

# Locating Key Views for Image Indexing of Spaces

Hongyuan Cai

Indiana University Purdue University Indianapolis  
723 W. Michigan Street  
Indianapolis, IN 46202-5132  
01-317-469-6872

hocai@cs.iupui.edu

Jiang Yu Zheng

Indiana University Purdue University Indianapolis  
723 W. Michigan Street  
Indianapolis, IN 46202-5132  
01-317-278-2365

jzheng@cs.iupui.edu

## ABSTRACT

Image is a dominant medium for visualizing spatial environment and creating virtual access on the Internet. Where to capture images is however subjective and relies on artistic sense of photographers so far. In this paper, we will not only visualize areas with images, but also determine where the most distinct viewpoints should be located. Starting from elevation data of an area, we present spatial and content information in ground based images such that (1) a given number of images can have maximum coverage on informative scenes; (2) a sequence of views can be concatenated with minimum continuity along most-exposed-paths. According to scene visibility, continuity, and data redundancy we evaluate viewpoints numerically with an object-emitting illumination model. Our view exploration may eventually reduce data to archive and transmit, facilitate image acquisition, indexing and interaction, and enhance human perception of spaces. Real images are captured based on our planned key positions to form a visual network to index the area.

## Categories and Subject Descriptors

H.2.4 [Database Management]: Systems - *multimedia database*.  
H.2.8 [Database Management]: Database Applications - *Image databases, spatial databases and GIS*. H.4.3 [Information Systems Applications]: Communications Applications - *Information browsers*. I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism- *Virtual Reality*. I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture - *Imaging, geometry, sampling*.

## General Terms

Algorithms, Management, Measurement, Documentation, Performance, Design, Experimentation, Human Factors.

## Keywords

Spatial visualization, key view, panorama, image indexing, virtual environments, camera placement, visibility

## 1. INTRODUCTION

Although numerous 3D approaches have been developed for spatial visualization, huge numbers of 2D images have been taken

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*MIR '08*, October 30–31, 2008, Vancouver, British Columbia, Canada.

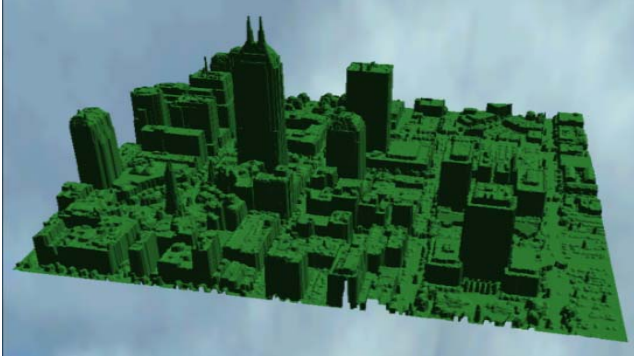
Copyright 2008 ACM 978-1-60558-312-9/08/10...\$5.00.

recent years for increased number of digital cameras, image quality, capacity of storage, and speed for transmission and display. Uploaded community images in many photo service sites have been tagged in maps for exploration of spaces [4,20]. Where to take images is empirical and relies on scenes of interest and artistic sense of users. Taking images pervasively in a space is not always possible for the infinite number of viewpoints and is sometimes unnecessary for redundant scene coverage.

On the other hand, efforts to archive large spaces systematically and completely with images have been made [24,26]. Visual spaces can thus be retrieved pervasively, which has a great value for disaster management, virtual navigation, geo-reference, etc. View selection problem becomes critical in such a system for image acquisition, indexing, and retrieval. We investigate the viewpoint allocation problem in a large area either for archiving scenes or for setting surveillance cameras. *Key views* that can present similar scenes in surrounding regions are effective for new visitors to access the area. Moreover, clickable image regions can be generated for indexing detailed information [22], and dynamic events can be monitored remotely via wireless transmission [27].

Because millions of images can be taken at infinitively dense positions and fine orientations in a high frequency, one has to assess the redundancy of pixels for indexing a space. Images can have similarity and difference in (a) color and signal level, (b) orientation and frame level, (c) position, motion, disparity, and 3D level, (d) visibility, occlusion, and scene coverage level, (e) dynamic object and event level, (f) time, day and night, season level, and (g) semantic, symbolic, and artistic level. Related works include a sea of images proposed by Aliaga et. al. [10] in an indoor environment and data compressed at a signal level; Panoramas [1,2], omni-directional image [8,13], and spherical views [3,16] removed the view difference at orientation and trim level. Along transitional spaces, parallel-perspective route panoramas [18,26], X-slit images [23], and multi-perspective views [19] provide continuous scenes. Snavely et. al. [20] has associated a collection of sightseeing images and recovered the 3D structures for photo tour, which is at position and disparity level. Google Street View has recorded panoramas in major US cities [24], organized at semantic and symbolic level.

This work analyzes the relation of images at visibility level. The key views will be planned to guide camera work so as to achieve a proper data size of image-base and continuity in virtual tour design. For a limited number of images, they should include as many scenes as possible to emphasize spatial layouts for viewer's spatial perception, and facilitate image indexing for flexible interaction as well. We first estimate *view significance* in a large area for key view selection or capturing images at evaluated locations. Such planned views can deliver spatial information



**Figure 1. Elevation maps of an urban area from LiDAR data**

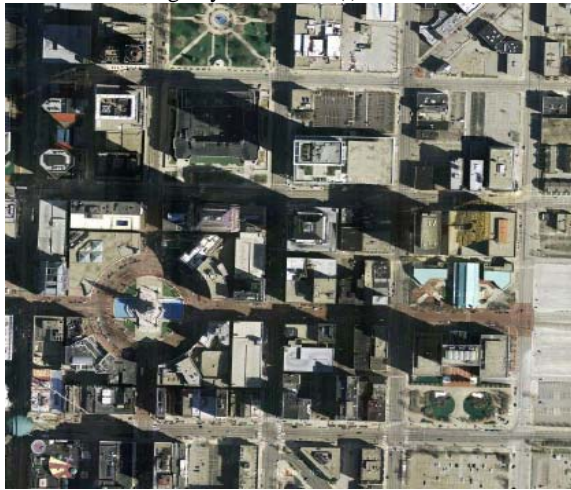
effectively with less redundant data. Our key views are not evenly distributed in spaces for two observations: (1) different viewpoints have different scopes of scene coverage; (2) a shift of viewpoint yields varied motion parallax and scene overlaps due to scene depths. These effects allow us to select less redundant yet more representative views, while keeping a certain degree of continuity among images for navigation.

Different from previous works, our key view planning may answer questions such as where the most distinct viewpoints are and how the most exposed paths can be connected in an area. The criteria for selecting key views are: (i) each view should cover as

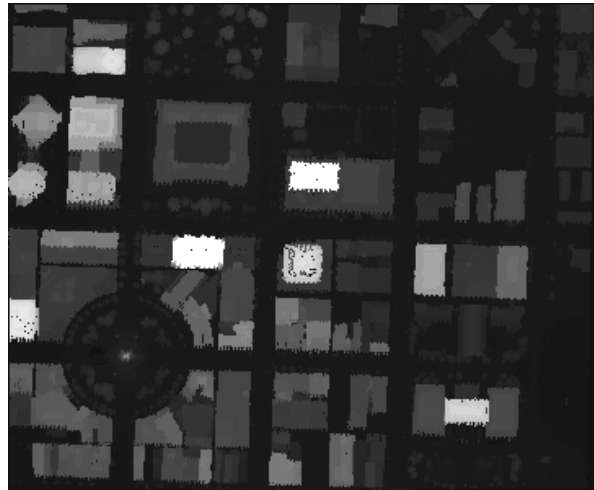
many 3D surfaces as possible such that viewers can understand his/her location by referring to the captured scene layout; (ii) consecutive points along walkthrough paths should share appropriate portions of scenes in order to enhance the perception of spaces in a virtual environment; (iii) the key view selection should reflect spots of interest to facilitate visual indexing and retrieval via clickable regions and embedded links in a multimedia system.

We start from a coarse elevation map [25] obtained from LiDAR data or map services (Figure. 1), and compute a view significance measure at all reachable positions. This measure also takes semantic information into account after specifying weights of importance on 3D surfaces. Based on this local measure, we further investigate global relations of viewpoints to guarantee the novelty of an added viewpoint. Moreover, for any two given positions, we find a path connecting them with the optimal view significance and, then, find viewpoints satisfying a minimum scene overlap requirement. Further, to verify the planned viewpoints, we capture real panoramas to form a virtual traversing system on the web.

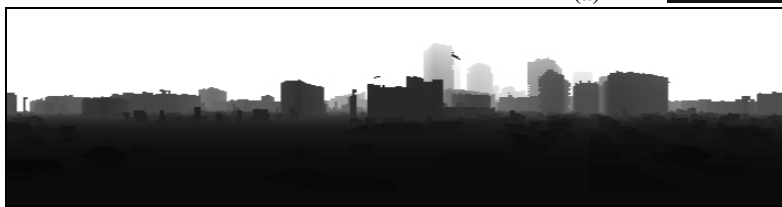
In the following, we discuss the view significance measure in Section 2 and propose an illumination model from planar light sources. Section 3 discusses the key view selection obtaining representative views, which produces discrete viewpoints from infinitely dense field. Section 4 shows the selection of a view



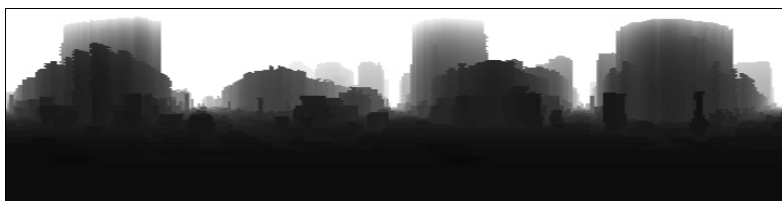
(a)



(b)



(d)



(e)



(c)

**Figure 2. Depth maps of viewpoints generated from LiDAR Data. (a) Satellite image of an area about  $700 \times 600 \text{m}^2$ . (b) LiDAR elevation map of  $360 \times 300$  pixels: intensity represents building height. (c) A hemisphere view at a point. (d,e) Depth of panoramas generated from viewpoints at a street corner and a parking lot. The intensity represents the depth from viewpoint.**

path while keeping a minimum sharing of scenes at consecutive viewpoints. Section 5 introduces a trial to capture real panorama images and the experimental results.

## 2. VIEW EVALUATION

### 2.1 Panorama Images for Scene Archiving

There are many parameters such as location, viewing angle, focal length, and resolution that can be taken into consideration in shooting photographs. Modeling all these parameters and optimizing them have a big order of complexity. In fact, the parameter selection is not only related to the Field of View (FOV) of a camera, but also related to targets of interest. For a particular scene, direction of a photo is more important than shooting distance and position, because distance can be compensated with a zoom lens or high resolution images. Although an aspect view of an object can record shape and color, and highlight details, it usually does not show the real object size and spatial location in its environment, which are important in navigation and virtual tour.

In this paper, we reduce those factors to basic ones – the viewpoint and FOV. We take high resolution panoramas at evaluated locations since (a) the scene continuity in orientations is greatly improved from normal discrete images; (b) images facing various directions can be generated from panoramas; (c) final images with a fixed resolution on the web or a multimedia window can be scaled from panoramas. To obtain panoramas, a fish-eye camera capturing hemispherical images can be used [14]. The images can be converted to cylindrical panoramas or discrete images easily. Other sensors such as omni-directional sensors [13] and LadyBug camera [11] are also good choices for obtaining spherical and cylindrical images.

### 2.2 View Significance from Scene Coverage

The criteria for evaluating a viewpoint at a reachable place at eye-level are to cover large 3D surfaces roughly, and take importance of scenes into consideration. Compared to an overhead image, ground-based-images capture more vertical surfaces and details in an area. Intuitively, a view with large portions at horizon is not as visually significant as a view with full of objects in conveying location information. Similarly, a view with a large sight from an overlook is more significant than a view from a narrow street in telling global location.

Our view significance is based on how large 3D surfaces an image can cover, which is measured in an area (for example, in Figure 2a) using its elevation data (Figure 2b). We compute the area of 3D surfaces covered by a panorama for the view significance. Let  $P(X, Y, Z)$  donate a position in a space and  $H(P)=Y$  is its height in the elevation map. We obtain ground reachable positions. Denoting 3D surfaces visible from  $P$  as  $S_i$ ,  $i=1, 2, 3 \dots m$ , and  $S_i \notin sky$ , we calculate the visible point set from  $P$  as the union of  $S_i$

$$S(P) = \bigcup_{i=1}^m S_i(P) \quad (1)$$

Denote a ray from  $P$  by  $r(\phi, \varphi)$  stretching in orientation  $\phi \in [0, 2\pi]$  and azimuth angle  $\varphi \in [-\pi/2, \pi/2]$  (also see Figure 2c). If  $r$  hits a surface at distance  $D(\phi, \varphi)$ , a sign function,  $\lambda(\phi, \varphi)$ , takes value 1 and otherwise 0. A viewpoint can have a depth map as in Figure 2d. The small area covered by a ray is then  $D(\phi, \varphi)d\phi d\varphi$ . We define view significance  $\sigma(P)$  to be the area of  $S$  accumulated by

$$\sigma(P) = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \lambda(\phi, \varphi) \frac{D(\phi, \varphi)}{D(\phi, \varphi) + D_0} d\phi d\varphi \quad (2)$$

where  $D_0$  is a large constant (e.g., 100m) and the denominator counts for the degradation of image contrast on distant scenes due to atmospheric haze. It discounts a close-to-infinity scene to be integrated into  $\sigma(P)$ .

Figure 3 shows the view significance maps evaluated for structures in the urban area of Figure 2. One can notice the influence from streets, high rises and open spaces. In general,  $\sigma(P)$  is high at a wide site surrounded with rich scenes. The calculation here treats all visible surfaces equally; the result of  $\sigma(P)$  is purely based on shapes and layouts of scenes. Although a spherical view covers entire of high rise even it is placed at bottom of building, the captured view can be severely distorted for its steep viewing angle.

### 2.3 A Lighting Model for View Significances

In presenting a real space, images are usually chosen to cover landmarks, cultural and historic sites, beautiful architecture, decorations, etc. To take these factors into account, we assign different weights to surfaces for the view significance evaluation, in addition to the visibility evaluation. The weights can be assigned manually in the map. As shown in Table 1, monuments, museums, and stations can receive higher weights than storage houses or office buildings. The weight is assigned building-wise in the space, unless some important facades or landmarks need to be emphasized particularly. Alternatively, we can assign high weights on trees and lawns if we are capturing images of green areas or ecology trails. Further, we can assign a river with a high value for setting surveillance cameras to monitor flood. The weight of importance thus incorporates functions and semantic information into the view evaluation and selection framework.

We propose a computing model to modify the view significance measure. Scenes have different irradiances illuminating reachable ground viewpoints. Each building plane is a planar light source with the intensity corresponding to its weight of importance. View significance  $\sigma(P)$  at viewpoint  $P$  is the accumulation of light from all visible points on the scene surfaces.

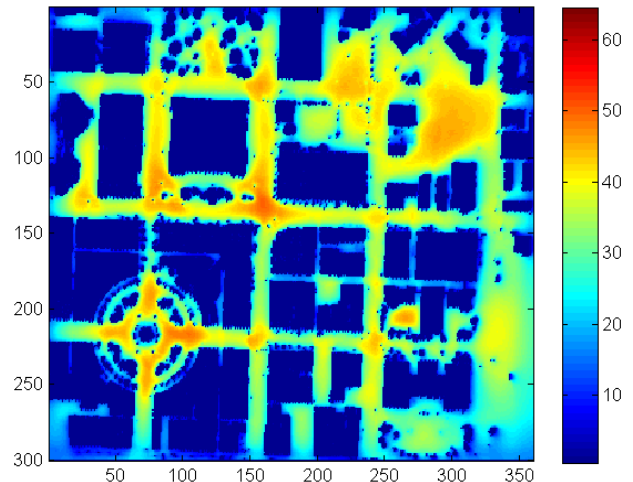


Figure 3. View significance at all positions in an area is displayed in leveled intensity.

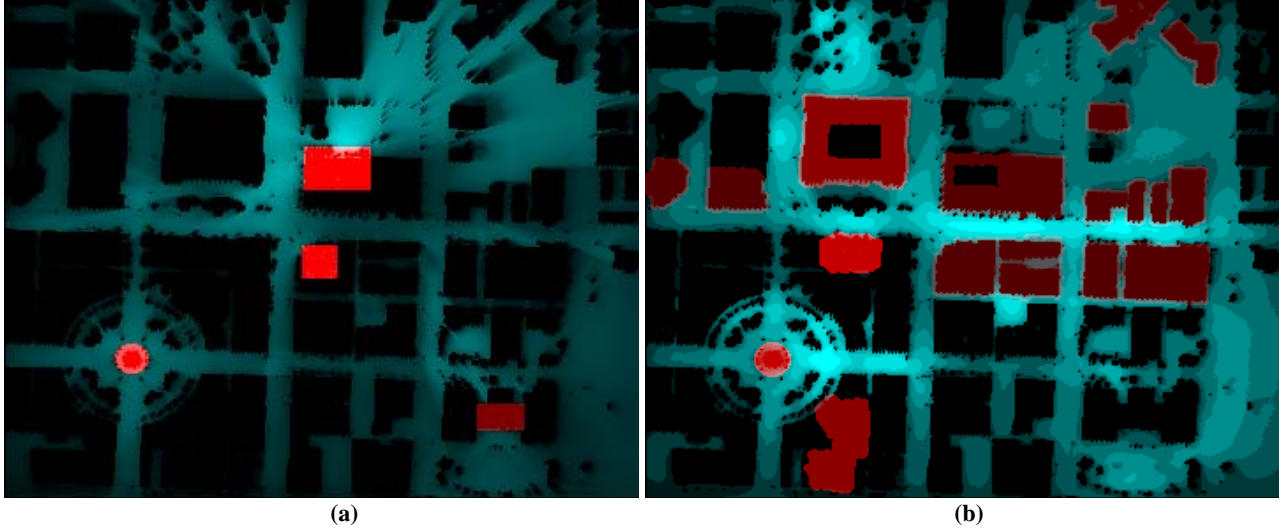


Figure 4. View significances from weighted scenes. Cyan intensity shows  $\sigma$  distribution and red intensity indicates the weight of importance. (a) Several landmarks weighted high and  $\sigma$  distribution is dominated by these strong light sources; (b) Leveled  $\sigma$  distribution from landmarks, public facilities, and a commercial street.

Table 1. An example of assigning weights to scene units.

| Type of scenes  | Weight |
|---|--------|
| Landmarks (monument, highest building in town, etc.)      | 200    |
| Cultural / sightseeing spots (museum, concert hall, etc.) | 150    |
| Commercial buildings (hotel, store, restaurant, etc.)     | 100    |
| Office buildings  | 50     |
| Park, green area  | 30     |
| Ground, road, parking lot                                 | 1      |

Assume building  $i$ ,  $i=1 \dots n$ , has intensity  $w_i$  on all of its surfaces. The computation of  $\sigma(P)$  is then

$$\begin{aligned} \sigma(P) &= \iint_{S(P)} w \lambda(\phi, \varphi) \frac{D(\phi, \varphi)}{D(\phi, \varphi) + D_0} d\phi d\varphi \\ &= \sum_{i=1}^n \iint_{S_i} w_i \lambda(\phi, \varphi) \frac{D(\phi, \varphi)}{D(\phi, \varphi) + D_0} d\phi d\varphi = \sum_{i=1}^n \sigma_i(P) \end{aligned} \quad (3)$$

As shown in Figure 4a, important landmarks with high weights emit strong light to their surroundings. The ground regions in the view significance map are illuminated by such strong “light sources” and even produce “shadows” behind low buildings and trees. Figure 4b shows another example that assigns weights on buildings along a commercial street and sightseeing attractions. The view significance is changed in the resulting distribution.

The computation of  $\sigma(P)$  from different buildings is additive; changing  $w_i$  on building  $i$  only varies  $\sigma_i(P)$ , which can be generated locally and then added to  $\sigma(P)$  distribution. This allows the update of  $\sigma(P)$  field in an inexpensive way.

### 3. KEY VIEW BASED INDEXING

#### 3.1 Selecting Representative Views

For visual access of spaces on the web, two approaches are normally employed. One is traversing spaces across neighbors in a network of images with the same resolution (e.g., map access), and another is in-depth exploration in coarse-to-fine resolution. Different from a spatial map with a fixed resolution and street views with constant intervals [24], our view network is constructed with non-uniformed intervals according to the coverage of views. Our hierarchy of viewpoints is ranked according to the view importance.

The view significance measure can lead to proper allocation of key views for distinct scenes, which can yield an efficient image index. Besides the view significance, another important parameter is the scene overlap between neighboring viewpoints. It can control viewpoint intervals and density view network. Although the motion parallax or disparity is an even finer property between images, we only measure the scene overlap briefly because our static views are largely separated and acquired images are not redundant as a video. A proper amount of scene overlap can

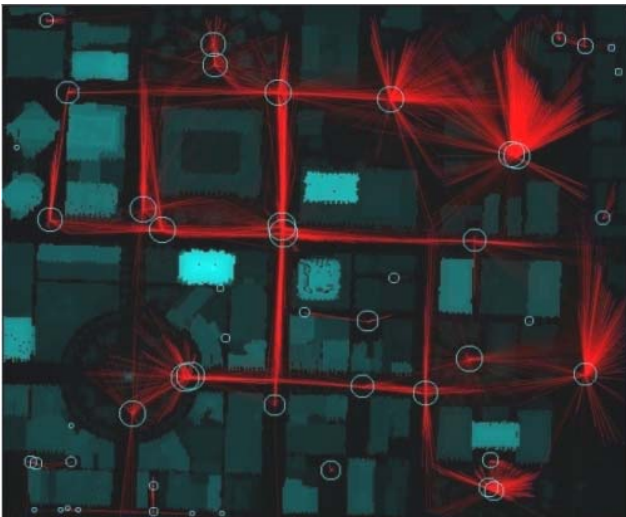


Figure 5. Relation of viewpoints. Selected key views are marked with circles and their radii are proportional to the view significances. Blue points (at ends of red lines) are peak candidates, but not selected as key views because of their large scene overlaps with the selected ones. The amount of overlap is illustrated in red intensity of line connecting peak candidates.

maintain the continuity in traveling the view network and avoid spending data on nearly identical scenes as well.

We select viewpoints  $V=\{P_1, P_2, \dots, P_k, \dots, P_m\}$  for key views as follows. To avoid examining vast number of the candidates and ensure a good visibility, we extract peaks in  $\sigma(P)$  in order of view significance, starting from the maximum  $\sigma(P)$ . As a level  $\delta$  lowers down, more peaks emerge and some nearby peaks may share a large portion of scenes. Therefore, we enforce a new viewpoint to have a scene overlap less than threshold  $\alpha$  with selected ones. Peak  $P_{k+1}$  is selected as a viewpoint if it has more than  $(1-\alpha)\%$  new scenes. Assuming the projected area of  $S$  in panorama at  $P$  is  $A(S)$ , the condition for scene overlapping is then

$$E(P) = \frac{\bigcup_{j=1}^k A(S(P_j) \cap S(P))}{A(S(P))} < \alpha \quad (4)$$

in the panorama.

The computation of overlapping area uses a ray stretching strategy. After ray  $r$  initiated from  $P$  hits a surface, reflecting rays  $r_i, i=1, \dots, k$ , are extended from the surface point towards all viewpoints already selected. If such a ray is blocked by an obstacle, no overlap with the corresponding viewpoint is counted.

One can also choose different values of  $\alpha$  for dispersed viewpoints in a large area or concentrating viewpoints that highlight a small area. Table 2 gives experiments where different  $\alpha$  yield numbers of panoramas. The selected key viewpoints are shown in Figure 5, where peaks and selected key viewpoints are connected with red lines in examining the candidates. The intensity of a red line is proportional to the amount of scene overlap between peaks. Viewpoints occluded by intermediate buildings may have weak links because they can still view common buildings at a distance. A peak connected with a cluster of bright lines can be summarized as a key viewpoint.

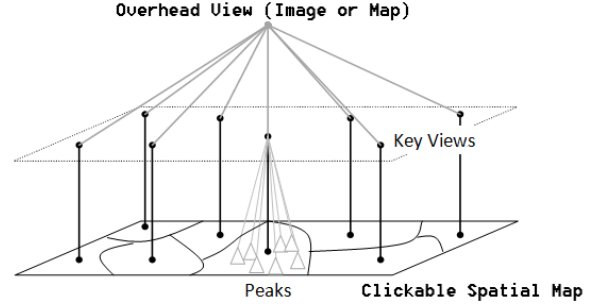
**Table 2. Number of viewpoints selected from 1165 peaks of the view significance for maximum scene overlap allowed.**

| $\alpha$      | 10% | 30% | 50% | 70% | 90% |
|---------------|-----|-----|-----|-----|-----|
| Num of images | 31  | 41  | 50  | 61  | 67  |

### 3.2 Searching Key Views for Scene Display

The framework of visual retrieval of spaces begins with a map clicking (at lower plane in Figure 6). Any points clicked will lead to a peak nearby in the shortest distance. From the peak, the algorithm finds a key view that shares scenes with the peak the most. The amount of overlap has been obtained in Figure 5. The key view is then displayed to briefly represent the scenes at the clicked position.

Because the distribution of view significance is a smooth function at reachable positions, there should always have a peak near a clicked point. The peak point is either a selected key view, or has a large overlap with a selected key viewpoint (for being rejected as a key view). We choose a key view sharing the maximum scene with the peak rather than at closest distance, since an obstacle in between two close viewpoints may create completely different views of them. Because the unique linking from peaks to key views, all the points can be associated to key views. The reachable area thus can be divided into regions represented by corresponding key views. Presenting key views allows further navigating to neighbour spaces or subspaces.



**Figure 6. One possible framework of visual retrieval**

## 4. PATH SELECTION FOR VIRTUAL WALKTHROUGH

In this section, we create a route for visitors to virtually move from one place to another in an area. Different from the representative views above, the views along a path should maintain a certain degree of continuity. We examine view overlaps at successive key views and enforce a minimum scene overlap requirement. The resulting panoramas at discrete viewpoints are still much sparser than a video sequence. We design a real path that exposes to landmarks and attractions as much as possible, and at the same time maintain the minimum number of images for continuity.

We first calculated a path called the *most-exposed path* in the area for sightseeing (similarly, the *most-green path* for recreation). The result can be used to guide video acquisition, data compression, and real tour design. We employ the *shortest path algorithm* to extend paths from a starting point to every position in the map by minimizing objective function

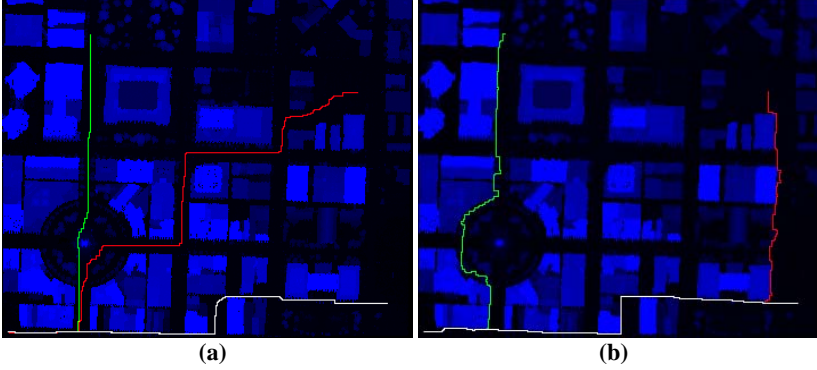
$$Dis = \min_s \left\{ \int_s \text{cost}(P) ds \right\} \quad (5)$$

among all possible paths  $s$ . The cost at each position is calculated as  $\max(\sigma) - \sigma(P)$ , roughly corresponding to sum of sky and ground areas in the panorama. Paths extend to every reachable position and a path field is generated.

Figure 7 shows an example of paths from a starting point of  $H(X, Z)$  to three other positions. Figure 7a are the paths mostly exposed to or viewed from scenes (e.g., from building windows). One can notice the paths through middle of streets along which a panoramic video can be taken. In addition, we also calculate the most hidden path for security to avoid being noticed from large areas. The shortest path algorithm for such a path calculation uses  $\sigma(P)$  as the cost to expend path segments. Besides a detoured route as shown in Figure 7b, the most hidden paths stretch along building corners towards the same destinations as in Figure 7a.

For a virtual walkthrough along a path with discrete images, the physical distance between viewpoints is not really a factor to consider, whereas the scene continuity in the image and the number of images become important. Given a viewpoint  $P_j$ , we can lineout a region,  $R_\beta(P_j)$ , in the map, where all points  $P$  in it share at least  $\beta$  percent of scenes with  $P_j$ , i.e.,

$$\frac{A(S(P) \cap S(P_j))}{A(S(P))} > \beta, \quad \text{if } P \in R_\beta(P_j) \quad (6)$$



**Figure 7. Optimal paths from a starting position (lower left corner) to three destinations, respectively. (a) The most exposed paths. (b) The most hidden paths.**

As shown in Figure 8, such regions with minimum scene overlaps are associated to two designated viewpoints. Yellow regions have more than 50% scene overlaps with red positions. The additional green regions guarantee 30%. It is obvious that a shift of viewpoint in an open area has less view changes than in a narrow street.

To assign a small number of viewpoints along a path while maintaining the continuity between images, we combine the most exposed path and the scene overlap condition. From each selected viewpoint, the algorithm searches along the path and picks the farthest location that satisfies the overlapping constraint as next viewpoint. These selected viewpoints lead to minimum number of images on the path with non-uniformed intervals.

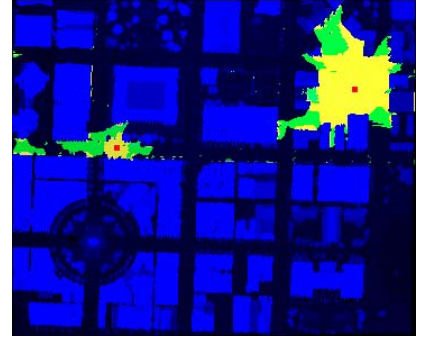
## 5. EXPERIMENTS

The cost computation of the methods is in estimating depth and view overlaps by stretching rays. However, the total time is about ten several minutes with current PC. Although modeling buildings with patches may reduce the ray computation, no depth is provided in such method so that the overlap computation will be difficult. On the other hand, the illuminating model with planar light sources on buildings can help the view significance computation with graphics hardware.

Because the algorithms work on close-to-raw sensor data, there is no surface patch needed to be formed. We calculate a continuous  $\sigma(P)$  field using elevation map  $H(X,Z)$ . LiDAR data are reduced in resolution to such a map first. At each small grid region (e.g.,  $5m^2$ ), non-zero elevation points are median-filtered to yield an integer value in metrics  $H(X,Z)$ . Second, all reachable points at eye-level are marked, if  $Y>0$ . Third, we compute rays in all orientations from  $P(X,Y,Z)$ , until they hit obstacles. The front tip  $P_f(X_f, Y_f, Z_f)$  of the ray satisfies  $Y_f < H(X_f, Z_f)$  as it stretches out. The distance is then determined for the view significance.

To compute view overlaps, the algorithm emits a hemisphere of rays from each peak, and each ray is advanced in a fixed interval. When hitting a surface point, rays are further bounced towards the previously selected viewpoints to measure the visibility from them.

The algorithm is implemented in local areas. The view significance is related to the viewing distance in terms of  $1/(D+D_0)$ . Beyond a range of 700m~1km, the visibility in the image decays significantly and the scenes are ignored in our view significance evaluation. The algorithm thus can be extended to a larger area in the same order of complexity. In a real lighting



**Figure 8. Viewpoints that have scene overlaps more than  $\beta\%$  with designated positions (red dots) are collected as regions.**

model, the irradiance is deducted by  $1/D^2$  in quadric manner as the distance increases. The image sharpness has a larger decay with respect to the viewing distance. Thus, key views may be planned at close ranges to scenes in order to gain resolution, depending on whether images are for location guidance or for arts and architecture appreciation.

Real key views are taken as panoramas at the planned positions to constructing a visual network of the area as shown in Figure 9. The panoramic images are consistent with the predicted depth images from  $H(X,Z)$  except on trees and vehicles, while the real images have much higher resolutions than depth images. The key views are selected from ranked peaks in the view significance distribution. The scene overlap of each key view with previously selected ones are calculated and displayed in red. Because of the large coverage of scenes in the key panoramas, many regions can be embedded into them for clicking and linking. A visual index can be established from a key panorama to detailed spaces and neighboring spaces.

As shown in Table 2, the change of overlap threshold does not affect the number of images dramatically in our results; meaning that the distinct views are stable and a significant reduction of images have been achieved. Users can cut off the number of key views accordingly.

In planning views along a path for virtual navigation, Figure 10 shows the results of viewpoints on one of the most exposed paths in Figure 7, with the minimum overlaps of 10% and 30% respectively. The selected views are more effective than those views arbitrarily taken or taken with a fixed interval. We can notice that the path has sparse viewpoints at wide and open areas and dense viewpoints at narrow streets.

## 6. DISCUSSION

Our key view planning is different from taking artistic photos based on appearances, which is hard to parameterize numerically. The key views selection emphasizes visibility, spatial layout, and functionality of scenes. The employed panoramas also deliver continuous spatial information as compared to discrete photos with unique orientations. Therefore, discrete photos should not be evaluated at the same level as panoramas for the space perception and navigation purposes. Even though, this approach does not prevent adding images on special spots of interest. On the other hand, our key view selection yields a much smaller data set than a video, and sea of images at pervasively distributed viewpoints. Those approaches

are compressed or not at a signal level based on the motion parallax or spatial frequency. The criteria are much tighter than the scene overlap constraint. For virtual space perception with

static images, the view coverage and scene overlap are one reasonable level to enforce the view continuity.

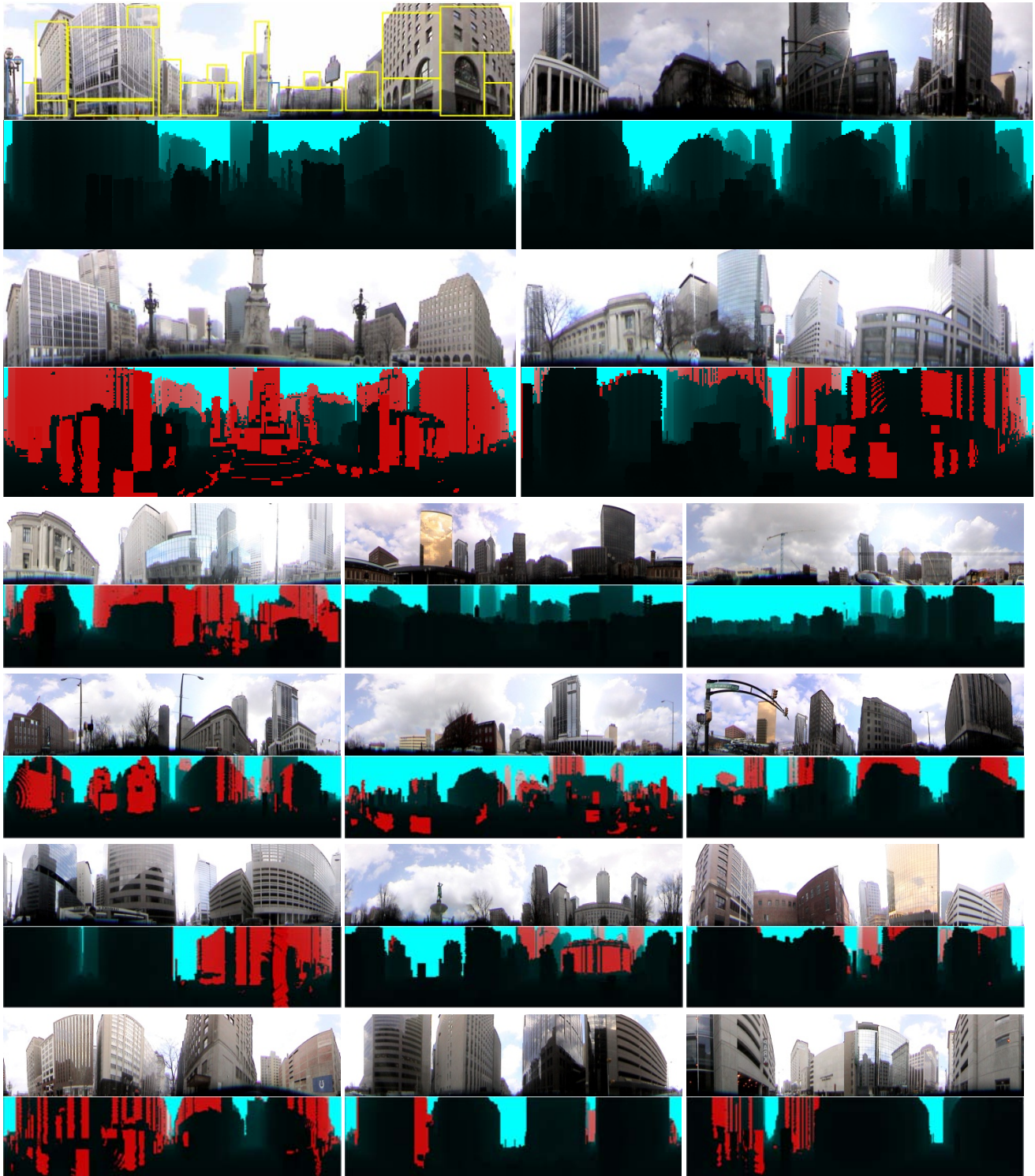


Figure 9. Calculated 16 key panoramas in depth and captured real panoramas in color. The overlapped parts with other selected key views are indicated in red in the depth maps. The real panoramas have vertical FOV  $\varphi \in [-5^\circ, 60^\circ]$  to avoid major ground area, cars and people. Clickable regions are embedded in the key views for visual indexing (blue regions for transition to other panoramas and yellow regions to websites of buildings and facilities).

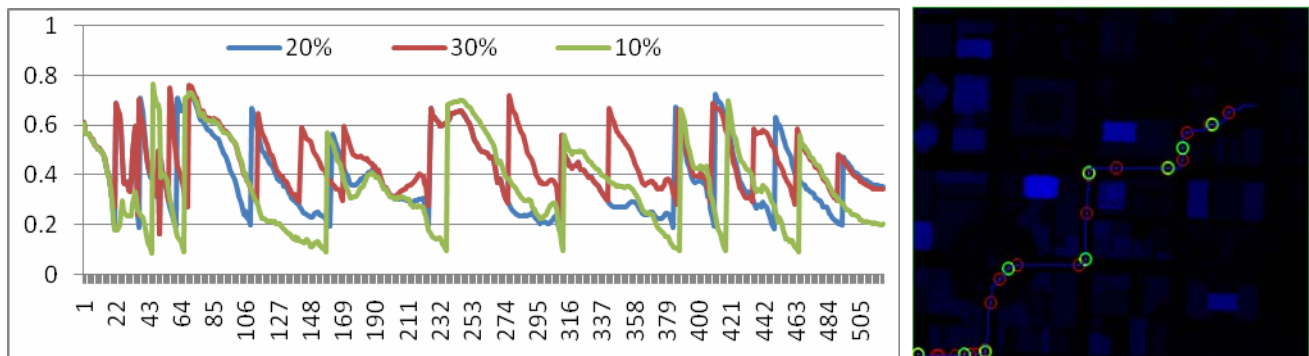
Compared to satellite images, our planned key views provide a brief synopsis of an area observable from the ground. Also, increasing the weight on a scene can attract viewpoints to closer positions. Extending from LiDAR data, the framework here can also be applied to web based graphics [27] to show high quality 3D data on the web. The key views evaluated in this work can help to cache frequently viewed images for transmission directly.

## 7. CONCLUSION

In this paper, we have explored a view planning scheme for scene indexing and visualization. Key views are selected according to appearance of scenes calculated from elevation data of a large area. This allows us to find representative views with the maximum scene coverage or maintain continuity in a view network and along a path for walkthrough. We model the scene visibility, functions, and distinctiveness numerically. Then view overlaps at significant positions are examined for locating key views or minimum number of images along the path. The results may guide image acquisition for spatial perception, help camera setting for surveillance, and improve image indexing for virtual tour.

## 8. REFERENCES

- [1] J. Y. Zheng, S. Tsuji, Panoramic Representation for route recognition by a mobile robot. *Int. J. Computer Vision*, Vol.9, no.1, 55-76, (1992)
- [2] S. E. Chen, L. Williams, "Quicktime VR: An image-based approach to virtual environment navigation", *SIGGRAPH95*, pp. 29-38, 1995.
- [3] R. Szeliski, H. Y. Shum, "Creating full view panoramic image mosaics and environment maps", *SIGGRAPH97*, pp. 251-258, 1997.
- [4] L. Kennedy, M. Naaman, S. Ahern, R. Nair, T. Rattenbury, How flickr helps us make sense of the world: context and content in community-contributed media collections, *ACM Multimedia 07*, 631-640, 2007.
- [5] S. Coorg, M. Master, S. Teller, Acquisition of a large pose-mosaic dataset, *IEEE CVPR98*, 23-25, 1998.
- [6] S. Teller. Toward urban model acquisition from geoLocated images. In *Proc. of Pacific Graphics'98*, pages 45-52, 1998.
- [7] S. Coorg S. Teller. Extracting textured vertical facades from controlled close-range imagery. *IEEE CVPR*, pages 625--632, 1999.
- [8] S. K. Nayar, V. Peri, Folded catadioptric cameras. *IEEE CVPR*, Vol. II, pp. 217-223, 1999.
- [9] S. K. Nayar, A. Karmarkar, 360 × 360 mosaics. *IEEE CVPR* vol. II, pp. 388-395, 2000.
- [10] D. G. Aliaga, T. Funkhouser, D. Yanovsky, and I. Carlbom. Sea of images. In *Proc. of IEEE Visualization*, pp. 331-338, 2002.
- [11] <http://www.ptgrey.com/products/legacy.asp>
- [12] D. G. Aliaga, T. Funkhouser, D. Yanovsky, I. Carlbom, "Sea of images," *IEEE Computer Graphics and Applications*, vol. 23, no.6, pp. 22 - 30, 2003.
- [13] Y. Yagi, M. Yachida: Real-Time Omnidirectional Image Sensors. *International Journal of Computer Vision* 58(3): 173-207 (2004)
- [14] S. Li, M. Nakano, N. Chiba, Acquisition of spherical image by fish-eye conversion lens, *IEEE Virtual Reality 2004*, pp. 235-236.
- [15] A. Roman, G. Garg, M. Levoy, Interactive design of multi-perspective images for visualizing urban landscapes, *IEEE Visualization*, 2004.
- [16] M. Uyttendaele, *et al.*, Image-based interactive exploration of real-world environments. *IEEE CGA*, 24(3), 2004
- [17] Y. Yagi, K. Imai, K. Tsuji, M. Yachida: Iconic Memory-Based Omnidirectional Route Panorama Navigation. *IEEE PAMI* 27(1): 78-87 (2005)
- [18] J. Y. Zheng, Y. Zhou, P. Mili, Scanning Scene Tunnel for City Traversing, *IEEE Transaction on Visualization and Computer Graphics*, Vol. 12, no. 2, 155-167, 2006.
- [19] A. Agarwala, *et. al.* Photographing long scenes with multi-viewpoint panoramas, *SIGGRAPH*, 853-861, 2006.
- [20] N. Snavely, S. M. Seitz, R. Szeliski, Photo tourism: exploring photo collections in 3D. *ACM Trans. Graph.* 25(3): 835-846 (2006)
- [21] T. N. Thanh, *et al.*, Robust and real-time rotation estimation of compound omni-directional sensor, *IEEE Int. Conf. Robotics and Automation*, pp. 4226-4231, 2007.
- [22] B. Russell, A. Torralba, K. Murphy, W. Freeman, LabelMe: a database and web-based tool for image annotation, *IJCV*, vol. 77, no. 1-3, pp. 157-173, 2008.
- [23] A. Rav-Acha, G. Engel, S. Peleg, Minimal Aspect Distortion (MAD) Mosaicing of Long Scenes, *IJCV* vol. 78, no. 2-3, pp. 187-206, 2008.
- [24] Google StreetView: <http://maps.google.com/>
- [25] J. Hu, S. You, U. Neumann: Integrating LiDAR, aerial image and ground images for complete urban building modeling. *3DPVT 2006*, 184-191.
- [26] J. Y. Zheng, X. Wang, Pervasive Views: Area exploration and guidance using extended image media, *ACM Multimedia 05*, 986-995, 2005.
- [27] B Chen, A Kaufman, Q Tang, Image-Based Rendering of Surfaces from Volume Data, *Volume Graphics*, pp. 279-295, 2001.



**Figure 10. Planned consecutive panoramas satisfying the minimum scene overlap along a path. Left: Overlap changes along the path based on 10%, 20% and 30% constraints. New viewpoints are added at where the overlaps reach the minimum requirements. Right: Selected viewpoints with 10% (9 green positions) and 30% (19 red positions) scene overlaps, respectively.**