Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Hongyuan Cai

Entitled

Locating Key Views for Image Indexing of Spaces

For the degree of Master of Science

Is approved by the final examining committee:

Jiang Yu Zheng
Chair

Shiaofen Fang

Mihran Tuceryan

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Jiang Yu Zheng

Approved by: Shiaofen Fang 4/16/2010
Head of the Graduate Program Date

Graduate School Form 20
(Revised 1/10)

PURDUE UNIVERSITY
GRADUATE SCHOOL

Research Integrity and Copyright Disclaimer

Title of Thesis/Dissertation:
Locating Key Views for Image Indexing of Spaces

For the degree of Master of Science

I certify that in the preparation of this thesis, I have observed the provisions of *Purdue University Teaching, Research, and Outreach Policy on Research Misconduct (VIII.3.1)*, October 1, 2008.*

Further, I certify that this work is free of plagiarism and all materials appearing in this thesis/dissertation have been properly quoted and attributed.

I certify that all copyrighted material incorporated into this thesis/dissertation is in compliance with the United States' copyright law and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of the law. I agree to indemnify and save harmless Purdue University from any and all claims that may be asserted or that may arise from any copyright violation.

Hongyuan Cai
Printed Name and Signature of Candidate

4/16/2010
Date (month/day/year)

*Located at http://www.purdue.edu/policies/pages/teach_res_outreach/viii_3_1.html

LOCATING KEY VIEWS

FOR IMAGE INDEXING OF SPACES

A Thesis

Submitted to the Faculty

of

Purdue University

by

Hongyuan Cai

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

May 2010

Purdue University

Indianapolis, Indiana

ACKNOWLEDGMENTS

This study and graduate thesis could not be completed without the kind attention and careful guidance from my advisor Dr. Jiang Yu Zheng. His serious scientific attitude, the spririt of rigorous scholarship and his work style of always pursuing improvement deeply influenced and inspired me. From the topics selection to the final completion of the thesis, Dr. Jiang Yu Zheng was always careful to give me guidance and tireless support. Over these three years, Dr. Jiang Yu Zheng not only carefully guided me on academic field, but also took great caHre and be concerned about my life here, here I would like to extend my sincere thanks and great respect to my advisor Dr. Jiang Yu Zheng. I also need to thank him for providing me the opportunity to study in the US, which extended my vision of the world and will benefit all my life. I would remember his kindness to me in my heart.

Here, I would also like to thank my committee members Dr. Shiaofen Fang and Dr. Mihran Tuceryan, for their academic assistance and expertise. It is precisely because of their help and support; I can overcome the difficulties one by one until the successful completion of the paper.

Thesis will be completed in time, now I feel unable to calm down. From the beginning of this thesis subject to the successful completion, many honorable teachers, classmates, friends and department staff gave me speechless help, here please accept my sincere thanks! Finally, I would also like to thank my family who educated me to be strong in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Cai, Hongyuan M.S., Purdue University, May 2010. Locating Key Views for Image Indexing of Spaces . Major Professor: Jiang Yu Zheng.

Image is a dominant medium among video, 3D model, and other media for visualizing environment and creating virtual access on the Internet. The location of image capture is, however, subjective and has relied on the esthetic sense of photographers up until this point. In this paper, we will not only visualize areas with images, but also propose a general framework to determine where the most distinct viewpoints should be located. Starting from elevation data, 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 set of key views can be selected with certain continuity for representing the most distinct views. According to the scene visibility, continuity, and data redundancy, we evaluate viewpoints numerically with an object-emitting illumination model. Our key view exploration may eventually reduce the visual data to transmit, facilitate image acquisition, indexing and interaction, and enhance perception of spaces. Real sample images are captured based on planned positions to form a visual network to index the area.

1 INTRODUCTION

Although numerous 3D approaches have been developed for spatial visualization, huge numbers of 2D images have been taken recent years for increased number of digital cameras, image quality, capacity of storage, and speed for transmission and display rate. Community images uploaded in many photo service sites have been tagged in maps for exploration of spaces [4,20]. However, where to take landmark images is empirical and it relies on scenes of interest and the esthetic sense of photographers. In addition, efforts to archive large spaces systematically and completely with images have also been made [24,26]. Visual spaces can thus be retrieved pervasively, which has a great value for disaster management, virtual navigation, Geo-referencing, etc. However, setting observation points pervasively in a space is not always possible for the infinite number of viewpoints and is sometimes unnecessary for redundant scene coverage between images to be transmitted and displayed. The view selection problem thus becomes critical for image acquisition, indexing, and retrieval. Our views are not distributed in spaces at an equal interval for two important observations: (1) different viewpoints have different scopes of scene coverage; (2) a shift of viewpoint yields different motion parallax and overlaps in the images on the scenes with different depths. These effects indicate that the total scene coverage is not simply proportional to the number of viewpoints, which allows us to select less redundant yet more representative views.

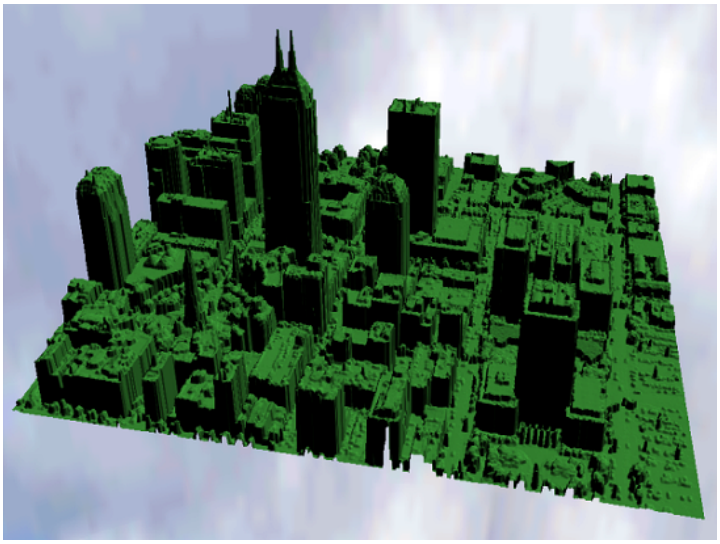
We investigate the viewpoint allocation problem in a large area for archiving scenes and setting live surveillance cameras. We select *key views* that can effectively present similar scenes in surrounding regions for virtual access to the area. The key views should be planned systematically to guide camera work so as to achieve a proper data size of image-base in designing a virtual tour and setting surveillance cameras. In a graphics interface, more clickable regions thus can be embedded in the images

for indexing detailed information [22]; users can click or select a part of the image to "zoom" or "go" there. Moreover, if cameras are mounted there, dynamic events can be monitored remotely via wireless transmission.

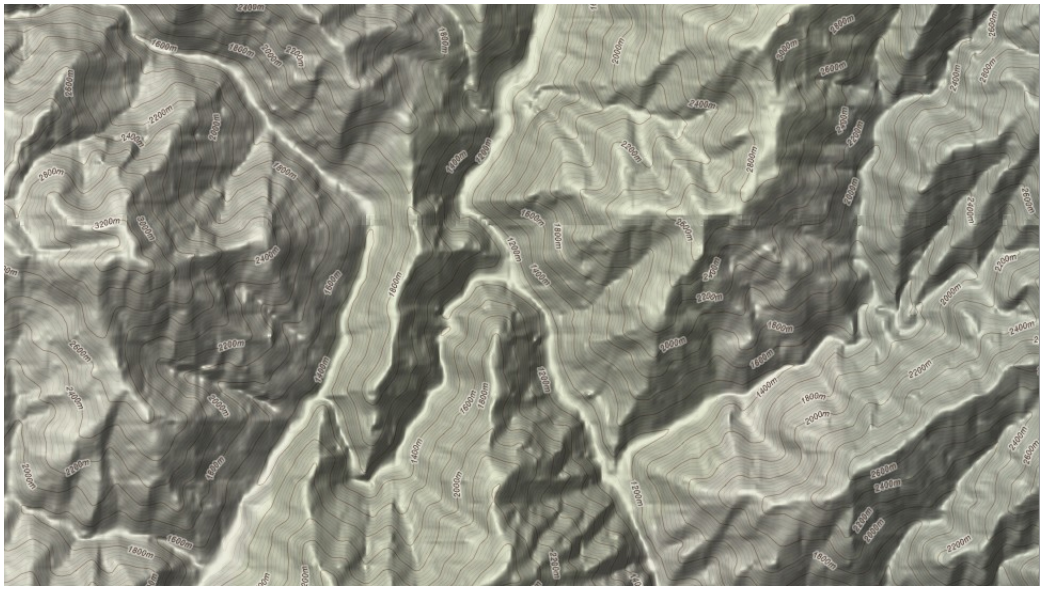
Because millions of images may be taken at infinitively dense positions and fine orientations in an area, it is necessary to assess the redundancy of pixels in presenting spaces. Images taken on the same scene can have similarity and difference at many levels, such as (a) *signal* level described by color [12] and camera parameters, (b) *event* level that scenes are affected by dynamic objects, (c) *temporal* level where scenes vary in time, day and night, season, (d) *FOV* (*field of view*) level in terms of camera orientation and image frame, (e) *viewpoint* level that generates motion, disparity, and varied 2D shape, (f) *visibility* level that varies in scene coverage due to occlusion, and (g) *semantic* level with functional, symbolic and artistic descriptions. This work analyzes the relation of images at visibility level so as to achieve a proper data size of image-base and the continuity in a virtual tour. We estimate *view significance* in a large area and select *key views* for distinct scenes. Such planned views can deliver spatial information effectively with less redundant data. Related works include a sea of images proposed by Aliaga et al. [10] in an indoor environment and data compressed at a signal level. In establishing an urban model [5,6,7], viewpoints were selected more for the purpose of texture mapping and rendering than for data reduction. Other approaches also tackled environments with various image projections and sensors: panoramas [1,2], omni-directional images [8,13,17], and spherical views [3,9,16] well fit a virtual environment framework and uniform the variation in FOV or camera direction. Along transitional spaces, parallel-perspective route panoramas [15,18,26], X-slit images [23], and multi-perspective views [19] provide continuous scenes of linear paths, which are viewpoint invariant. Snavely et al. [20] has associated a collection of sightseeing images and recovered the 3D structures for photo tour, which is at viewpoint and disparity level. Google Streetview has recorded panoramas in major US cities [24], organized by map coordinates at semantic and symbolic level. Different from previous work, our key view planning may find where

the most distinct viewpoints are. The criteria for selecting key views are: (i) each view should cover as large 3D surfaces as possible such that viewers can understand more on his/her location by referring to the scene layout; (ii) the key view selection should reflect spots of interest to facilitate visual indexing of objects via clickable regions and embedded links in the images.

We start from a coarse elevation map [25] obtained from LiDAR data or map services (Figure 1.1), and compute a view significance measure at all reachable positions, e.g., streets of an urban area. This measure also takes semantic information into consideration if weights of importance are specified on 3D scenes. Based on this local measure, we further investigate global relations of viewpoints to guarantee the novelty of added key views. 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 introduces a trial to capture real panorama images in the experiments, and concludes with a discussion of these results.



(a)



(b)

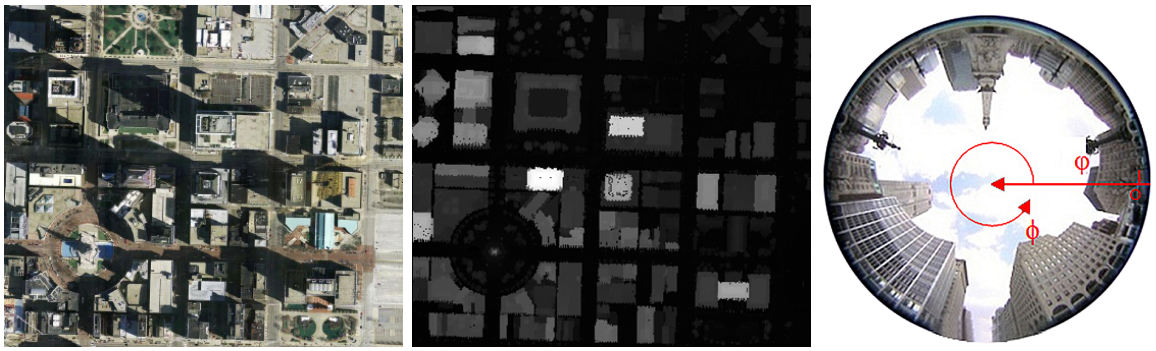
Figure 1.1.: Elevation maps of an urban area (a) from LiDAR data and a town struck by a recent earthquake (b), where wireless surveillance cameras were set for monitoring collapse of mountains and runoff of a dangerous standby reservoir.

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 has a high 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, the direction of a photo is more important than the shooting distance and position, because the distance can be compensated with a zoom lens or high image resolution. Although a local perspective image can record shapes and colors and can highlight details, it usually does not show real sizes and spatial locations of objects in the environment, which makes virtual navigation much more difficult than real navigation.

In this paper, we simplify the factors for image acquisition to basic ones - the viewpoint and FOV. High resolution panoramas are taken at evaluated locations instead since the panoramas achieve the scene continuity in camera orientation as compared to normal discrete images. Low-resolution discrete images facing various directions can be generated in a multimedia window from these panoramas. To obtain panoramas, fish-eye cameras 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 cameras [11] are also good choices for obtaining spherical and cylindrical images.

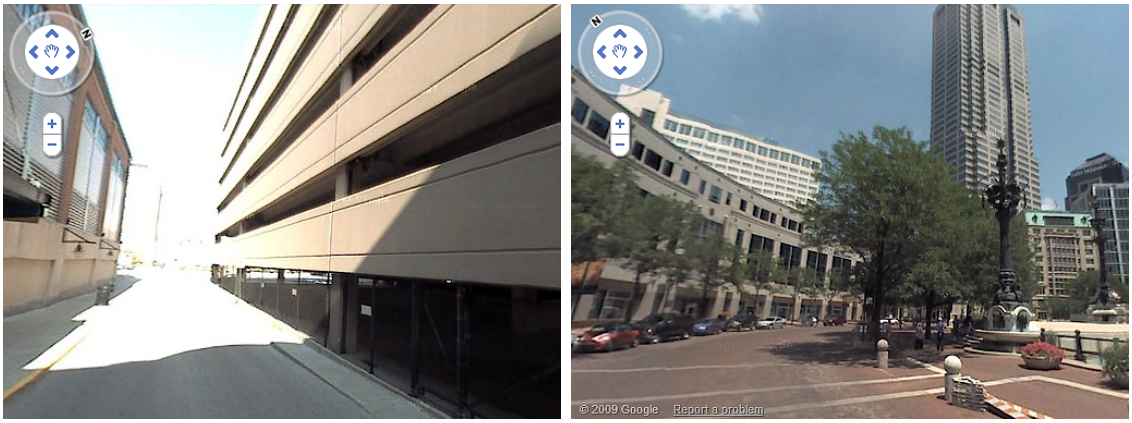


(a) (b) (c)

(d)

(e)

Figure 2.1.: Depth maps of viewpoints generated from LiDAR Data. (a) Satellite image of an area about $600 \times 700m^2$. (b) LiDAR elevation map of 360×300 pixels: intensity represents building height. (c) A hemisphere view at a point, where $\phi \in [0°, 360°]$ and $\varphi \in [-5°, 90°]$. (d,e) Depth of panoramas generated from viewpoints at a street corner and a parking lot. The intensity represents the depth from a viewpoint.



(a) (b)

Figure 2.2.: Two real images captured from street view of google map. The first one is in a narrow alley, while the second one is in a large square seeing more scenes for location understanding.

2.2 View Significance Based on Scene Coverage

The criteria for evaluating a viewpoint at a reachable place at eye-level are to cover large 3D surfaces roughly. Compared to an overhead image, ground-based images capture more vertical surfaces and details in an area. Intuitively, a view with horizon only is not as visually significant as a view with full of objects in conveying location information. Similarly, as shown in Figure 2.2, a view with a large sight from an overlook is more significant than a view from a narrow street in ascertaining one's global location.

Our view significance is based on how large 3D surfaces an image can cover, which is measured at every reachable position in an area (for example, in Figure 2.1a) using its elevation data (Figure 2.1b). 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. Denoting 3D surface patches visible from P as $S_i, i = 1, 2, 3...m$, and $S_i \notin sky$, we calculate the visible point set from P as the union of S_i

$$S(P) = \bigcup_{i=1}^{m} S_i(P) \tag{2.1}$$

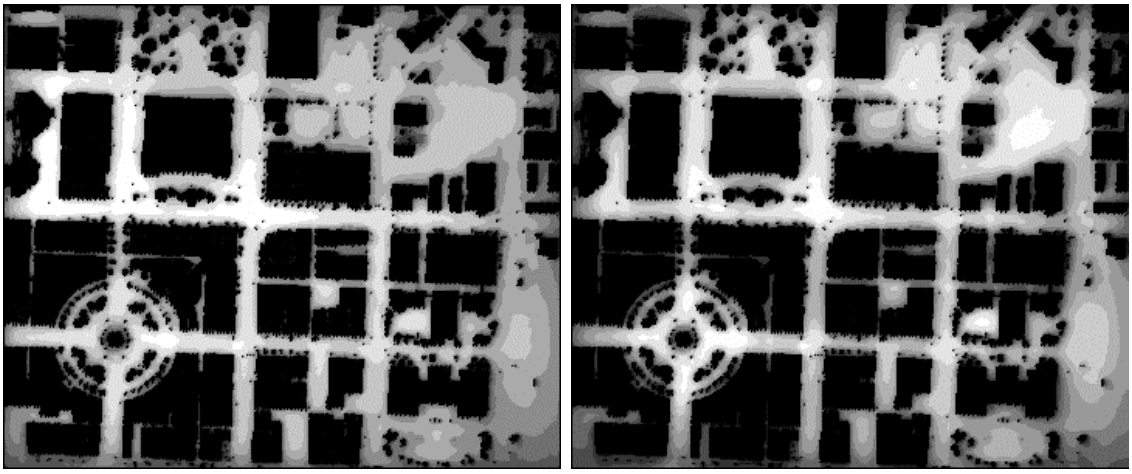
Their formed 3D area is accumulated for the view significance $\sigma(P)$.

To compute $\sigma(P)$, a ray $r(\phi, \varphi)$ is stretched from P in orientation $\phi \in [0, 2\pi]$ and azimuth angle $\varphi \in [\varphi_0, \varphi_1]$ (also see Figure 2.1c). If r hits a surface at distance $D(\phi, \varphi)$, a 0-1 function, $\lambda(\phi, \varphi)$, takes value 1 and otherwise 0. A viewpoint thus can have a depth map as in Figure 2.1d,e. The small 3D area covered by a ray is then $D^2(\phi, \varphi)cos\varphi d\phi d\varphi$. In a real lighting model, the irradiance of a ray is deducted by $1/D^2$ in quadric manner as the viewing distance increases. The image sharpness has a larger decay on distance scenes. We can thus model the degradation of the image contrast on a ray by $D^2(\phi, \varphi) \cos\varphi d\phi d\varphi / (D^2(\phi, \varphi) + C)$, where C is a constant. To balance the accuracy and computational cost, we simplify the formula by replacing D^2 with D. As shown in Figure 2.4, we can obtain similar results by toggling constants C and D_0 properly. In that sense, We define view significance $\sigma(P)$ to be the area of S accumulated by

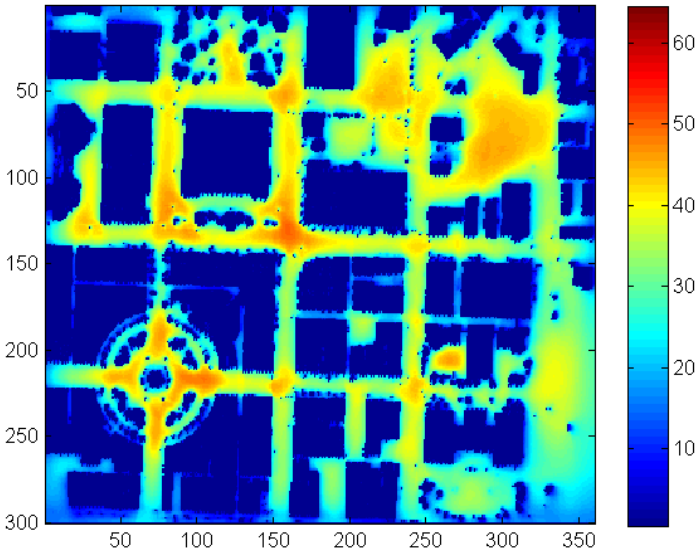
$$\sigma(P) = \int_0^{2\pi} \int_{\varphi_0}^{\varphi_1} \lambda(\phi, \varphi) \frac{D(\phi, \varphi)}{D(\phi, \varphi) + D_0} \cos\varphi d\phi d\varphi \tag{2.2}$$

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 $\sigma(P)$. φ_0 is a lower bound of φ and $\varphi_1$is an upper bound ($-5^\circ \leq \varphi_0 < \varphi_1 \leq 90^\circ$). The calculation here treats all visible surfaces equally; the result of $\sigma(P)$ is purely based on shapes and layouts of scenes.

Figure 2.3 shows the view significance fields evaluated from structures in the urban area of Figure 2.1. One can notice the influence from streets, high-rise buildings and open spaces. In general, $\sigma(P)$ is high at a wide site surrounded with rich scenes. We can also notice that, in $\sigma(P)$ field from cylindrical views (Figure 2.3b), local maxima are at locations slightly away from buildings as compared to that in $\sigma(P)$ field from spherical views (Figure 2.3a). This is because a cylindrical panorama with a limited vertical FOV may cut off high buildings as the viewpoint moves too close to buildings.



(a) (b)



(c)

Figure 2.3.: View significances at all possible positions in an area are displayed in leveled intensity (with contours more distinguishable than original distribution). The brighter the intensity, the higher the significance value is. Black regions are unreachable locations occupied by buildings and trees. (a) Reachable positions and their view significances evaluated with spherical views. (b) View significance evaluated with cylindrical panoramas. (c) Color display of (b) for the continuous distribution of view significance.

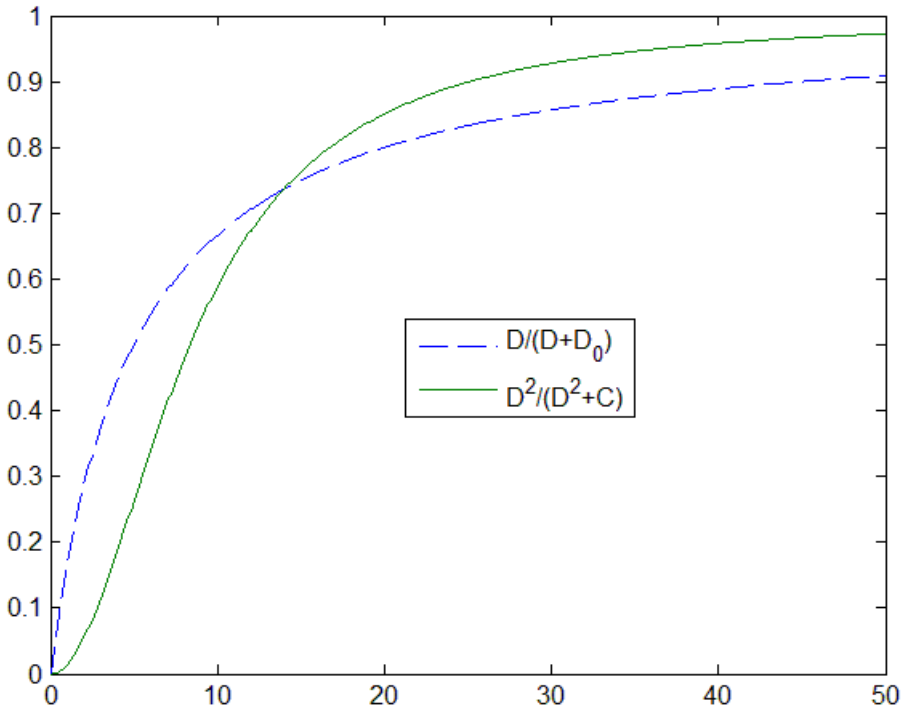


Figure 2.4.: Curves of two formulas in calculating view significance. In presenting a real space, images are usually chosen to cover landmarks, cultural and historic sites, beautiful architecture, decorations, etc. To take these factors into account, we assign different weights to 3D surfaces for the view significance evaluation, in addition to the visibility evaluation. The weights can be assigned manually in the map.

Table 2.1: 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

Although a spherical view covers an entire 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.

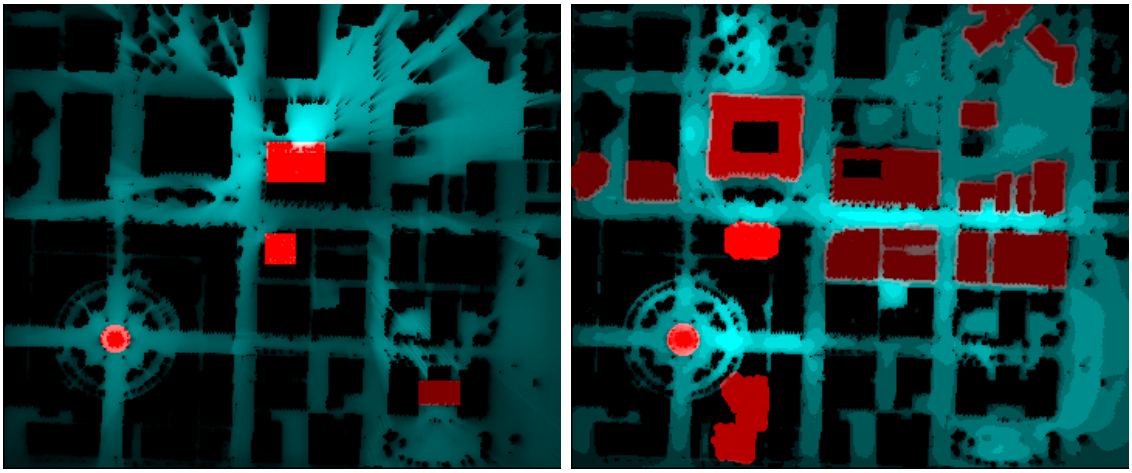
As shown in Table 2.1, monuments, museums, and stations can receive higher weights than storage houses or office buildings. The weight is assigned building-wise in the space, unless some important facades or landmarks need to be emphasized particularly. Alternatively, we can assign high weights on trees and lawns if we are capturing images to archive (record) green areas or ecology trails. Further, we can assign a river with a high value for setting surveillance cameras to monitor flood (Figure 1.1b). The weight of importance thus incorporates scene functions and semantic information into the view evaluation and selection framework.

We propose a computing model to modify the view significance measure by giving the following setting. Scenes have different irradiances that can illuminate ground-based viewpoints. Each building plane can be taken as 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.

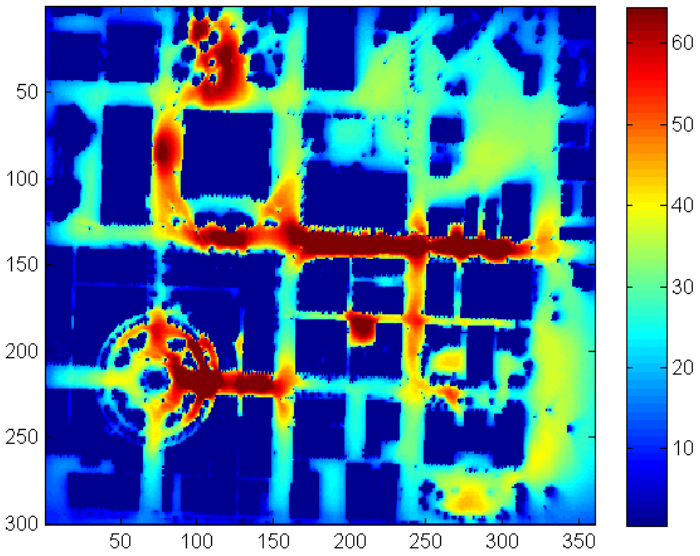
Assume building $i, i = 1...n$, has illuminant intensity w_i on all of its surfaces. The computation of is then

$$\begin{aligned}\sigma(P) &= \iint_{S(P)} w_{S(P)}\lambda(\phi,\varphi)\frac{D(\phi,\varphi)}{D(\phi,\varphi)+D_0}\cos\varphi d\phi d\varphi \\ &= \sum_{i=1}^{n}\iint_{S_i} w_{S_i}\lambda(\phi,\varphi)\frac{D(\phi,\varphi)}{D(\phi,\varphi)+D_0}\cos\varphi d\phi d\varphi = \sum_{i=1}^{n}\sigma_i(P) \end{aligned} \tag{2.3}$$

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 the $\sigma(P)$ distribution. This allows the update of $\sigma(P)$ field locally in an inexpensive way. As shown in Figure 2.5a, 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"; lower buildings and trees even produce "shadows" behind. Figure 2.5b shows another example that assigns weights to sight-seeing attractions and buildings along a commercial street. The view significance, then, is changed in the resulting distribution. This approach provides the possibility of treating scenes differently in a computational frame for view planning. The weight assignment is flexible and interactive, and the resulting view significance is a continuous function of the weight.



(a) (b)



(c)

Figure 2.5.: View significances from weighted scenes. Cyan intensity shows distribution and red intensity indicates the weight of importance. (a) Several landmarks weighted high and view significance distribution is dominated by these strong light sources; (b) Leveled distribution of view significance calculated from landmarks, public facilities, and a commercial street. (c) Pseudo color display of the distribution in (b).

3 KEY VIEW BASED SPATIAL INDEXING

3.1 Selecting Representative Views

For visual access of spaces on the web, two approaches are routinely employed. One is traversing across neighboring spaces in a network of linked images with a similar resolution (e.g., map access), and another is in-depth exploration with coarse-to-fine resolution. Different from a spatial map with a fixed resolution and Google Streetviews with a constant interval [24], our view network is constructed with non-uniformed intervals according to the scene coverage of views. Our viewpoints are ranked according to the view significance to form a hierarchy. A set of *key views* will be extracted as representative views of many similar ones available at nearby positions, while keep roughly the equivalent scene coverage as those distributed views with a constant interval.

The view significance measure can lead to proper allocation of key views for distinct scenes. In addition, another important parameter to consider is the scene overlap between neighboring key viewpoints, which can control the viewpoint intervals and the density of the viewpoint network. Although the motion parallax or disparity is an even finer property in an image sequence or video, we only measure the scene overlap briefly because our static views are largely separated and are not as redundant as video frames. A proper amount of scene overlap can maintain the continuity of virtual travel in the view network and prevent spending data on nearly identical scenes.

The algorithm to select viewpoints $V = \{P_1, P_2, ..., P_k, ..., P_m\}$ for key views is as follows. To avoid examining vast number of candidates and ensure a good visibility of key views, we extract peaks in $\sigma(P)$ distribution in order of its value, starting

Table 3.1: Number of viewpoints selected from 1165 peaks of the view significance for maximum scene overlap allowed.

α	10%	30%	50%	70%	90%
Num of images	31	41	50	61	67

from the maximum $\sigma(P)$. As a level δ decreases, more peaks in $\sigma(P)$ emerge and some nearby peaks may share a large portion of scenes. Therefore, we enforce a new viewpoint to have proper portion of novel scenes included, i.e., scene overlap with selected viewpoints should be less than threshold α. Peak P_{k+1} is selected as a key view if it has more than $(1-\alpha)\%$ new scenes. Assuming the projected image area from 3D surface set S to panorama P is *A(S)*, the condition to select a new key view is then

$$E(P_{k+1}) = \frac{A(\bigcup_{j=1}^{k} S(P_j) \cap S(P_{k+1}))}{A(S(P_{k+1}))} < \alpha \tag{3.1}$$

in the panorama. With known $\sigma(P)$ distribution and maximum allowance α, the algorithm to select key views is

Set V as an empty set
For δ decrease from $max(\sigma)$ to $min(\sigma)$
For every point P satisfying $\sigma(P) = \delta$ *and* $P \notin V$
If P is a local maximum in $\sigma(P)$ and satisfies $E(P) < \alpha$ then $V = P \cup V$;
Mark P as an examined peak

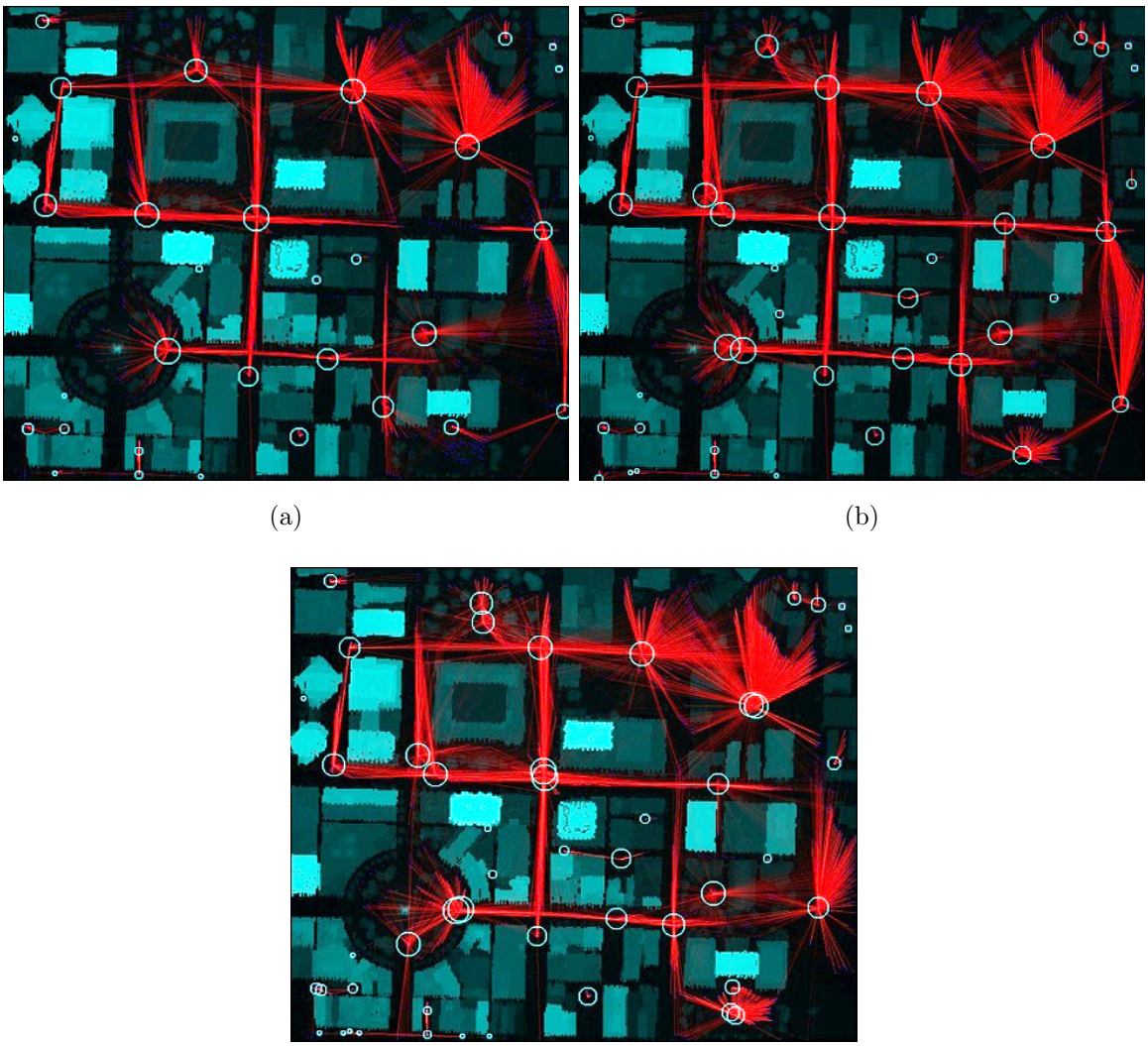
The computation of overlapping area of views uses a ray reflecting and stretching strategy. After ray r initiated from P hits a surface, reflecting rays $r_i, i = 1, ...k$, are extended from the surface point towards all viewpoints already selected. If such a ray is blocked by an obstacle, no overlap with the corresponding viewpoint is counted in $S(P_i) \cap S(P_{k+1})$.

One can choose different values of α to restrict the view overlaps. A small α may yield dispersed viewpoints with relatively smaller $\sigma(P)$ in a large area, while a large α may yield concentrating viewpoints with high $\sigma(P)$ in order to highlight a crowded area. Table 3.1 gives experiments where different α yield numbers of panoramas. The selected key viewpoints are shown in Figure 3.1, where candidate peaks of $\sigma(P)$ and selected key viewpoints are examined (connected with red lines). The intensity of a red line is proportional to the amount of scene overlap between a pair of $\sigma(P)$ peaks. Viewpoints occluded by intermediate buildings may still have weak links because they can view common buildings at a distance. A peak connected with a cluster of bright lines can be summarized as a key viewpoint.

3.2 Searching Key Views for Scene Display

The framework of visual indexing of spaces begins with a map clicking (at lower plane in Figure 3.2). Any point clicked will lead to a key view that hosts an area that the clicked point belongs to. The key view is then transmitted and displayed to briefly represent the scenes around the clicked position. Because the view significance distribution is a smooth function at all reachable positions, there should always have a peak of $\sigma(P)$ near a clicked point. The peak point is either a selected key viewpoint, or has a large scene overlap with a key view (for being rejected as a key view). We choose a key view sharing the maximum scene with the peak.

Because the computed links from peaks to key views, all the points can be associated to proper key views. Presenting key views allows navigating to neighboring spaces or subspaces. Furthermore, the view at the clicked point can even be generated roughly by transforming images from multiple key views nearby.



(a) (b)

(c)

Figure 3.1.: Relation of viewpoints with scene overlaps less than 10%, 30%, 50%, respectively. 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.

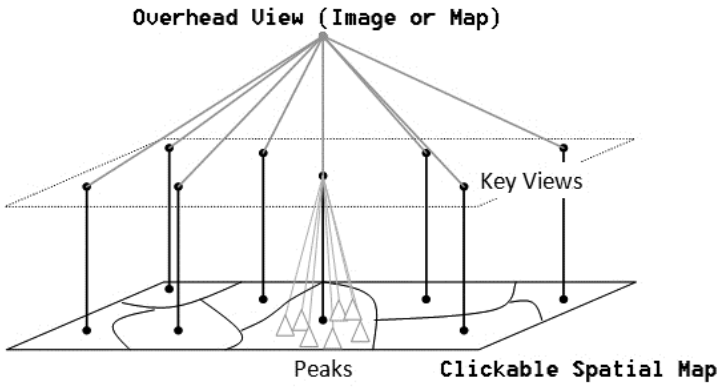


Figure 3.2.: One possible framework of visual retrieval.

4 EXPERIMENTS

The cost computation of the method is in estimating depth and view overlaps by stretching and reflecting rays. However, the total time to select key views is approximately 10 minutes with a normal desktop PC. Because our algorithms work on close-to-raw sensor data, there is no surface patch needed to be formed. Although modeling buildings with graphics patches may reduce the ray computation, no depth is provided directly in such a graphics rendering method. This makes the computation for scene overlap difficult.

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 $0 < Y < 2m$. Third, we compute rays in all orientations from *P(X, Y,Z)*, until they hit obstacles. The front tip $P_1(X_1, Y_1, Z_1)$ of the ray satisfies $Y_1 < H(X_1, Z_1)$ as it hits an object. The distance $D(\phi, \varphi)$ is then determined for the view significance. Due to the local computation of the view significance based on the view stretching method, our method can work in both crowded and sparse areas.

To compute view overlaps, the algorithm emits hemispherical rays from each peak of $\sigma(P)$, and each ray is advanced at a fixed interval. When hitting a surface point, rays are further bounced back towards the previously selected viewpoints to measure the visibility from them.

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

To verify the effectiveness of key view selection, real panoramas are taken at the planned positions to constructing a visual network of the area as shown in Figure 4.1. The panoramas are consistent with the predicted depth maps 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 is calculated and displayed in red. Because key panoramas have large scene coverage, many clickable regions can be embedded into panoramas for links and view transition. A visual index can be established from key panoramas to detailed spaces and neighboring spaces.

As shown in Table 3.1, the change of overlap threshold does not affect the number of panoramas dramatically in our results; meaning that the distinct views are stable and a significant reduction of image numbers has been achieved. Users can decide the number of key views in the visualization. Users can decided the number of key views properly for visualizing an area.

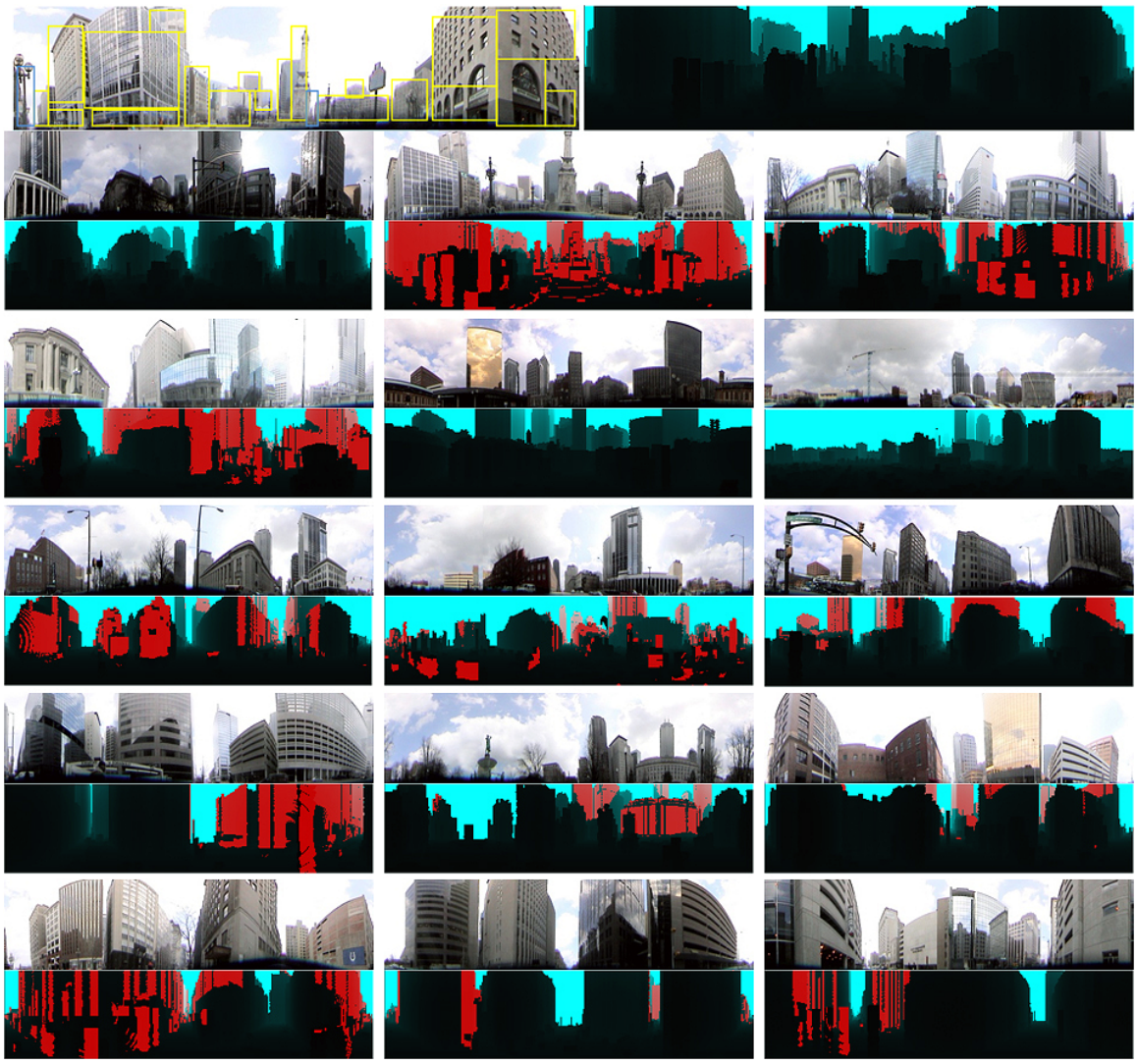


Figure 4.1.: Calculated 16 key panoramas in depth map and captured real color images. The overlapped parts in each view with other selected key views are indicated in red in the depth maps. The real panoramas have vertical FOV $\varphi \in [-5°, 45°]$ degree to avoid major ground area, cars and people. 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).

5 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 more continuous spatial information as compared to discrete landmark photos with single orientation. Therefore, discrete landmark photos should not be evaluated equally as panoramas for the space perception and navigation purposes. However, this approach does not prevent adding discrete images on special spots of interest. The key views can be considered as a type of landmark views but viewer centered.

On the contrary, our key view selection yields a much smaller data set than a video, sea of images, and Google Streetview at pervasively distributed viewpoints. Those approaches are compressed at a signal level based on the motion parallax or spatial frequency, or are not compressed. For virtual space perception using static images, the view coverage and scene overlap are at a reasonable level to enforce the view continuity.

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 closer. If a scene is weighted high for art or architecture appreciation rather than for location finding, our algorithm may place a key view close to it in order to gain resolution. 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 direct transmission.

6 SUMMARY

In this paper, we have explored a view planning scheme for scene visualization with images and surveillance cameras. Viewer centered key views, which has a function of landmark, are selected according to elevation data of a large area. This allows us to find limited representative views with the maximum scene coverage, and maintain scene novelty or continuity in a view network. We first model the scene visibility, functions, and distinctiveness numerically by the view significance. Then view overlaps at significant positions are examined for locating key views. The results may guide image acquisition for spatial perception, help camera setting for surveillance, and improve image indexing for virtual tours.

LIST OF REFERENCES

LIST OF REFERENCES

[1] J. Y. Zheng and S. Tsuji. Panoramic representation for route recognition by a mobile robot. *International Journal of Computer Vision*, 9(1):55–76, 1992.

[2] S. E. Chen and L. Williams. Quicktime VR: An image-based approach to virtual environment navigation. In *SIGGRAPH 95*, pages 29–38, 1995.

[3] R. Szeliski and H. Y. Shum. Creating full view panoramic image mosaics and environment maps. In *SIGGRAPH 97*, pages 251–258, 1997.

[4] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury. How flickr helps us make sense of the world: context and content in community-contributed media collections. In *ACM Multimedia 07*, pages 631–640, 2007.

[5] S. Coorg, M. Master, and S. Teller. Acquisition of a large pose-mosaic dataset. In *IEEE Conference on Computer Vision and Pattern Recognition 98*, pages 23–25, 1998.

[6] S. Teller. Toward urban model acquisition from geo-located images. In *International Proceedings Of Pacific Graphics 98*, pages 45–52, 1998.

[7] S. Coorg and S. Teller. Extracting textured vertical facades from controlled close-range imagery. In *IEEE Conference on Computer Vision and Pattern Recognition 99*, pages 625–632, 1999.

[8] S. K. Nayar and V. Peri. Folded catadioptric cameras. In *IEEE Conference on Computer Vision and Pattern Recognition 99*, volume II, pages 217– 223, 1999.

[9] S. K. Nayar and A. Karmarkar. 360 mosaics. In *IEEE Conference on Computer Vision and Pattern Recognition 00*, volume II, pages 388–395, 2000.

[10] D. G. Aliaga, T. Funkhouser, D. Yanovsky, and I. Carlbom. Sea of images. In *International Proceedings of IEEE Visualization 02*, pages 388–395, 2002.

[11] Point grey. http://www.ptgrey.com/products/legacy.asp.

[12] D. G. Aliaga, T. Funkhouser, D. Yanovsky, and I. Carlbom. Sea of images. *IEEE Computer Graphics and Applications*, 23(6):22–30, 2003.

[13] Y. Yagi and M. Yachida. Real-time omnidirectional image sensors. *International Journal of Computer Vision*, 58(3):173–207, 2004.

[14] S. Li, M. Nakano, and N. Chiba. Acquisition of spherical image by fish-eye conversion lens. In *IEEE Virtual Reality*, pages 235–236, 2004.

[15] A. Roman, G. Garg, and M. Levoy. Interactive design of multi-perspective images for visualizing urban landscapes. In *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, and M. Yachida. Iconic memory-based omnidirectional route panorama navigation. *IEEE PAMI*, 27(1):78–87, 2005.

[18] J. Y. Zheng, Y. Zhou, and P. Mili. Scanning scene tunnel for city traversing. *IEEE Transaction on Visualization and Computer Graphics*, 12(2):155–167, 2006.

[19] A. Agarwala, et. al. Photographing long scenes with multi-viewpoint panoramas. In *SIGGRAPH'06*, pages 853–861, 2006.

[20] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics*, 25(3):835–846, 2006.

[21] T. N. Thanh, et al. Robust and real-time rotation estimation of compound omni-directional sensor. In *IEEE International Conference on Robotics and Automation*, pages 4226–4231, 2007.

[22] B. Russell, A. Torralba, K. Murphy, and W. Freeman. Labelme: a database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1-3):157–173, 2008.

[23] A. Rav-Acha, G. Engel, and S. Peleg. Minimal aspect distortion (mad) mosaicing of long scenes. *International Journal of Computer Vision*, 78(2-3):187–206, 2008.

[24] Google street view. http://maps.google.com/.

[25] J. Hu, S. You, and U. Neumann. Integrating lidar, aerial image and ground images for complete urban building modeling. In *3D Data Processing, Visualization, and Transmission*, pages 184–191, 2006.

[26] J. Y. Zheng and X. Wang. Pervasive views, area exploration and guidance using extended image media. In *ACM Multimedia'05*, pages 986–995, 2005.

[27] B Chen, A Kaufman, and Q Tang. Image-based rendering of surfaces from volume data, volume graphics. In *Volume Graphics*, pages 279–295, 2001.