature area. The feature areas are played at normal speed and the non-feature areas are played more quickly.

Many visitors asked questions about the image enforme calculation and the error rate. Although they could easily imagine it working effectively in small devices, they imagined that much computational power would be needed to deform images. Actually, the image segmentation and fisheye calculation require only moderate computational power. However, since we can adjust the image resolution and mesh flexibly, we can adapt them to the capabilities of the target device. The error rate is related to the size of the feature area and is important because a deformed image is not effective if the original features cannot be retrieved precisely. Since our deformation is based on fisheye calculation, objects close to deformed features are also deformed. Moreover, both the feature and non-feature areas still reside in the deformed image without correct features. In contrast, with cropping, unrecognized feature areas are cropped.

 ISBN: 978-1-61208-177-9

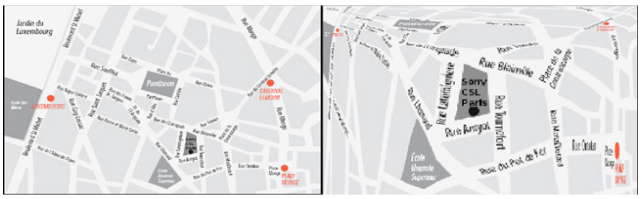

Figure 12. Deformed map: A simple example of a deformed map created by combining the contents-enforme technique with optical character recognition.

Users also suggested many possible applications such as deformed CD jackets and maps. Thumbnails of deformed CD jackets could be used in applications for browsing the contents of music storage devices. Moreover, maps could be deformed to give prominence to areas of interest. A simple example of a deformed map is shown in Figure 12.

B. Video Enforme

Our concept for deformed video contents was well received. Many deformation techniques for video contents are available, especially in the product field [12]. Although they also focus on the feature areas, they create a digest by collecting only the feature areas. While users can quickly view the contents, they can understand only parts of the contents, so they may have trouble understanding the relationships between the whole story and the features. They are thus quite similar to image cropping in that the non-feature areas are discarded. In contrast, our technique focuses on the non-feature areas and maintains the relationships between them and the feature areas. The contents in non-feature areas are played more quickly, enabling users to view the entire video and understand the importance of the contents based on the playback speed. By adjusting the threshold level, a user can set the total playtime.

It is difficult to define the features of TV programs by using one universal method because TV programs cover a very wide range of areas. We demonstrated different types of feature retrieving for three different types of TV programs. We can retrieve some information about a TV program through the Web or TV listings. Sound and face detection are again useful features for deformation. When an exciting moment occurs, such as the chance for a home run or a goal, the announcer's voice generally becomes louder. Thus, a part where the voices are louder often contains an interesting or emotional scene, so voice-based deformation is promising for use with video contents. Moreover, scenes with the user's favorite actor, i.e., someone previously registered to the system by the user, would play at normal speed.

Our deformation focused on the time axis can be used effectively for sound deformation in, for example, a song. Most sound data contains refrain or chorus parts [13]. These parts often include the "hook" of a song and are a kind of feature area. It should be possible to deform sound data based on such features and control the playing speed accordingly. A sound digest should be an effective way to browse music available through the Web because it focuses on feature areas and plays more quickly than the original sound data.

VI. CONCLUSION

We have described our contents enforme technique for deforming and reducing the features of contents (image and video). The technique uses image processing (face detection and image segmentation) and zooming (non-linear and linear). We showed some examples of our deformation and discussed its effectiveness. Since this technique has potential applications in various areas, including video content browsing and net-meeting systems, we mentioned these applications in addition to describing our concepts.

We are planning to improve our technique by combining it with conventional information visualization systems, which focuses on the information layout. Combining our approach with VelvePpath [14]

REFERENCES

[1] S. H. Ling, B. B. Bederson, and D. W. Jacobs, Automatic Thumbnail Cropping and its Effectiveness, *In Proceedings of ACM UIST 2003*, pp. 95–104. 2003.

[2] X. Fan, X. Xie, W. Ma, H. Zhang, and H. Zhou, Visual Attention Based Image Browsing on Mobile Devices, *In Proceedings of ACM Multimedia 2003*, pp. 148–155, 2003.

[3] V. Setlur, S. Takagi, M. Gleicher, and B. Gooch, Automatic Image Retargeting, *In Conference Abstracts and Applications of ACM Siggraph 2004,* pp. 4, 2004.

[4] F. Liu, and M. Gleicher, Automatic Image Retargeting with Fisheye View Warping, *In Proceedings of ACM UIST 2005*, pp. 153–162, 2005.

[5] G. W. Furnas, Generalized fisheye views. *In Proceedings of the ACM Tran. on Computer-Human Interaction* 1, 2, pp. 126–160, 1994.

[6] K. Sabe and K. Hidai. Real-time Multi-view Face Detection using Pixel Difference Feature, Recognition, *In proceedings of SSII 2004*, 2004.

[7] R. Nock and F. Nielssen. Statistical Region Merging, Transactions on Pattern Analysis and Machine Intelligence (TPAMI) IEEE CS Press 4, pp. 557–560, 2004.

[8] S. Tanaka, S. Inoue, M. Ishikawa, and S. Inoue. A Method for Extracting and Analyzing "Kansei" Factors from Pictures, *In Proceedings of IEEE Workshop on Multimedia Signal Processing 1997*, pp. 251–256, 1997.

[9] S. Tanaka, A. Planete, and S. Inoue. A Foreground-Background Segmentation Based on Attractiveness. *In Proceedings of CGIM 1998*, pp. 191–194, 1998.

[10] M. Sarkar, S. S. Snibble, O. J. Tversky, and S. P. Reiss. Stretching the Rubber Sheet: A Metaphor for Viewing Large Layouts on Small Screens, *In Proceedings of ACM UIST 1993*, pp. 81–91, 1993.

[11] L. Bartram, A. Ho, J. Dill and F. Henigman. The Continuous Zoom: A Constrained Fisheye Technique for Viewing and Navigating Large Information, *In Proceedings of ACM UIST 1995*, pp. 207–215, 1995.

[12] Iitokomi
http://prius.hitachi.co.jp/go/prius/pc/2005sep/iitokomi/index.html [retrived: December, 2011]

[13] M. Goto. SmartMusicKIOSK: Music Listening Station with Chorus-Search Function, *In Proceedings of ACM UIST 2003*, pp. 31–40, 2003.

[14] H. Tobita. VelvetPath: A Layout Design System with Sketch Manipulations, *In Proceedings of EuroGraphics 2003 Short Presentation*, pp. 137-144, 2003.

[15] H. Tobita, K. Shibasaki. EnforManga: Interactive Comic Creation Using Drag-and-Drop and Deformation, *In Proceedings of IEEE Multimedia 2009*, pp. 269-274, 2009.

[16] H. Tobita, and F. Nielsen, Image Enforme: Automatic Deformation of Image for Multi-features without Information Loss, Pervasive 2009 (late breaking result), 2009.

 ISBN: 978-1-61208-177-9