Object Modeling from Multiple Images Using Genetic Algorithms

Hideo SAITO and Masayuki MORI
Department of Electrical Engineering, Keio University
E-mail: saito@ozawa.elec.keio.ac.jp

Abstract

This paper describes about an application of genetic algorithms (GAs) to modeling of multiple object from CCD images. Shape modeling is a very important issue for shape recognition for robot vision, representing 3-D shapes in the virtual world, and so on. In this paper, we propose a method for object modeling from multiple view images using genetic algorithms (GAs). In this method, similarity between the model and the image at each view angle is evaluated. The model having the maximum evaluation is found by GAs. In the proposed method, sharing *scheme is used for finding multiple solutions efficiently. Some results of object modeling experiments from synthetic and real multiple view images demonstrates the proposed method can robustly generate model by using the GA.*

1 Introduction

Object modeling is very important issue of computer vision, which can be applied to object recognition, environment recognition, construction of virtual 3-D space, and so on. Current techniques for object modeling mostly requires human operators [1]. Therefore, automatic object modeling techniques from real images can reduce the cost of model construction.

Many methods have recently been studied for recovering range information from a sequence of CCD images of multiple views. In such methods, some corresponding points between each images must be detected. If the detection of the corresponding points can successfully be performed, the range information can be accurately recovered by using some methods [2, 3, 4] based on principal component analysis, and so on. However, some mistaken correspondence are usually detected, since there are a lot of candidates for the corresponding features in the real scene images. Then recovered range information includes some errors.

If we think that the object problem is not estimating accurate range information but generating accurate object models, we don't have to explicitly recover the range information from the images. For example, interactive operation system for object modeling [5] efficiently provides the model from multiple view real images. Using this system, operator can be obtain accurate object modeling by repeating generation of hypothesis of models and evaluation of the hypothesis on the multiple view images. Although this system can give good object models, the system requires human interaction.

In this paper, we propose an application of genetic algorithms (GAs) to object modeling from multiple view images. In this method, the model matching to every input image is efficiently found by applying GAs which repeat evaluation of hypotheses of the models. There are some applications of GAs to finding registrated patterns from images [6], obtaining superquadrics models from shading images [7], and so on [9]. As described in the previous applications of GAs, GAs can effectively provide almost optimized solution from variety of candidates. In our modeling method by GAs, the efficiency in the optimization is the same. Shape, position, and pose of the object are simultaneously determined by optimizing the evaluation of the similarity between the model and the images.

2 Proposed Method

2.1 Definition of the Problem

In this study, we assume the object is a building which can be regarded as the polyhedral model. Then, the problem is how to estimate shape, position, and pose of the model from the input multiple view images. Figure 1 shows an example of the assumed scheme in the proposed method. Here, several images of the object are taken from several view directions. In the experiment performed in this paper, the camera parameters are previously known.

(a) Flow of object modeling.

(b) Modeling scheme.

Figure 1. Scheme of the shape modeling assumed in this paper.

2.2 Modeling by GAs

The modeling from the multiple view images is performed by maximizing an evaluation function by the use of GAs. The evaluation function is determined by the similarity between the input multiple view images and the projected images of the estimated object model shape. The flow of the modeling process by the GA is shown in figure 2.

In this method, an object is regarded as a set of simple primitives as shown in figure 1 (b). Because the actual objects are supposed to artificial buildings, the primitives can be limited to rectangular solids or triangular pillars which have arbitrary position, size and pose. As shown in figure 3, parameters of each primitive are position of the object barycenter, size and pose of the primitive, which are defined as (xc, yc, zc), (AX, AY, AZ), and θ, respectively. The pose of the primitive has only one degree of freedom because the assumed object is buildings which are constructed parallel to the ground plane.

2.2.1 Encoding of parameters

A set of the parameters is encoded into a string in the GA. For each parameters, eight bits are used. One bit is allocated to distinguish rectangles and triangles. Then, a string consists of 57 bits binary. The gray coding scheme is employed for making the hamming distance of the strings to correspond to the difference of the parameter values.

2.2.2 Evaluation of string

Each string is evaluated by the similarity between the input multiple view images and virtually synthesized model images as shown in figure 4. We define the following criterion.

Figure 2. The flow of the modeling process by GA.

1. Correlation between blurred gradient images of the input images and the wireframe images of the primitive represented by a string ($E1$).

2. Consistency of texture patterns on the model plane which are provided by every input images ($E2$).

The first criterion can be defined as the following equation.

$$E1 = \sum_i \frac{\sum_{x,y} B_i(x, y) S_i(x, y)}{\sqrt{\sum_{x,y} B_i^2(x, y)}\sqrt{\sum_{x,y} S_i^2(x, y)}}, \quad (1)$$

where $S_i(x, y)$ represents wireframe image of the model synthesized at ith view, and $B_i(x, y)$ is blurred gradient image of the input image at ith view. $B_i(x, y)$ is calculated as

$$B_i(x, y) = G(x, y) * \sqrt{\{\frac{\partial I_i(x, y)}{\partial x}\}^2 + \{\frac{\partial I_i(x, y)}{\partial y}\}^2}, \quad (2)$$

Figure 3. The encoding scheme of primitive.

where $G(x, y)$ is Gaussian distribution and $I_i(x, y)$ is input image at ith view.

The second criterion is defined by the variance in the texture images back-projected on the model plane as shown in figure 5. Then the criterion $E2$ is defined as

$$E2 = \sum_{x,y} \sum_i \{T_i(x, y) - \frac{\sum_j T_j(x, y)}{p}\}^2 \qquad (3)$$

where $T_i(x, y)$ represents texture pattern back-projected from ith input image, and p is the number of the images back-projected on the model plane. If the model parameters are accurate, every back-projected texture must be the same and then $E2$ must be 0.

The total evaluation function E is

$$E = E1 - \alpha \times E2, \qquad (4)$$

where α is a weighting constant.

2.2.3 Sharing

Because the object may consist of multiple primitives, they must be obtained simultaneously. The GA is suitable for finding multiple solutions, because the GA hold many strings in the population. For finding multiple solutions efficiently, we employ *sharing* scheme [8].

By the sharing, evaluation of the string is decreased if the string is similar to other strings in the population. This prevents the one point convergence of the solution, and then the multiple solutions can be obtained. The relationship between the modified evaluation and the original evaluation is shown in the following equation.

Figure 4. Evaluation of string.

$$E_s(x_i) = \frac{E(x_i)}{\sum_{j=1}^{n} s(d(x_i, x_j))}, \qquad (5)$$

$$\text{where } s(d) = \max(1 - d/\sigma, 0),$$

x_i, x_j: ith and jth strings
$E(x_i)$: original evaluation of x_i
$E_s(x_i)$: modified evaluation of x_i
n: number of the strings in a population
$d(x_i, x_j)$: distance between x_i and x_j
$s(d)$: sharing function of d
σ: constant determining effect of sharing.

The distance between two strings $d(x_i, x_j)$ is evaluated by the hamming distance between the binary code of strings.

2.2.4 Genetic operations

First, The initial strings are generated at random. Next, each string is evaluated by E_s. According to the evaluation, some elite strings are selected. The selected elite strings are improved by a local search method and preserved as offspring strings. For the rest of strings, the parent strings are selected according to the selection probability which is proportional to the evaluation

Figure 5. Textures of correspondence areas are back-projected onto the model surface. These back-projected textures are represented by $T_{1,2,3..}(x, y)$.

Figure 6. By the sharing, the population tend to have multiple solutions by the alternation of generations.

E_s (Reproduction). Then the offspring strings are generated by the one-point crossover which is performed bit-wise. Some bits which are selected at random are reversed by the mutation. This process is repeated.

After repeating the alternation of generations, some strings gather around a solution, and some other strings gather around the other solution as shown in figure 6. According to the distribution of the strings in the searching space, we have to segment some solutions which represents object models. However, the segmentation is quite difficult problem because the strings are not perfectly separated in the searching space. In our experiments, we employ a heuristic way for the segmentation, but it does not work for every condition. This problem must be studied for the future work.

3 Experiments

For demonstrating the efficacy of the proposed algorithm, we try to obtain object models from both synthetic and real images.

In this experiments, three view images with 45° interval are used for modeling. In the GA, 256 strings are used in a population, and 100 generations are repeated for obtaining optimal solutions.

Figure 8 shows the object model obtained from the synthetic multiple view images shown in figure 7. As shown in figure 8 (a), two primitives of rectangular solids can be obtained simultaneously. However, the obtained model does not match with the input images completely, because the string segmentation does not work well. Figure 8 (b) shows an example of synthesized image of the object at other view using the obtained model.

Figure 10 shows the object model obtained from the real multiple view images shown in figure 9. The object is a toy building replica put on the floor of our laboratory. Because there are many features on the background scene in those multiple view images, it is difficult to successfully detect correct corresponding points on the object, and to recover accurate range information. By the proposed algorithm, however, the best hypothesis of the model is efficiently found without explicit recovery of the range information. Thus, the object modeling is successfully performed as shown in figure 10 (a).

4 Conclusion

In this paper, we propose a method for object modeling from multiple view images using genetic algorithms (GAs). In the proposed algorithm, the best hypothesis of the model is efficiently found without explicitly recovering the range information, the object modeling is successfully performed. Some results of object modeling experiments from synthetic and real multiple view images demonstrates the proposed method can robustly generate model by using the GA.

(a) (b) (c)

Figure 7. Examples of multiple view images ((a),(b), and (c)).

(a) (b)

Figure 8. The object model obtained from the synthetic multiple view images.

(a) (b) (c)

Figure 9. Real multiple view images.

(a) (b)

Figure 10. The object model obtained from the real multiple view images. White line in (a) represents the obtained object model. An example of the synthesized image at other view is shown in (b).

References

[1] F.Arman and J.K. Aggarwal, "Model-based object recognition in dense range images–a review", *ACM Computing Surveys*, vol.25, no.1, pp.5-43, 1993.

[2] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method", *Int'l J. Comput. Vision*, vol.9, no.2, pp.137-154, 1992.

[3] H.Y.Shum, K Ikeuchi, and R. Reddy, "Principal component analysis with missing data and its application to polyhedral object modeling", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.17, no.9, pp.854-867, 1995.

[4] H Yu, Q. Chen, G. Xu, and M Yachida, "Recovering 3-D shape and motion from image sequence under perspective projection by SVD", *Proc. 2nd. Asian Conf. on Comput. Vision*, vol.2, pp.171-175, 1995.

[5] M Watanabe and S. Shintani, "Machine supporting function of Sketch Vision, CG modeling tool based on video-images", Symposium of Information Processing Society of Japan, vol.95, no.5, pp.9-16, 1995.(in Japanese)

[6] E Lutton and P.Martinez, "A genetic algorithm for detection of 2D geometric primitives in images", *12th IAPR International Conference on Pattern Recognition*, vol.1, pp.526-528, 1994.

[7] H Saito and N. Tsunashima, "Estimation of 3-D parametric models from shading image using genetic algorithms", *12th IAPR International Conference on Pattern Recognition*, vol.1, pp.668-670, 1994.

[8] D Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

[9] E.S.Gelsema, Special Issue on Genetic Algorithms, *Pattern Recognition Letters*, vol.16, no.8, 1995.