3D and Sequential Representations of Spatial Relationships among Photos

Mahoro Anabuki

Canon Development Americas, Inc.

E15-349, 20 Ames Street

Cambridge, MA 02139 USA

mahoro@media.mit.edu

Hiroshi Ishii

MIT Media Laboratory

E15-328, 20 Ames Street

Cambridge, MA 02139 USA

ishii@media.mit.edu

Abstract

This paper proposes automatic representations of spatial relationships among photos for structure analysis and review of a photographic subject. Based on camera tracking, photos are shown in a 3D virtual reality space to represent global spatial relationships. At the same time, the spatial relationships between two of the photos are represented in slide show sequences. This proposal allows people to organize photos quickly in spatial representations with qualitative meaning.

Keywords

3D visualization, slide show, structure analysis, photo organization, photomicrograph, camera tracking

ACM Classification Keywords

H5.1. [Information Interfaces and Presentation]: Multimedia Information Systems—artificial, augmented, and virtual realities.

Introduction

When people take several photos of a subject, they take them from different points of view. For example, a material researcher takes photomicrographs of one material with several magnification levels to analyze its structure. An architectural photographer takes still

photos of a house in every room in order to record its floor plan.

Spatial relationships among such photos have important meaning for structure analysis and review of the photographic subject. However, these relationships are not always clear by viewing the photos. Photomicrographs with totally different magnification levels provide few visual cues to reveal the corresponding areas. Still photos in different rooms rarely include the same object. Instead of taking several photos, creating a video seems better for recording a structure of a subject because adjacent video images of one sequence would include the same visual area. Therefore additional effort is sometimes required for efficient analysis and review of the structure. When use of a video is not convenient, people take just several photos of one subject and make annotations or manual layouts to show the spatial relationships. However, these manual tasks are time-consuming, and the results cannot be easily converted to other formats. They also usually show only qualitative relationships.

In this paper, we propose automatic representations of spatial relationships among photos based on camera tracking. In this proposal, spatial relationships are calculated from camera parameters (position, orientation, and focal length) at the time of shooting. Based on the calculation, various representations can be generated without time-consuming manual tasks. Since multiple representations allow for multi-perspective analysis and review, we propose two simultaneous contrasting representations for one set of photos. One representation is a 3D virtual space that shows global spatial relationships among the photos
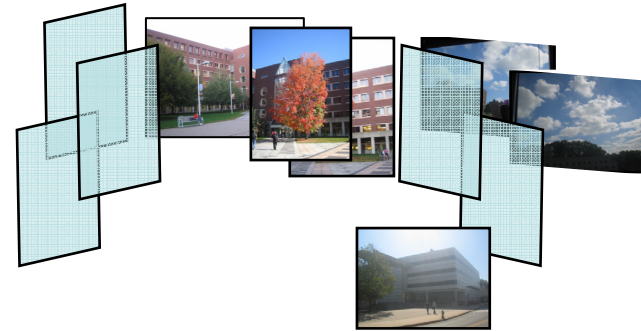

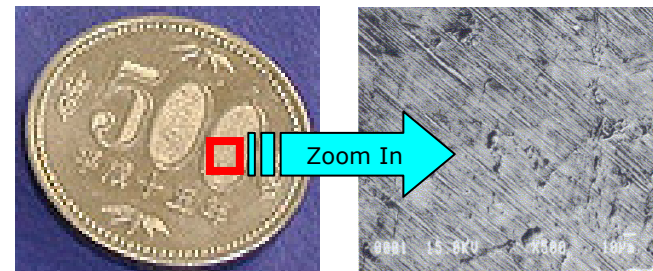**Figure 1.** 3D virtual space representing the global spatial relationship among photos.

Figure 2. Slide show (zoom in) representing the spatial relationship between two photos.

(Figure 1). The other is a slide show sequence of two (or more) photos showing local spatial relationships among the photos (Figure 2). These representations show spatial relationships quantitatively so that they help people to analyze and review the subject in different ways from qualitative manually-created representations. For example, empty spaces in the 3D space can remind people of non-captured areas of a material that should be captured. The slide show will provide a sense of speed and direction, for example, a sense of walking in a house.

Related Work

There are some previous works that represent spatial relationships among photos.

Salient Stills is a technique that merges frame images of a movie into a high resolution still image using optical flow [4]. If the camera is moving, the resulting still image will represent a wider view than the camera image. If an object is moving, a trajectory of the object's motion will be represented in the resulting still image. This technique allows people to see multiple points of view and/or one sequence at a glance, so it has been used for surveillance. *STAMP* is a technique for constructing a pseudo-3D virtual space [3]. With *STAMP*, several photos are shown with corresponding areas overlapping one another. A user can switch a main photo with morphing animations that maintain the overlaps so as to provide pseudo spatial movements, which are helpful for understanding the spatial relationships among photos. This technique has been used for sequential directions on a web site. *Salient Stills* and *STAMP* have a common characteristic, which is that corresponding visual areas are required between photos, while the corresponding areas are not required in our proposal.

Smith et al. proposed an educational application using a digital camera with a global positioning system (GPS) and digital compass to record its position and orientation [2]. A photo taken with this camera is shown with a past photo at the same location so that it allows students to see the historical changes. While they focused on the relationships between photos shot at different times at the same location, we are focusing on the relationships between photos shot at different locations.

System Design

Our proposal can be applied to various types of cameras, such as microscopes and still cameras. This section describes common system design issues.

Camera parameter tracking
For any camera, its parameters (position, orientation, point of focus, etc.) are tracked to calculate spatial relationships among the captured photos. In the case of a microscope, the system tracks the camera position relative to a microscope slide and magnification value. Because the camera (or the slide) might change its position slightly when high powered magnification is used, accurate camera position tracking is required. In the case of a still camera, we assume that the photographic subject is the point of focus of the camera. Therefore, the system tracks the camera position and orientation in real space, as well as its focal length. Because some photos may be taken at locations near each other, accurate camera tracking is needed for still cameras, as well as microscopes. At the same time, wide area camera tracking is also required because some photos may be taken at locations far from one another.

Photo arranging in a 3D virtual space
Captured photos are placed in a 3D virtual space automatically, based on camera tracking. A user can browse global spatial relationships among the photos by walking through the space. In the case of a microscope, one axis of the space is the magnification value. The photograph size in the space is decided based on the magnification value (Figure 3). All photos have the same orientation and show a kind of tree structure. In the case of a still camera, the 3D space
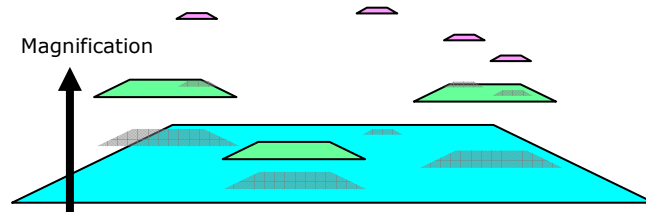
Magnification

Figure 3. An example of the 3D virtual space in the case of microscope.

corresponds to real space. The size of photos is decided according to the field of view.

Selection of slide show effects
When a user selects one photo and then selects another photo, the system represents the spatial relationship between the first one and the second one with a slide show effect. For example, when the second photo is a part of the first one, the second one appears from the corresponding point in the first one with a *zoom in* effect (Figure 2). When the second photo is next to the first one, the second one appears from the side and the first one disappears to the opposite side (Figure 4). The slide show continues if more photos are selected. Below are examples of calculations for selecting a slide show effect in the case of a still

camera: p_1 and p_2 are position vectors and o_1, o_2 are orientation vectors of two photos.

- If $o_1 \approx o_2$ and $o_1 \cdot (p_2 - p_1) \approx 0$ and $p_{1, height} \approx p_{2, height}$, the system assumes that two photos line up side by side and then *slide in from side (left/right)* effect is selected.

- If $p_1 + k\, o_1 \approx p_2$ (k > 0) and $o_1 \cdot o_2 / |o_1||o_2| \approx 1$, the system assumes that the first photo is in front of the second one and then *box out* effect is selected.

Prototype Implementation
For the first prototype, we implemented a system for still cameras. As we mentioned in the system design section, high quality camera tracking in real space is required. Since such tracking technologies have been proposed in the augmented reality (AR) research area [1], we decided to use a camera tracking method developed for AR systems.

We used an AR software development kit (SDK) called "MR Platform SDK" [5] for camera tracking. This SDK allows us to track a camera in real space in real time

Moving to Right

Figure 4. Slide show (moving to right) representing the spatial relationship between two photos.

using a six-degree-of-freedom sensor (POLHEMUS FASTRAK). The sensor has a limited sensing area, but it has enough accuracy to distinguish spatial differences between shooting locations even if they are very close to one another. Because the SDK is designed for easy enhancement by users, other tracking methods, such as GPS, ultrasound sensor, inertia sensor, and computer vision technique can be used with the SDK. We will use these methods for camera tracking in future prototype systems.

We attached a sensor receiver with a button to a CCD camera (WATEC WAT 221S + M96001l). The button is used for taking a photo. Because this camera has a fixed point of focus, we did not implement focus length tracking, using a fixed value as the focus length parameter in this prototype. The camera and the sensor are connected to a Linux PC. When the button is pressed, a photo is taken and sent to the PC. The sensor data is also sent to the PC at the same time. The photo and its position and orientation that are calculated using the sensor data are related to each other.

Our software has four views on a PC display (Figure 5). The top left view shows the real-time camera view. The top right view shows a 3D virtual space that represents the global spatial relationship among the captured photos. The bottom left view shows the captured photo thumbnails in the order in which they were photographed. The bottom right view shows a selected photo. When another photo is selected, this view shows the spatial relationship between both the first and the second selected photos with a slide show effect. The prototype system has six effects. A suitable effect is selected through the flow in Figure 6.
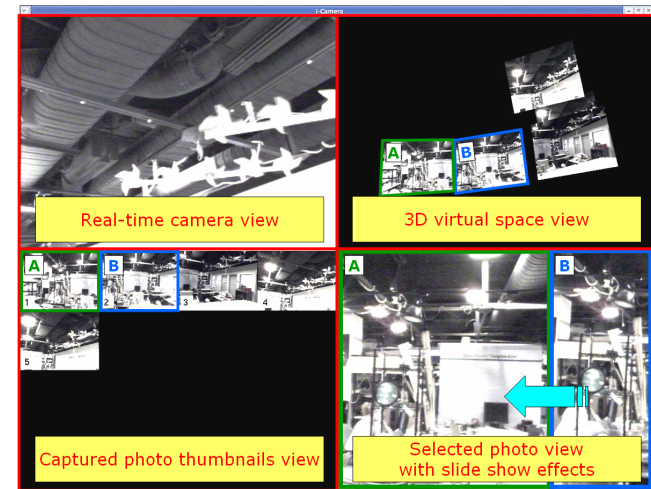

Figure 5. Our software has four views: the real-time camera view, the captured photo thumbnails view, the 3D virtual space view, and the selected photo view with slide show effects.
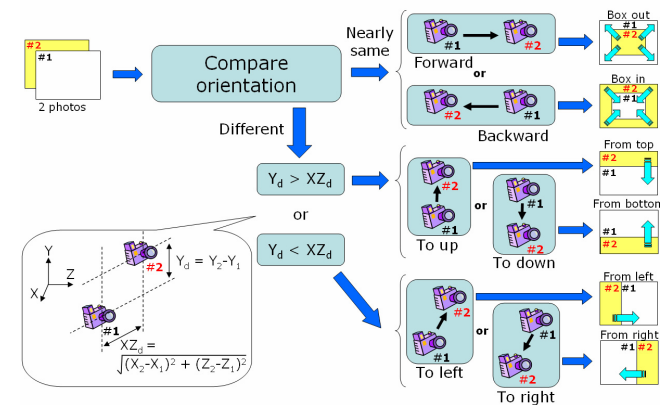

Figure 6. A suitable slide show effect is selected through this flow.

Discussion

The automatic photo indexing using camera parameters would be beneficial to those who organize several photos in spatial representations. Some material and biological researchers are looking forward to using our proposed system as they expect it will reduce the time it takes to make presentation materials with several photomicrographs.

Tight coupling of photo shooting and photo indexing enables real-time representations of spatial relationships among photos. We expect real-time representations to indicate which areas should be captured at the next shooting.

Spatial relationships can be represented in various formats. Each format has its own advantages and disadvantages. Our proposal attempts to enhance the advantages and reduce the disadvantages by computer-aided multi-format data representations.

Our proposed representations include empty areas/time in some cases. When the captured photos are far from one another, our proposed 3D virtual space is sparse. When a user selects two photos that are far from each other, to represent the distance between them the system may show nothing for a while in a sequence. But this emptiness has some meaning because it is based on actual qualitative spatial relationships. We believe this allows people to feel the space beyond the captured photos. We think it is a kind of information extension.

Future Work

As we develop this project, we will expand the first prototype system for still cameras and implement another prototype system for microscopes. It is also our goal to investigate better representations. After implementation, we will evaluate how people interpret the proposed representations, as well as utilize our system. In addition, we will determine if our proposed representations of spatial representation among photos make the effects described in the discussion section.

Acknowledgements

References

[1]   Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. Recent advances in augmented reality. *IEEE Computer Graphics and Applications 21*, 6 (2001), 34-47.

[2]   Smith, B. K., Blankinship, E., Ashford, A., Baker, M., and Hirzel, T. Inquiry with imagery: Historical archive retrieval with digital cameras. *In Proc. MM 1999*, ACM Press (1999), 405-408.

[3]   Spatio-Temporal Association with Multiple Photos (STAMP). http://home.csis.u-tokyo.ac.jp/~tanaka/stamp/stamp.html

[4]   Teodosio, L. and Bender, W. Salient stills. In *ACM Trans. Multimedia Comput. Commun. Appl. 1*, 1 (Feb. 2005), 16-36.

[5]   Uchiyama, S., Takemoto, K., Satoh, K., Yamamoto, H., and Tamura, H., MR Platform: a basic body on which mixed reality applications are built, In *Proc. ISMAR 2002*, IEEE Computer Society Press (2002), 246-253.