Creating Architectural Scenes from Photographs Using Model-based Stereo and Image Subregioning

Jitti Aphiboon and Borworn Papasratorn
School of Information Technology,
King Mongkut's University of Technology Thonburi,
91 Pracha-Uthit Road, Bangkok 10140, Thailand
Tel.: +66-2-470-9849, Fax.: +66-2-470-9849
e-mail : toanjoela@yahoo.com, borworn@it.kmutt.ac.th

Abstract: In the process of creating architectural scenes from photographs using Model-based Stereo [1], the geometric model is used as prior information to solve correspondence problems and recover the depth or disparity of real scenes. This paper presents an Image Subregioning algorithm that divides left and right images into several rectangular sub-images. The division is done according to the estimated depth of real scenes using a Heuristic Approach. The depth difference between the reality and the model can be partitioned into each depth level. This reduces disparity search range in the Similarity Function. For architectural scenes with complex depth, experiments using the above approach show that accurate disparity maps and better results when rendering scenes can be achieved by the proposed algorithm.

1. Introduction

The process of creating architectural scenes from photographs using Model-based Stereo is to map points in the key (left) image to the corresponding points in the warped offset (right) image. A disparity map can be obtained in the stereo matching process using simple correlation approach, even for the camera positions are relatively far apart (large baseline) [1]. Appropriate disparity search range is the key in creating a disparity map for the image pairs having many different depth levels. Although a coarse model simplifies the matching process, there are some chances of selecting false matches depending on the disparity search range used in the Similarity Function [2]. Therefore, the rendering scenes will contain holes and noise.

In this paper, we propose an Image Subregioning technique that divides the key (left) and the warped offset (right) images of architectural scenes into several rectangular sub-images with similar depth levels. The proposed technique will minimize error in the matching process caused by choosing improper disparity search range. Therefore, the better results of rendering scenes can be achieved.

2. Similarity Functions and Matching

Suppose that $I_l(x_l,y_l)$ and $I_r(x_r,y_r)$ are the intensity of the key (left) and the warped offset (right) images, respectively. Since each pixel in the key image is known, we can find a match of that pixel in the warped offset image from:

$$I_r(x_r,y_r) = I_l(x_l,y_l + D_h(x,y)) \quad (1)$$

where $D_h(x,y)$ is a disparity map at position (x,y); that is, (x_r,y_r) in the warped offset image corresponds to $(x_l,y_l + D_h(x,y))$ in the key image. SAD (Sum of Absolute Differences) is used as Similarity Function since it is simple and uses less computing power [3]. SAD uses Local Search method [2] as matching algorithm. The correlation of two windows of size $(2N+1)\times(2N+1)$, where N is any integer values, in the two images is performed along the same horizontal scan line. If for any points in the key image, the search window varies from the minimum disparity d_{min} to the maximum disparity d_{max} in the warped offset image as shown in Fig. 1. The correlation $S(x,y,d)$ can be calculated by:

$$S_{sad}(x,y,d) = \sum_{i=-N}^{N}\sum_{j=-N}^{N} \left| (I_l(x+i,y+j) - I_r(x+i,y+j+d)) \right| \quad (2)$$

The value of d in Eq. (2) is disparity search range. The total range of disparities is given by $drange = |d_{max} - d_{min}| + 1$ according to the disparity limit constraint [2]. Only one value of d that produces the minimum of the function is determined as the disparity for coordinates (x,y), according to the uniqueness constraint [4].

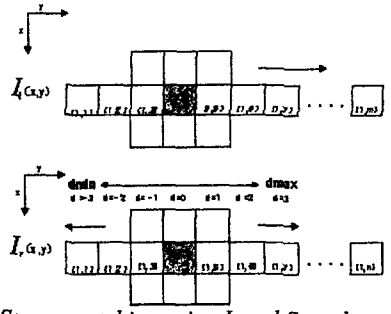

Figure 1. Stereo matching using Local Search method.

3. Image Subregioning and Rendering Algorithms

3.1 Using Sub-images

Using large disparity search range while matching between each horizontal scan line of the whole key image and the whole warped offset image, incorrect disparity values may occur in a disparity map. It is occurred where the depth level of architecture changed. Because the value of d is derived from the minimum $S(x,y,d)$ value in the Similarity Function, it is possible that d can be improperly selected. If the appropriate disparity search range at depth level z_1 and z_2, where $z_2 > z_1$, is from 0 to 3 (d_1) and from 0 to 10 (d_2), respectively, the disparity search range used in the matching process should be from 0 to 10 to cover the depth level from z_0 to z_2 as shown in Fig. 2. Thus, the improper disparity search range at depth level z_1, from 7 to 10, is

possible. For $m \times n$ image, the computation complexity is $m \times n \times d$, where m, n are the number of row and column, and d is the disparity search range over the whole image. By reducing the size of the key and the warped offset images before the matching process [3], the computation complexity can be decreased.

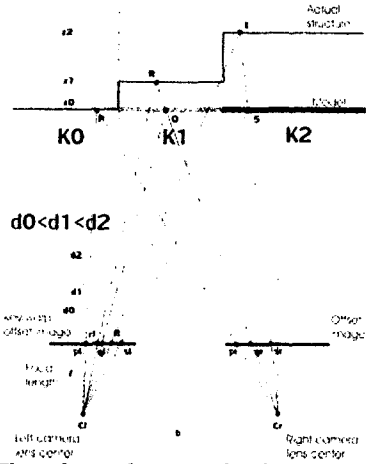

Figure 2. The relation between the disparity search range and the depth level.

Since the depth level of most architecture changes vertically along the column spans or the edges of the building, these can be properly used as the positions for dividing the key and the warped offset images into several sub-images. The division is done according to the estimated depth of real scenes using a Heuristic Approach as will be described below. If the image is partitioned into k vertical sub-images, the disparity search range is reduced to d_i for each sub-image of size $m \times n_i$ as shown in Fig. 3, where i is the order of the 0 th to the $k-1$ th sub-image. The computational complexity will be $\sum_{i=0}^{k-1}(m \times n_i \times d_i)$, which is smaller than $m \times n \times d$ [3].

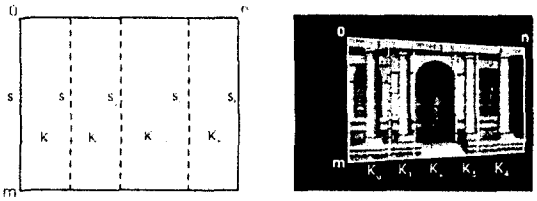

Figure 3. Sub-dividing the whole image into vertical sub-images.

3.2 Heuristic Approach

Heuristic is one of the problem-solving techniques that domain experts use to generate good solutions without exhaustive search [5]. In our proposed algorithm, we use a domain expert in Architecture and Image Processing to determine the positions of the division in the image pairs as the following steps:

1. Use thresholds T_l defined by visual inspection of the domain expert to classify the depth of real scene to L levels, where $l = 1$ to $L-1$;
2. Use s_1 to s_{k-1} as the positions of the column spans or the edges of the building, which can be seen in both the key and the warped offset images. s_0 and s_k are the left and the right borders of the images, respectively;
3. Partition the k_i th sub-image, where $i = 0$ to $k-1$. If $i = 0$, use s_0 as the beginning of the division and s_j as the end of the division, where $j = 1$ to k. If $i > 0$, use s_{j-1} as the beginning of the division and s_j as the end of the division;
4. If the depths of most area in the k_i th sub-image are the L th level, define this sub-image as the k_i th sub-image and be the L th level;
5. If the L th level of the k_{i-1} th sub-image is the same as the k_i th sub-image, merge these two sub-images into the k_{i-1} th sub-image and be the L th level; otherwise go to Step 6;
6. Set $i = i+1$ and $j = j+1$;
7. Repeat Step 3 through 6 until reaching the k_{k-1} th sub-image, using s_{k-1} as the beginning of the division and s_k as the end of the division.

3.3 Rendering Strategy

Thus, the sub-images will be used as input of matching algorithm described in Section 2. $D_{kl}(x,y)$ and $D_{rl}(x,y)$ will be disparity maps using the key and the warped offset sub-images as a reference image, respectively. We can render the scene from novel viewpoints using Geometric Transform [6]. The intensity of sub-images at position (x,y) will be moved to the new position (x',y') according to their disparity maps as the following equation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x + D \\ y + D \end{bmatrix} \tag{3}$$

Because disparity is inversely proportional to depth [7], the first visibility problem, image folding or image overlaps, can be handled. The intensity of sub-images moved by more disparity value will be in front of that of sub-images moved by less disparity value. The second visibility problem, image stretching or holes, are handled by a morphing technique [6]. Holes are filled with the combination of the intensity of the key and the warped offset sub-images using the following weighting function:

$$I_v = (1-s) \cdot I_l + s \cdot I_r \tag{4}$$

where I_v is the intensity of novel rendering, and s is the weights of the combination, which range from 0 to 1. In our proposed algorithm, the morphing technique is performed by adding the following conditions:

1. Identify holes, where the average of intensity values (RGB) is less than or equal to a predefined threshold T_h.
2. Define $I_l + D$ and $I_r + D$ as the images, which their intensities have been moved to the new position according to their disparity maps using Eq. (3).
3. For each pixel in $I_l + D$ and $I_r + D$, if the average of intensity values of $I_l + D > T_h$ and that of $I_r + D > T_h$, then I_v is obtained by combining the intensity of $I_l + D$ and $I_r + D$ using Eq. (4).
4. If the average of intensity values of $I_l + D > T_h$ and that of $I_r + D \le T_h$, then I_v is obtained by the intensity of $I_l + D$.
5. If the average of intensity values of $I_l + D \le T_h$ and that of $I_r + D > T_h$, then I_v is obtained by the intensity of $I_r + D$.
6. If the average of intensity values of $I_l + D \le T_h$ and that of $I_r + D \le T_h$, then I_v is obtained by combining the intensity of I_l and I_r using Eq. (4).

Since the novel camera position is defined as Current Frame (c), and the distance between the key and the warped offset camera positions is defined as Total Frame (T), the intensity of novel rendering ($I_{v_{C_i}}$) can be obtained by linear interpolating [6] as:

$$I_{v_{C_i}} = I_{k_i}(x, y + (1 - \frac{C}{T}) \cdot D_{k_i}(x, y)) + I_{r_i}(x, y + (\frac{C}{T}) \cdot D_{r_i}(x, y)) \quad (5)$$

By compositing of these sub-image renderings, the final result of architectural scene is created. The proposed algorithm can be summarized as shown in Fig. 4.

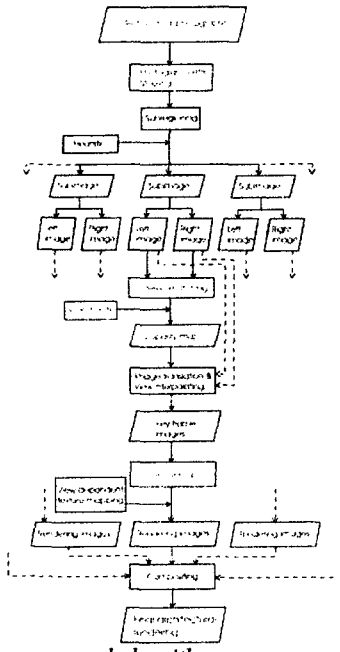

Figure 4. Our proposed algorithm.

4. Experimental Results

In our experiments, two groups of the synthetic architectural scenes; low complexity depth (L-1-1 to L-1-4), and high complexity depth (H-1 to H-4), have been used as the key and the warped offset images. By applying Caching Technique [8], each pair of the synthetic architectural scenes can be created efficiently. The time spent in the matching process using subregioning technique was recorded and compared to the one without using subregioning technique. The quality of rendering scenes is also measured using the objective fidelity criteria [9][10], Signal to Noise Ratio (SNR_{rms}) and Correlation Coefficient. Table 1 gives the summary results of the tested images using our proposed algorithm compared to the conventional one.

Table 1. The summary results from our experiments using the proposed algorithm compared to the conventional one.

		Difference of times spent in the matching process (%)	Difference of the number of disparities (%)	Difference of mean of SNR (%)	Difference of mean of Correlation Coefficient (%)
L-1-1		24.62	3.94	0.44	-0.09
L-1-2		34.89	5.27	1.88	0.33
L-1-3		16.55	-0.84	4.11	1.38
L-1-4		33.18	5.95	1.15	-0.23
H-1		34.89	11.08	1.34	0.44
H-2		49.38	-1.54	2.63	0.61
H-3		51.30	-0.05	2.29	0.61
H-4		31.46	-2.96	2.52	0.33

Fig. 5 shows some of rendering scenes, starting from frame 0 to frame 10, using subregioning technique and without using subregioning technique. Comparing the results at about frame 3 to frame 7, it can be seen that our proposed algorithm gives better rendering results.

5. Discussions and Conclusions

As can be seen in Table 3 and Fig. 5, the proposed subregioning algorithm can help reduce the time spent in the matching process significantly while the better results of rendering scenes can be obtained. For low complexity depth scenes, time saving is 16-34%, while high complexity depth scenes, time saving is up to 51% compared to the conventional matching method. By considering the width of the deepest area in architectural scene, the tested images having narrow segment of the deepest area (L-1-2, H-1, H-2, and H-3) can reduce more computation time. We also found that symmetrical architecture can help increase the number of disparities in the matching process and produce higher value of SNR_{rms} and Correlation Coefficient.

We can conclude the characteristics of architectural scene, the most appropriate for using our proposed algorithm, as a scene with high complexity depth, narrow segment of the deepest area, and symmetrical structure.

There are several improvements and extensions to our proposed algorithm. First, our matching techniques only use the intensity value in the Similarity Function, so there are not enough information to determine the minimum of the function correctly. By using or adding other features of the scenes in the matching process, more accurate disparity maps can be achieved. Second, our subregioning method needs a domain expert to determine the appropriate positions for dividing images. The system should be extended so that it can be able to partition images automatically by using some architectural knowledge as the input. Lastly, other directions of the division can be used to overcome the problems for dividing architectural scenes having complex structure such as Thai Architecture.

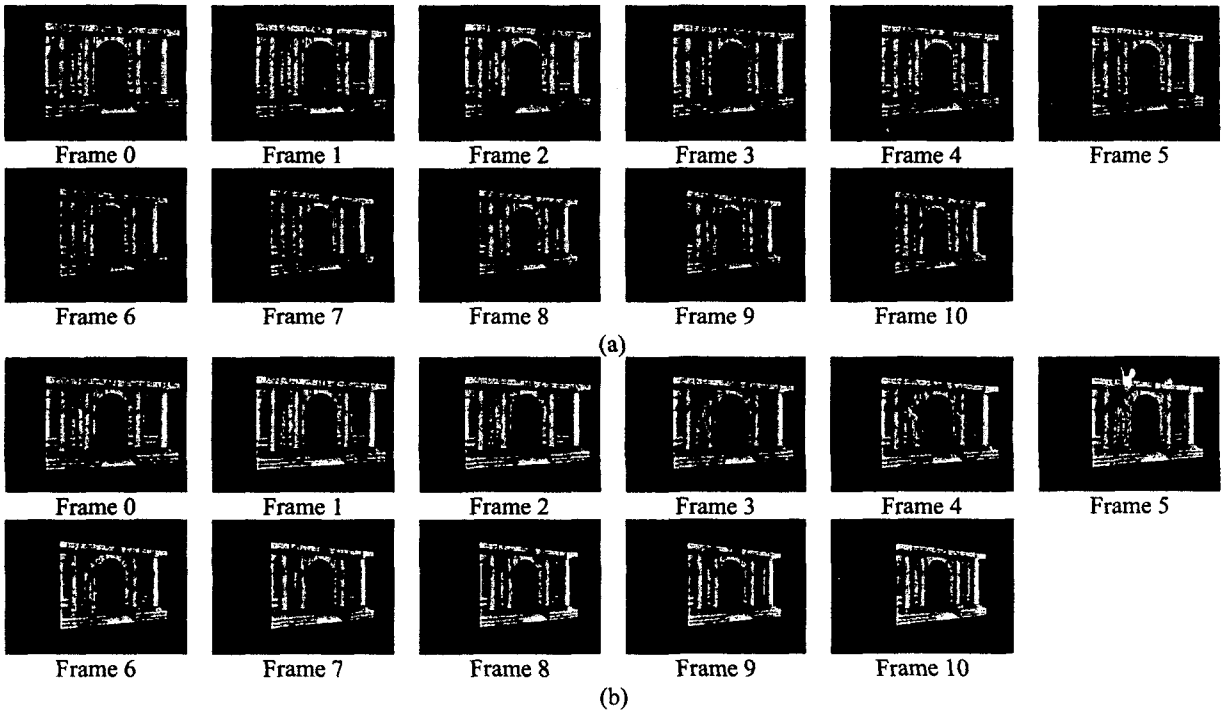

Figure 5.Some of rendering scenes (H-2). (a) using subregioning. (b) without using subregioning.

Acknowledgment

This research was supported by School of Information Technology, KMUTT.

References

[1] Debevec, P. E., 1996, "Modeling and rendering architecture from photographs", *Ph.D. thesis, University of California at Berkeley.*

[2] Fielding, G., and Kam, M., 1999, "Weighted matchings for dense stereo correspondence", *http://citeseer.nj.nec.com/fielding99weighted.html.*

[3] Sun, C., 1998, "Multi resolution rectangular subregioning stereo matching using fast correlation and dynamic programming techniques", *CMIS report*, No. 98/246.

[4] Faugeras, O., 1993, *Three dimensional computer vision*, MIT press.

[5] Russell, S., and Norvig, P., 1995, *Artificial Intelligence A Modern Approach*, Prentice Hall Inc.

[6] Watt, A., Policarpo F., 1998, *Computer image*, Addison wesley.

[7] Jain, R. K., Kasturi, R., and Schunck, B.G., 1995, *Machine vision*, McGraw-Hill Inc.

[8] Papasratorn, B., and Vanijja, V., 1998, "Stereographic with caching technique", *The National Computer Science and Engineering Conference.*

[9] Umbaugh, S. E., 1998, *Computer Vision and Image Processing a practical approach using CVIPtools*, International edition.

[10] Gonzalez, R. C., and Woods, R. E., 2002, *Digital Image Processing*, second edition, Prentice Hall Inc.