Computer Modeling and Simulation of an Active Vision Camera System

Ming-Chin Lu

Communications Excellence, Inc.
540 Marl Road, Colts Neck, NJ 07722
e-mail: mclu@jolt.att.com

Murali Subbarao

Department of Electrical Engineering
State University of New York, Stony Brook, NY 11794
e-mail: murali@sbee.sunysb.edu

ABSTRACT

Verification of computer vision theories is facilitated by the development and implementation of computer simulation systems. Computer simulation avoids the necessity of building actual camera systems; they are fast, flexible, and can be easily duplicated for use by others. In our previous work, we proposed a useful computational model to explore the image sensing process. This model decouples the photometric information and the geometric information of objects in the scene. In this paper, we further extend the proposed image sensing model to simulate the image formation of moving objects (motion) and stereo vision system. The simulation algorithms for curved objects, moving objects, and stereo imaging are presented. Based on the proposed model and algorithms, a computer simulation system called Active Vision Simulator (AVS) has been implemented. AVS can be used to simulate image formation process in a monocular (MONO mode) or a binocular (STEREO mode) camera system to synthesize the images. It is useful for research on image restoration, motion analysis, depth from defocus, and algorithms for solving the correspondence problem in stereo vision. The implementation of AVS is efficient, modular, extensible, and user-friendly.

Keywords: image sensing and modeling, computer simulation, camera modeling, active vision

1 INTRODUCTION

Many theories have been developed in computer vision during the past three decades for recovering scene information. Verification of such computer vision theories often require expensive and accurate camera systems, and laboratory facilities for calibration and experimentation. As an alternative, it is possible to develop computational models of the camera system, and simulate the system on a computer. This is not only faster and cheaper than building actual camera systems and setting up expensive laboratories, it also provides flexibility and accuracy. The physical parameters of the camera system (*e.g.* focal length, sampling rate, quantization level, noise characteristics, optical aberrations, etc.) are easily changed and they can be set to desired values to very high accuracy. The only major limitation is the amount of computational resources required for simulation.

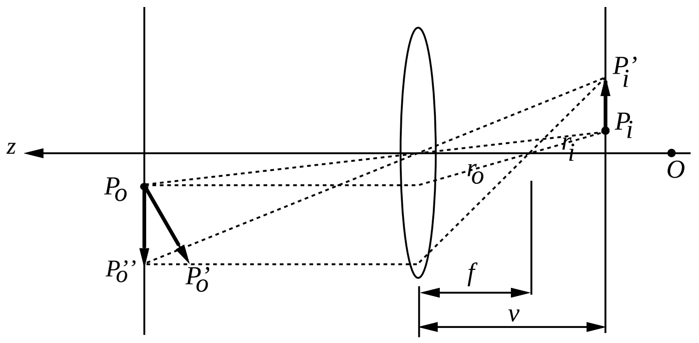


Figure 1: Relationship between the displacement of a point in the scene and the corresponding point in the image.

In active vision, changing the direction of view and the visual parameters facilitates and makes efficient the computational stage of machine vision. An active vision system can be considered as a system that integrates visual sensing and action. There are two common tasks to be solved in active vision systems: one is the correspondence problem in stereo imaging, the other is motion estimation to dynamically track the objects in the scene. Many researchers have proposed algorithms[1–6,8,11] for these tasks. Our objective here is to provide researchers a simulation environment to simulate image sensing process in motion and stereo systems.

In our previous work,[7,10] we proposed a computational model on the image sensing and formation process of a CCD camera system. This model decouples the photometric properties of a scene from the geometric properties of the scene in the input to the camera system. In this paper, we further extended the proposed computational model to simulate the image formation of moving objects (motion) and stereo vision system. Based on the extended computational model, a computer simulation system called Active Vision Simulator (AVS) is developed. AVS is an extension of the Image Defocus Simulator (IDS) presented in.[7,10] It can be used to simulate image formation process in a monocular (MONO mode) or a binocular (STEREO mode) camera system. The simulation of curved objects is also included in AVS. The user interfaces for AVS are similar to those in IDS.[7]

This paper is organized as follows: Section 2 presents the computational model for motion and stereo simulation; Section 3 describes the simulation algorithms used for curved objects, motion simulation, and stereo imaging; Section 4 describes the user interfaces of AVS; Section 5 presents the simulation results; and finally, Section 6 concludes this paper.

2 CAMERA MODEL

In this section, we will extend the computational model presented in[10] to simulate the image sensing process for moving objects and binocular stereo camera systems.

2.1 Motion simulation

When objects move in front of a camera, or when a camera moves through a fixed environment, there are corresponding changes in the images. The displacement of a point in the environment will cause a displacement of the corresponding image point. In motion simulation, we assume that all the objects in the scene are rigid objects. Therefore, the shape of the objects will not change during motion.

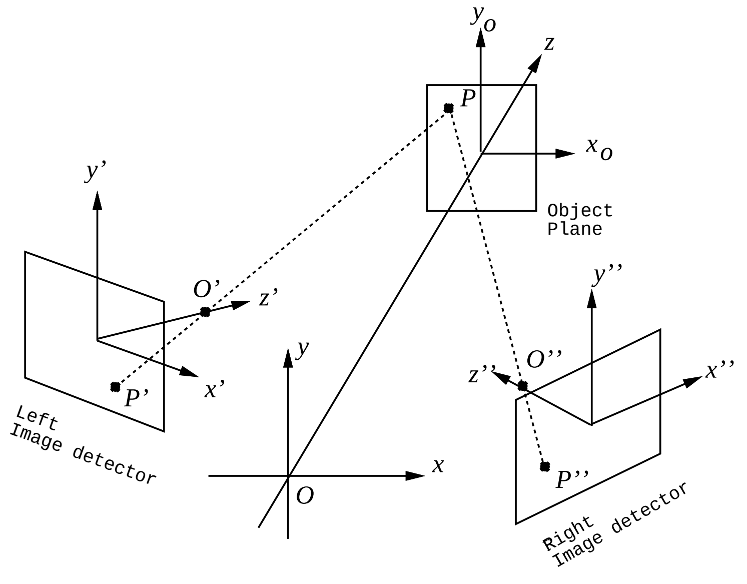


Figure 2: A general stereo system model.

Figure 1 shows the relationship between an object motion vector $\vec{m_o} = P_o\vec{P}_o' = [V_{x0} \quad V_{y0} \quad V_{z0}] \quad \Delta t$ and the image motion vector $\vec{m_i} = P_i\vec{P}_i'$. For simplifying the discussion, the image plane is placed at the focused position and is perpendicular to the optical axis (z-axis). The vector $\vec{m_o}$ can be decomposed into two components, one parallel to the x-y plane ($P_o\vec{P}_o''$) which shifts the object, and another parallel to the z-axis which changes the size of the object.

Consider the translation vector $P_o\vec{P}_o''$. Let $P_o\vec{P}_o'' = \vec{V_t} \quad \Delta t = [V_{xo} \quad V_{yo} \quad 0] \quad \Delta t$ for a fixed time interval Δt. This corresponds to a motion vector $\vec{m_i} = \vec{V_i} \quad \Delta t = [V_{xi} \quad V_{yi} \quad 0] \quad \Delta t$ in the image plane. The amount of displacement is $||P_o\vec{P}_o''|| = ||\vec{V_t}\Delta t||$ in the scene which corresponds to a displacement of $||P_i\vec{P}_i'|| = ||\vec{V_i}\Delta t||$ in the image plane.

From the geometry in Figure 1, we have

$$\frac{||\vec{V_i}\Delta t||}{||\vec{V_t}\Delta t||} = \frac{r_i}{r_o} = \frac{v-f}{f} \tag{1}$$

The displacement of points in the image plane can be computed using Equation (1). For z-axis movement, *i.e.* $V_{zo} \neq 0$, there will be a change in the size of the objects in the scene. This results in image magnification or shrinking which requires image interpolation and resampling.

2.2 Stereo vision system

A general stereo system model is shown in Figure 2 where O is the global origin, O' and O'' are the entrance pupil origin of the left and the right cameras, respectively. The left and the right cameras can be treated as monocular camera systems similar to that in Figure 3. The global origin O is introduced as a reference point for the positions of the object, the left camera, and the right camera.

In this generalized stereo system, the optical axes of the two cameras are not parallel, but intersect at some point in the scene. In order to simplify computations in our simulation, we restrict the optical axes of the cameras

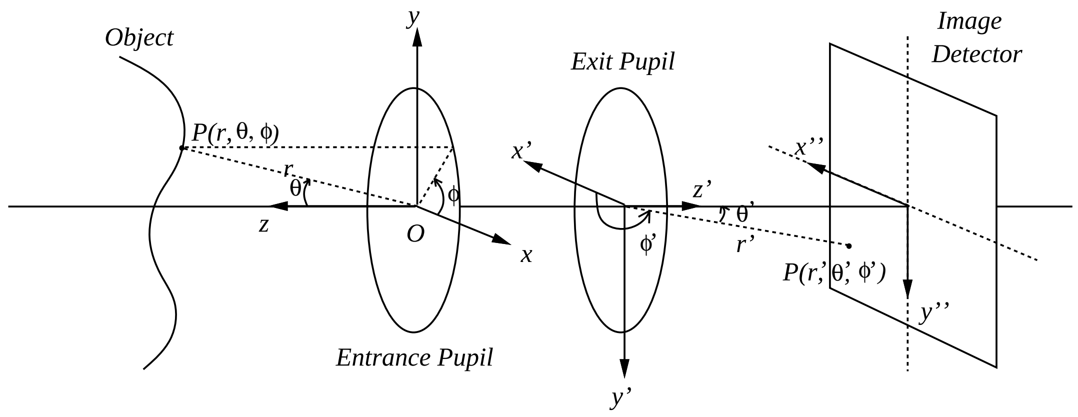


Figure 3: Entrance pupil coordinate system.

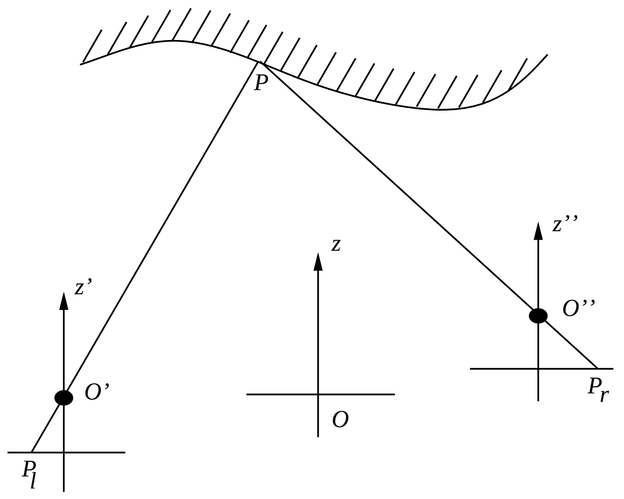


Figure 4: Global coordinate system used in stereo simulation.

to be parallel. In this model, there is a relative translation between the two cameras, but no relative rotation. This restriction can be removed at the expense of more computation.

Figure 4 is the global coordinate system used in our current stereo simulation where z-, z'-, and z"-axes are parallel to each other. Based on this configuration, the stereo vision system can be modeled as shown in Figure 5. The scene information is first translated, and scaled with respect to the origin of each camera. After this transformation, the photometric information $f(\theta, \phi, \lambda, t)$ and the geometric/depth information $r(\theta, \phi)$ are transformed to $f_l(\theta, \phi, \lambda, t)$, $f_r(\theta, \phi, \lambda, t)$ and $r_l(\theta, \phi)$, $r_r(\theta, \phi)$ for the left and the right cameras. These functions are the input to the camera system. The remaining functional blocks are the same as those presented in.[10]

3 SIMULATION ALGORITHMS

3.1 Curved objects

Consider the photometric information $f(\theta, \phi, \lambda, t)$ and the geometric information $r(\theta, \phi)$. $r(\theta, \phi)$ contains the depth information of objects in the scene. For curved objects, $r(\theta, \phi)$ can not be approximated by a constant

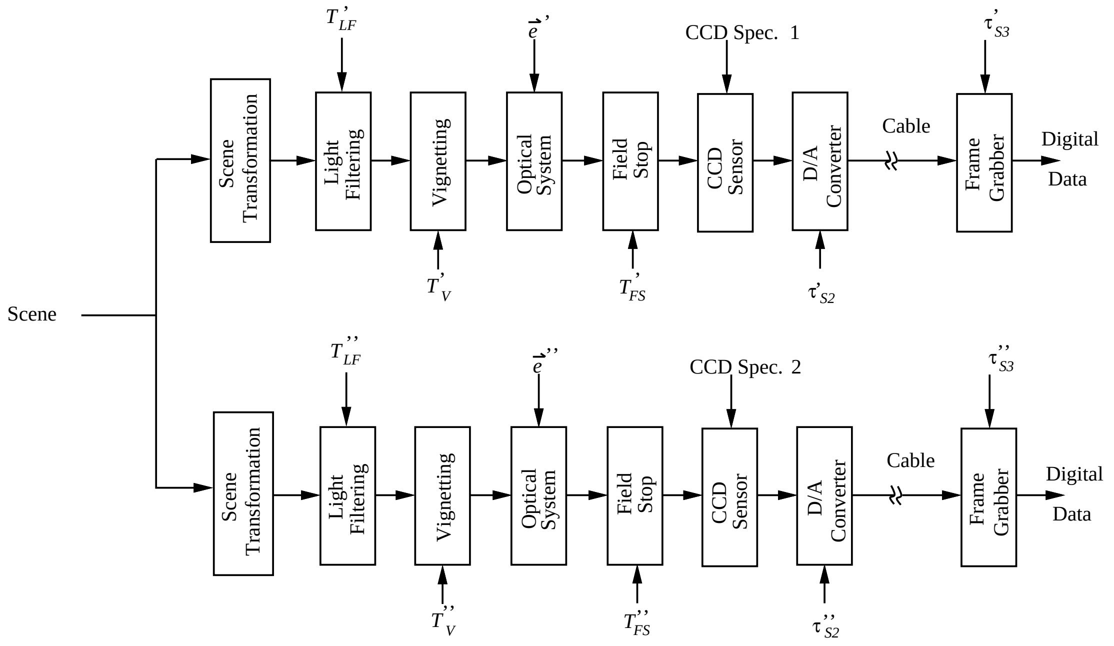


Figure 5: Block diagram of a stereo vision system.

u. Under this situation, the point spread function is space-variant and is specified by $h(\theta, \phi, \theta', \phi', r(\theta, \phi), \vec{e})$ as discussed in.[10] In a Cartesian coordinate system, the geometric information and the point spread function can be represented as $r(x, y)$ and $h'(x, y, x', y', r(x, y), \vec{e})$, respectively, under the assumption that all CCD elements have the same characteristics. In this case, the output of the optical system will be:

$$f_3(x, y, \lambda, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h'(x - x', y - y', r(x, y), \vec{e}) \cdot f_2'(x', y', \lambda, t) \quad dx' \, dy' \tag{2}$$

For computational purposes, assume there are N different distances $(r_i, i = 1, \cdots, N)$ in the scene. Using superposition, $f_2'(x, y, \lambda, t)$ can be decomposed into N components as:

$$f_2'(x, y, \lambda, t) = \sum_{i=1}^{N} f_{2i}(x, y, \lambda, t) \tag{3}$$

where

$$f_{2i}(x, y, \lambda, t) = \begin{cases} f_2'(x, y, \lambda, t), & \text{if } r(x, y) = r_i \\ 0, & \text{elsewhere} \end{cases}$$

Thus, Equation (2) becomes:

$$\begin{aligned} f_3(x, y, \lambda, t) &= \sum_{i=1}^{N} h'(x, y, r(x, y), \vec{e}) \star f_{2i}(x, y, \lambda, t) \\ &= \sum_{i=1}^{N} h_i(x, y, \vec{e}) \star f_{2i}(x, y, \lambda, t) \end{aligned} \tag{4}$$

where, $h_i(\cdot)$ is the point spread function for the planar object at distance r_i and $\star$ is the convolution operator. Note that, if the profile of the scene in a small field-of-view is smooth, we have $N = 1$ and

$$f_3(x, y, \lambda, t) = h'(x, y, \vec{e}) \star f_2'(x, y, \lambda, t)$$

```
Step 1: Decompose the object into $N$ planes, $f_{2i}(x, y, \lambda, t)$, of distance
            $r_i, i = 1, \cdots, N$, according to the depth map information;
Step 2: for $i = 1$ to $N$ do
        begin
            Compute and store the point spread function $h_i$
        end;
Step 3: $f_3 \leftarrow 0$;
        for $i = 1$ to $N$ do
        begin
            $f_3 \leftarrow f_3 + f_{2i} \star h_i$
        end;
```

Figure 6: Simulation algorithm for curved objects.

as derived in.[10] Therefore, the algorithm for the simulation of curved objects can be summarized as in Figure 6 where FFT algorithm can be applied in Step 3 to reduce the large amount of computations needed. The depth map information can be obtained from, *e.g.*, range scanner or the ray casting algorithm.[9]

3.2 Motion simulation

The motion parameters used in the simulation are specified by the vector $\vec{m} = [V_x \quad V_y \quad V_z] \quad \Delta t$, where V_x, V_y, and V_z are the velocity components of the motion; Δt is the duration of the motion. Here, we assume that the scene contains only rigid objects so that the object will not change its shape while it is moving.

For objects moving perpendicular to the optical axis, *i.e.* $V_z = 0$, the size of the objects in the scene will remain unchanged. However, part of the original image will move out of the field-of-view and will not appear in the image plane. This will also introduce other objects into the scene which are not in the original image. Therefore, the original input image must include the objects that may come into the camera's field of view due to motion. This problem can be avoided by assuming a dark background. When parts of the objects move out of the camera's field of view, the dark background appears in the field of view. Here, we use this approach for its simplicity and efficiency in memory management.

When an object moves toward or away from the camera, the objects in the scene will be enlarged or shrunk. Therefore, resampling must be done to compensate for this effect. In AVS, we use bi-linear interpolation to compute the value $g(m, n)$ from its four neighbors $f(i, j)$, $f(i+1, j)$, $f(i, j+1)$, and $f(i+1, j+1)$. The result is:

$$g(m, n) = a \cdot (m - i) + b \cdot (n - j) + c \cdot (m - i)(n - j) + d \tag{5}$$

where $i \leq m \leq i + 1$, $j \leq n \leq j + 1$, and

$$\begin{aligned} a &= f(i+1, j) - f(i, j) \\ b &= f(i, j+1) - f(i, j) \\ c &= f(i+1, j+1) + f(i, j) - f(i, j+1) - f(i+1, j) \\ d &= f(i, j) \end{aligned}$$

The simulation of an object moving with an arbitrary motion vector $\vec{m}$ is done by a shift operation if $V_x \neq 0$ or $V_y \neq 0$, and then a resampling operation if $V_z \neq 0$ to get the synthesized image. The algorithm is shown

```
Step 1: if $V_x \neq 0$ or $V_y \neq 0$ then
        begin
            Shift the object horizontally by the amount $V_x \Delta t$;
            Shift the object vertically by the amount $V_y \Delta t$
            Append dark background if part of the object is
                moved out of the view;
        end;
Step 2: if $V_z \neq 0$ then
        begin
            if $V_z > 0$ then
                move the object toward the camera and upsample;
            else /* $V_z < 0$ */
                move the object away from the camera and down-sample;
            Append dark background if the neighbor of the object
                appears in the view;
        end;
```

Figure 7: Simulation algorithm for moving objects.

in Figure 7. Note that, during up-sampling process, the image might be smoothed, while in the down-sampling process, some image details might be lost.

3.3 Stereo system

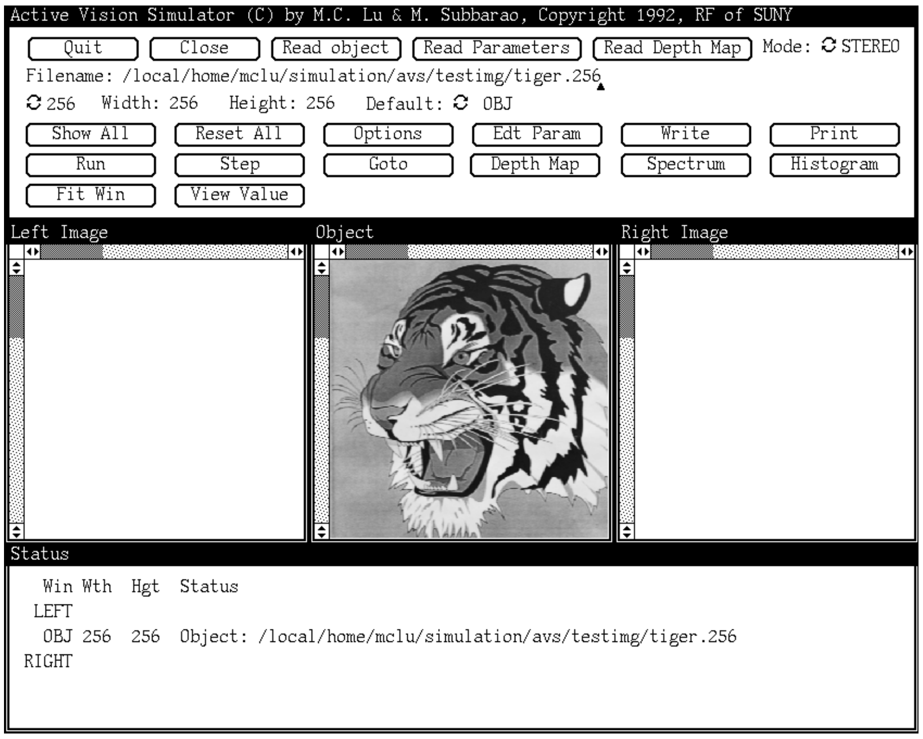
For a binocular camera system, the camera positions are specified by the vectors $\vec{O_l} = [x_l \quad y_l \quad z_l \quad \theta_{xl} \quad \theta_{yl} \quad \theta_{zl}]$ and $\vec{O_r} = [x_r \quad y_r \quad z_r \quad \theta_{xr} \quad \theta_{yr} \quad \theta_{zr}]$ with respect to the global origin O in Figure 2. The components of these vectors specify the positions and the orientations of the two cameras.

Assuming that the three coordinate systems parallel (i.e. $\theta_{xl} = \theta_{yl} = \theta_{zl} = \theta_{xr} = \theta_{yr} = \theta_{zr} = 0$ the stereo image pairs can be generated using the motion algorithm presented in Figure 7 where the motion displacement corresponds to $(-x_l, -y_l, -z_l)$ for the left camera and $(-x_r, -y_r, -z_r)$ for the right camera.

4 THE USER INTERFACES

Both graphical and non-graphical user interfaces are provided in AVS. The appearance and the basic functions of these user interfaces are similar to those in IDS.[7] AVS has all the functions of IDS plus one more window and some additional features as shown in Figure 8. Besides, the single parameter window in IDS is now three parameter windows in AVS – one each for the left and the right camera (camera parameters), the other for object-specific parameters as shown in Figure 9 .

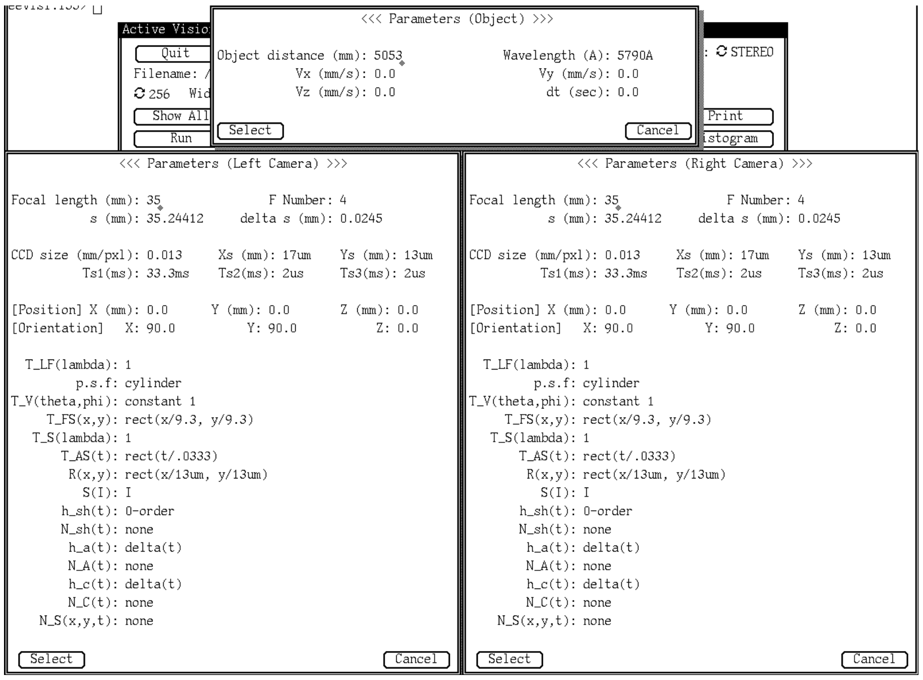
For curved object simulation, the depth map is read from a file by using the "Read DepthMap" command. The depth information stored in the file is a relative value, $\Delta r(x, y)$, with respect to the object distance u which is the shortest distance between the global origin O and the scene ($i.e.$, $\min\{d_i, i = 1, \cdots, N\}$). The object distance



```
Active Vision Simulator (C) by M.C. Lu & M. Subbarao, Copyright 1992, RF of SUNY
Mode: STEREO
Filename: /local/home/mclu/simulation/avs/testimg/tiger.256
256  Width: 256  Height: 256  Default: OBJ

  Win Wth  Hgt  Status
 LEFT
  OBJ 256  256  Object: /local/home/mclu/simulation/avs/testimg/tiger.256
RIGHT
```

Figure 8: AVS graphical user interface.



```
<<< Parameters (Object) >>>

Object distance (mm): 5053            Wavelength (A): 5790A
          Vx (mm/s): 0.0                    Vy (mm/s): 0.0
          Vz (mm/s): 0.0                     dt (sec): 0.0
```

```
<<< Parameters (Left Camera) >>>

Focal length (mm): 35            F Number: 4
          s (mm): 35.24412    delta s (mm): 0.0245

CCD size (mm/pxl): 0.013    Xs (mm): 17um    Ys (mm): 13um
        Ts1(ms): 33.3ms     Ts2(ms): 2us     Ts3(ms): 2us

[Position] X (mm): 0.0      Y (mm): 0.0      Z (mm): 0.0
[Orientation]   X: 90.0         Y: 90.0         Z: 0.0

  T_LF(lambda): 1
         p.s.f: cylinder
T_V(theta,phi): constant 1
     T_FS(x,y): rect(x/9.3, y/9.3)
   T_S(lambda): 1
       T_AS(t): rect(t/.0333)
        R(x,y): rect(x/13um, y/13um)
          S(I): I
       h_sh(t): 0-order
       N_sh(t): none
        h_a(t): delta(t)
        N_A(t): none
        h_c(t): delta(t)
        N_C(t): none
    N_S(x,y,t): none
```

```
<<< Parameters (Right Camera) >>>

Focal length (mm): 35            F Number: 4
          s (mm): 35.24412    delta s (mm): 0.0245

CCD size (mm/pxl): 0.013    Xs (mm): 17um    Ys (mm): 13um
        Ts1(ms): 33.3ms     Ts2(ms): 2us     Ts3(ms): 2us

[Position] X (mm): 0.0      Y (mm): 0.0      Z (mm): 0.0
[Orientation]   X: 90.0         Y: 90.0         Z: 0.0

  T_LF(lambda): 1
         p.s.f: cylinder
T_V(theta,phi): constant 1
     T_FS(x,y): rect(x/9.3, y/9.3)
   T_S(lambda): 1
       T_AS(t): rect(t/.0333)
        R(x,y): rect(x/13um, y/13um)
          S(I): I
       h_sh(t): 0-order
       N_sh(t): none
        h_a(t): delta(t)
        N_A(t): none
        h_c(t): delta(t)
        N_C(t): none
    N_S(x,y,t): none
```

Figure 9: Categorized parameters in AVS.

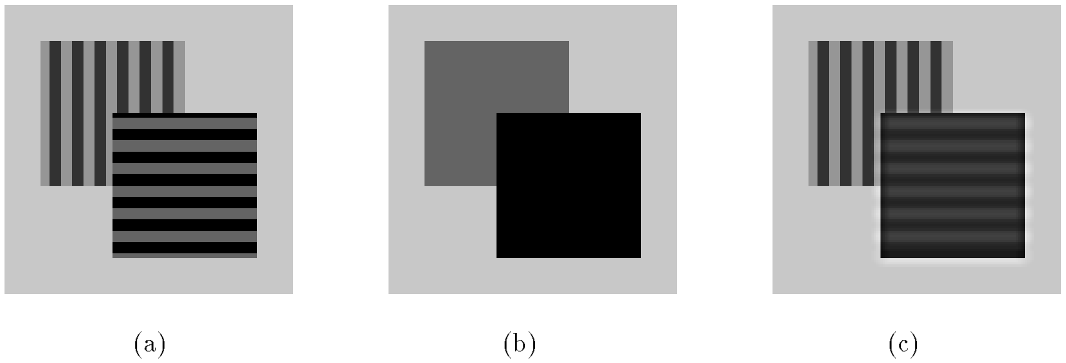


Figure 10: Simulated images for two planar boxes placed at different distances.

$r(x, y)$ is computed as

$$d(x, y) = \Delta r(x, y) \cdot k + u$$

where k is the scaling factor option in the option menu popped up by the "Option" command. The format of the depth map file is also specified in this menu.

When the depth map is loaded, depth information $r(x, y)$ can be viewed by using the "Depth Map" command which will pop up a window with depth profile. The value of the depth at each point can then be viewed on the screen by moving the mouse pointer to the desired location. In non-graphical user interface, the value is displayed according to the command line arguments used.

The parameters can be edited/viewed by the "Edt Param" command which searches the "Default" field for target window. The target can be object parameters, left camera parameters, or the right camera parameters as shown in Figure 9. The object parameters contain object distance, wave length of illuminating light, and V_x, V_y, V_z, dt for motion information. The camera parameters are basically the same as those in IDS except that (i) the object distance and wave length information are moved to the object parameters window; and (ii) the camera position and orientation information ($\vec{O}_r$, $\vec{O}_l$) are added.

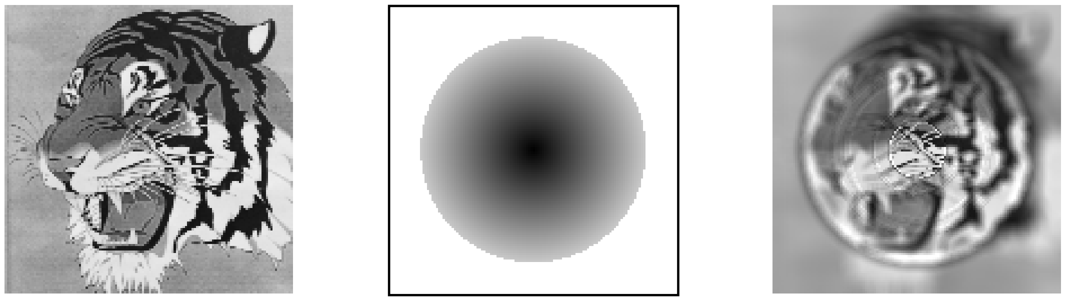
Another added feature is the "Mode" choice in Figure 8 which can be toggled between MONO and STEREO mode to simulate monocular and binocular image formation process. In "MONO" mode, "Left Camera" window will disappear. Therefore, the image will be synthesized in the "right camera" window by default. All the other commands are borrowed from IDS and carry the same functions.

5 SIMULATION RESULTS

5.1 Curved objects

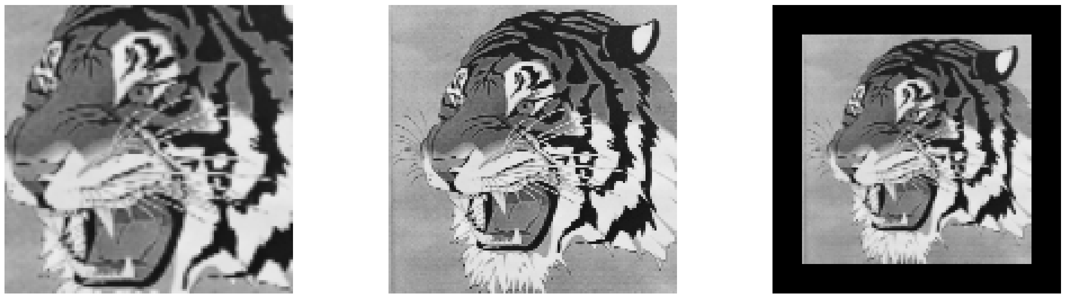
Figure 10 gives a simulated image of two striped boxes placed at two distances. The scene and the depth map are shown in Figure 10(a) and Figure 10(b), respectively. In Figure 10(b), the darker the appearance of the depth map, the closer the object is to the camera. The horizontal-striped box is located near the camera, while the vertical-striped box is located away from the camera. The camera parameters are adjusted to focus at the vertical-striped box. The resulting image is shown in Figure 10(C).

Another example is the tiger face placed on a cone-shape depth map as shown in Figure 11(a) and Figure 11(b),



(a) (b) (c)

Figure 11: Simulated images for object placed on a cone-shaped depth map.



(a) $\vec{m}_1$ (b) (c) $\vec{m}_2$

Figure 12: Resampled images in motion simulation.

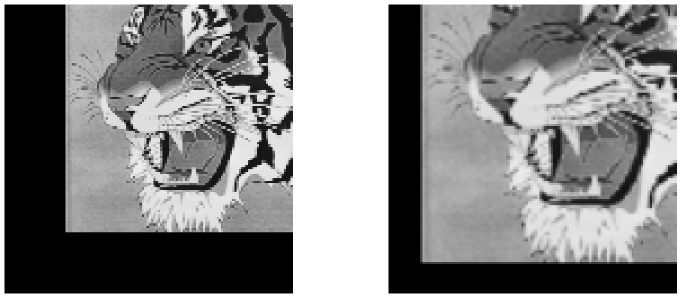
respectively. The depth range is from 2000mm to 3600mm (inside the cone) and the camera parameters are adjusted to focus at an object distance of 2000mm. The resulting image is shown in Figure 11(c). Note that, the depth outside the cone (white area) is assumed to be infinity. Therefore, a circle is visible in Figure 11(c).

5.2 Motion

Figure 12 shows the simulated images of moving objects. The center image is the original one. The left and the right images are generated with motion vector $\vec{m}_1 = [0 \quad 0 \quad -2.5]m/s \quad 1sec$ and $\vec{m}_2 = [0 \quad 0 \quad 2.5]m/s \quad 1sec$, respectively. All other parameters are the default ones shown in Figure 9.

The simulation of the shift operation (motion with $V_z = 0$) and the combined operation are shown in Figure 13 with $\vec{m}_3 = [100 \quad 100 \quad 0]m/s \quad 1sec$ and $\vec{m}_4 = [100 \quad 100 \quad 2.5]m/s \quad 1sec$. All other parameters remain the default ones.

Note that in Figure 12(c), Figure 13(a) and (b), dark background is introduced because the object is moved away from camera or part of the object moves out of the field of view as mentioned in Section 3.



(a) $\vec{m_3}$ (b) $\vec{m_4}$

Figure 13: Simulated images under shift operation and the general motion vector.

5.3 Stereo

The simulation of the stereo image pairs for the left camera position $\vec{O_l} = [-100mm \quad y_l \quad z_l \quad 0° \quad 0° \quad 0°]$ and the right camera position $\vec{O_r} = [100mm \quad y_r \quad z_r \quad 0° \quad 0° \quad 0°]$ is shown in Figure 14 where the first row is the image on the left camera, the second row is the image on the right camera. In Figure 14(a), $y_l = z_l = y_r = z_r = 0$ which corresponds to the shift operation; in Figure 14(b), the left lens is moved toward the object with $y_l = -100mm, z_l = 500mm$ while the right camera is moved toward the camera with $y_l = 100mm, z_l = -500mm$. The image resampling and the dark background effect are visible in these simulations.

6 CONCLUSION

In this paper, we have implemented the curved object, motion, and stereo image sensing simulation in a computer simulation package called Active Vision Simulator. AVS is a natural extension of IDS presented in our previous work.[10] It can be used to synthesize the images for research on image restoration, motion analysis, depth from defocus, and algorithms for solving the correspondence problem in stereo vision area.

The efforts spent on extending the IDS to AVS is limited – two added modules on motion and stereo, and some changes in the user-interfaces – because of the modular design and embedded extensibility of our original design of IDS. Again, AVS can also be easily extended if needed and can be used by other researchers on the verification of various vision theories.

7 REFERENCES

[1] S. T. Barnard and M. A. Fischler, "Computational Stereo," *Computing Surveys*, Vol. 14, No. 4, Dec. 1982.

[2] P. Bouthemy and P. Lalande, "Motion Detection in an Image Sequence Using Gibbs Distributions," *Proceeding of IEEE International Conference on Acoustic, Speech, and Signal Processing,* pp. 1651-1654, May 1989.

[3] B. K. P. Horn, *Robot Vision,* McGraw-Hill, New York, 1986.

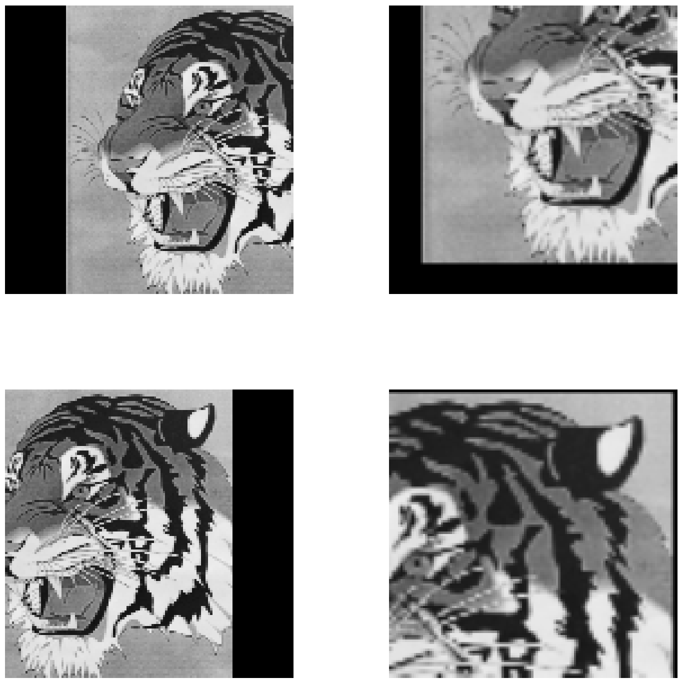


Figure 14: Simulation of stereo image pairs.

[4] J. Konrad and E. Dubois, "Bayesian Estimation of Motion Vector Fields," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* Vol. 14, No. 9, pp. 910-927, Sep. 1992.

[5] E. P. Krotkov, *Exploratory Visual Sensing for Determing Spatial Layout with an Agile Stereo Camera System,* Ph.D. dissertation, Department of Computer and Information Science, University of Pennsylvania, PA, April 1987.

[6] Y. Liu and T. S. Huang, "A Linear Algorithm for Determining Motion and Structure From Line Correspondences," *Computer Vision, Graphics, and Image Processing,* Vol. 44, No. 1, pp. 35-57, 1988.

[7] M.-C. Lu and M. Subbarao, "Image Defocus Simulator: A Software Tool," Technical Report No. 92.05.27, Computer Vision Laboratory, Department of Electrical Engineering, State University of New York, Stony Brook, May 1992.

[8] D. Murray and B. Buxton, "Scene Segmentation From Visual Motion Using Global Optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-9, pp. 220-228, March 1987.

[9] S. D. Roth, "Ray Casting for Modeling Solids," *Computer Graphics and Image Processing,* **18**, pp. 109-144, 1982.

[10] M. Subbarao and M.-C. Lu, "Computer Modeling and Simulation of Camera Defocus," *Proceedings of Optics, Illumination, and Image Sensing for Machine Vision VII,* SPIE Vol. 1822, pp. 110-120, Boston, Nov. 1992)

[11] J. Weng, T. Huang, and N. Ahuja, "Motion and Structure from Line Correspondences: Closed-Form Solution, Uniqueness, and Optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 14, No. 3, pp. 318-336, March 1992.