Synthetic Ground Truth Dataset to Detect Shadows Cast by Static Objects in Outdoors

Cesar Isaza
Instituto Politecnico Nacional
cisazab0900@ipn.mx

Joaquin Salas
Instituto Politecnico Nacional
jsalasr@ipn.mx

Bogdan Raducanu
Computer Vision Center
bogdan@cvc.uab.es

ABSTRACT

In this paper, we propose a precise synthetic ground truth dataset to study the problem of detection of the shadows cast by static objects in outdoor environments during extended periods of time (days). For our dataset, we have created a virtual scenario using a rendering software. To increase the realism of the simulated environment, we have defined the scenario in a precise geographical location. In our dataset the sun is by far the main illumination source. The sun position during the simulation time takes into consideration factors related to the geographical location, such as the latitude, longitude, elevation above sea level, and precise image capturing day and time. In our simulation the camera remains fixed. The dataset consists of seven days of simulation, from 10:00am to 5:00pm. Images are captured every 10 seconds. The shadows' ground truth is automatically computed by the rendering software.

Keywords

synthetic ground truth dataset, static objects shadow detection.

Keywords

Synthetic ground truth dataset, sun position, shadow detection.

1. INTRODUCTION

The problem of shadow detection has been studied extensively by researchers to improve the performance of computer vision algorithms, including segmentation [11, 12], recognition [14] and tracking [2, 3]. The problem has been shown to be difficult as the nature of the shadows is influenced by factors such as their dimensions, shape, orientation, location, date, time and translucent features of the object. In our research, we are particularly interested in studying the shadows cast by static objects in outdoor environments, such as buildings, lamp posts, traffic signs, among others. Assuming that during daytime the main light source is the sun, the shadows cast by static objects is different to the one cast by moving objects [4, 8] by the fact that most of the time, except before dusk or after dawn, the shadows move very slowly almost imperceptibly. Thus, depending on the scenario, the presence of shadows may represent a challenge and a problem to overcome by computer vision methods.

Our research focus on the automatic detection of shadows cast by static objects in outdoors as the scene is viewed for extended periods of time (days, weeks), from a fixed camera. In this context, a very important task to either compare or develop an efficient algorithm corresponds to measuring its performance. However, to evaluate the efficiency of a method, ground truth data is required. Even in trivial scenarios, building a ground truth dataset by hand is a daunting task that demands a lot of effort. Some of the aspects that make it difficult include the large number of frames needed, the dynamics of the scenario, the translucent properties of some objects, that in some cases project shadows without well-defined boundaries, and large shaded regions.

In the literature, there are several datasets that include the ground truth images to study shadows cast by moving objects such as cars or pedestrians, like the ones used by [7, 1, 13]. Nonetheless, none of these sequences have shadows cast by static objects in outdoors during long intervals of time. Actually, the aforemention datasets include events lasting for a few seconds. Another dataset to study the problem of illumination normalization for video surveillance was proposed by Matsushita *et al.* [5]. The authors consider synthetic images with changes in the position of the shadows cast by static objects in outdoors. In their dataset the shadows' ground truth images are not provided.

In this article, we propose a ground truth dataset of synthetic images to study the detection of shadows cast by static objects in outdoor environments during extended periods of time (days). The synthetic images have the potential to overcome the previous described difficulties for the detection of the ground truth by hand, as this process is done automatically by a computer. In our approach, we have created a virtual scenario using a rendering software. To increase the realism of the simulated environment, we have defined the scenario in a precise geographical location. In our dataset the sun is by far the main illumination source. The sun position during the simulation time takes into consideration factors related to the geographical location, such as the latitude, longitude, elevation above sea level, and precise image capturing day and time. In our simulation the camera remains fixed. The dataset consists of seven days of

VIGTA'12, May 21 2012, Capri, Italy.

simulation, from 10:00am to 5:00pm. Images are captured every 10 seconds. The shadows' ground truth is automatically computed by the rendering software. More realistic scenarios could include higher order interreflections, variable reflectance properties for the objects in the field of view, and light source attenuations due to clouds and fog. To render the images we used the *POV-Ray* software [6]. The synthetic shadow ground truth image is obtained computing the difference between two rendered images: One with shadows and the other one shadowless.

The rest of the paper develops as follow. In section 2, we described how the sun position is computed. Next, in section 4, we describe how the image ground truth dataset is obtained automatically. Finally, we summarize our contributions and conclude our article.

2. COMPUTING THE SUN POSITION

In our outdoors scenarios, we assume that the sun is by far the main illumination source. Therefore, to simulate the effect of shadows cast by static objects in a rendering software, we need to compute the sun position as the time passes by. To obtain these values, we assume that the virtual scenario is centered in a specific geographic location where latitude, longitude, and elevation above sea level, are provided. Then, the sun position in a real scenario at a given location, time and date for long intervals of time (days) is computed using the *Matlab* implementation [10] of the Reda and Andreas [9]. This method computes the sun position with an uncertainty of $\pm 0.0003°$ in the azimuth and elevation angles.

Table 1: Geographical location of the scenario. Latitude, longitude, and elevation have an uncertainty of $\pm$ 4m.

Latitude	Longitude	Elevation
20.57391**N**	100.37055**O**	1857 m

The sun position over time can be described by the azimuth and altitude angle vectors. The azimuth (measured from the north geographic pole in the horizontal plane) and altitude (measured from a line on the horizontal plane in the direction of the azimuth toward the zenith) angles vectors are projected to the rendering software coordinate system $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ by using an orthogonal decomposition (see Figure 1). In Figure 1 α is the azimuth and $\mathbf{A}_x$, $\mathbf{A}_z$ their corresponding projections to $\mathbf{X}$ and $\mathbf{Z}$ axes. β is the altitude, and $\mathbf{A}_y$ is its projection to the $\mathbf{Y}$ axis.

To obtain the geographical location, we used an off-the-shelf GPS (see Figure 2(a)). Its altitude, latitude and elevation are presented in Table 1. Figure 3 illustrates an example of the projected azimuth and altitude angles values corresponding to the sun position on October 26th 2011 from 11:32 to 16:38 hours. Another parameter to consider for the rendering process is the distance between the light source and the object in POV-Ray. The light source in the POV-Ray system requires a scalar value to simulate the distance between the light source and a given object. This distance defines the effect of parallel rays from the light source. To achieve parallel cast shadows from the simulation, the parameter should be large. We used *8.000*, *7.200* and *-6.500* as scalar values to the $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ axes respectively. After the synthetic images with shadows are rendered, the process to obtain the ground truth is applied. It may be worth to underline the fact that using the procedure we describe it is possible to simulate accurately the sun position of any geographical location, date and time for long periods (days, weeks, and months).

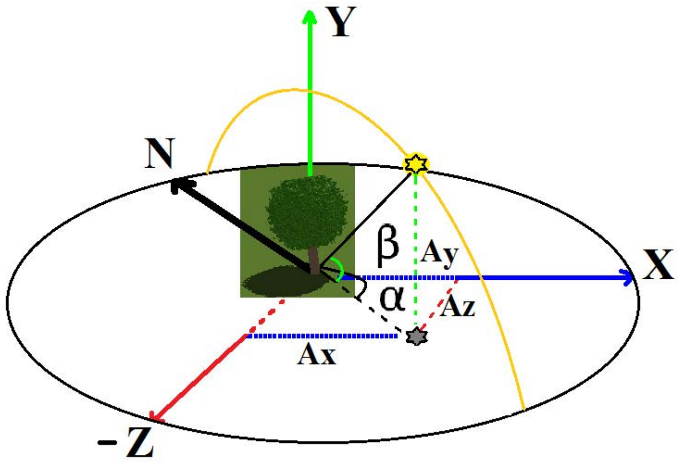
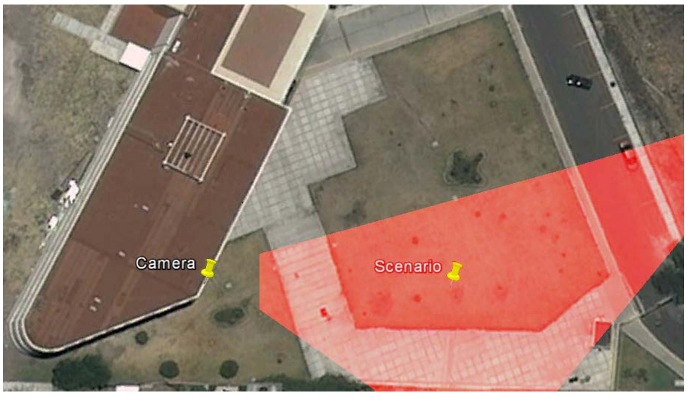
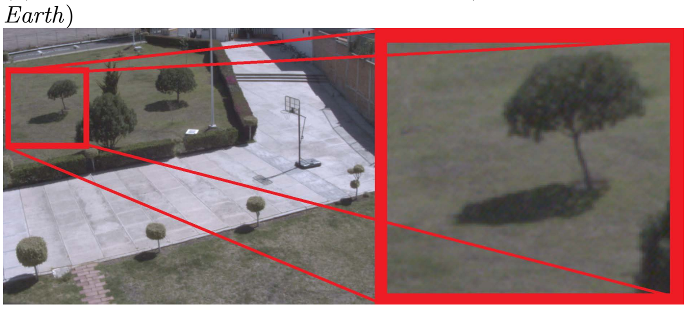


Figure 1: POV-Ray coordinate system and its relationship with the azimuth (α referred to the geographical North -N), and altitude (β referred to the horizontal coordinate system), angles of the sun position. An arbitrary XYZ-reference system is super imposed on an object in the scenario.



(a) Topview of the scenario under study (*Image from Google Earth*)



(b) Region of interest from an image acquired on October 26th 2011.

Figure 2: Illustrative example of the sun position algorithm

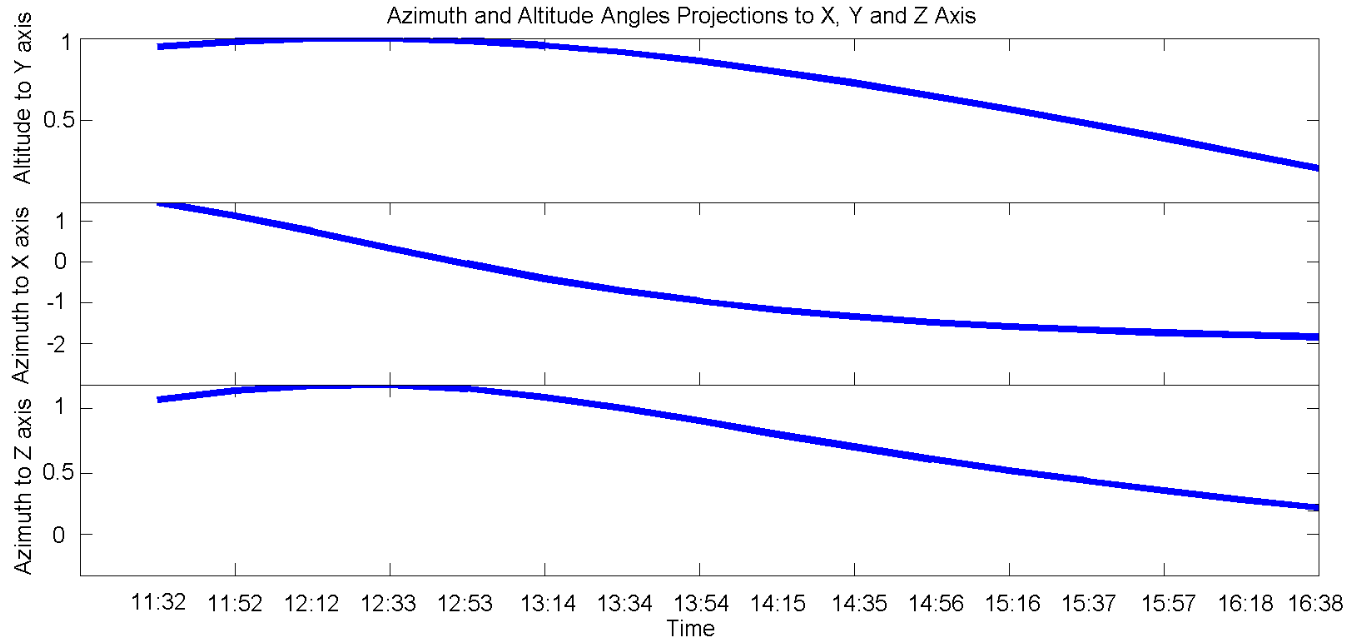


Figure 3: Projections of the azimuth and altitude angles to the X, Y and Z POV-Ray Axis, according to the sun position computed on October 26th 2011.

3. DATASET DESCRIPTION

An urban infrastructure environment is usually composed by buildings, streets, cars, trees and light poles. We considered the above elements to design and generate the virtual scenario in the rendering software. To create the dataset, we computed the sun position for one week. Then, the rendering process was achieved by changing the light source position into the *POV-Ray* coordinate system. We generated the images with insignificant changes in the position and shape of the shadows. To achieve the above, an interval of 10 seconds between frames was selected. Thus, each day is simulated by 2520 frames. We rendered seven consecutive days, deriving the total amount of 17640 images, including their corresponding ground truth. Each image has a resolution of 512×384 pixels (width×height), and were storage in *BMP* file format. However, it is possible to obtain images at higher rate. Figure 6 shows in the first and second rows examples of rendered images. In the third and fourth row the ground truth, while in the fifth and sixth the corresponding highlighted images.

4. BUILDING THE DATA SET

The procedure to automatically estimate the shadow ground truth image is based on a property of the rendering software, which has the capability to generate an image with or without shadows. Thus, we generated two images for each sun position at a given time; one with shadows and other without them. Then, the difference between the generated images is computed and the shadow ground truth is obtained. Figure 4 shows an example.

To make a qualitative comparison between the synthetic ground truth image and a real shadow cast by a static object in outdoors, a region of interest was selected from a frame of our real scenario of study. We analyzed the behavior of a real shadow in the location and time, against the derived

Figure 4: Example images. a) Image with shadow, b) shadowless, and c) shadows ground truth.

synthetic shadow. Figure 2(b) illustrates an image acquired on October 26th 2011 and a region of interest. We selected this region to illustrate the orientation of the shadow.

Figure 5 illustrates on the left column the shadow cast by a tree at 11:32, 12:12, 12:53, 13:43, 14:15, 14:56, 15:37 and 16:18 hours. The second column is the synthetic tree and the simulated shadow cast. The third and fourth columns are the ground truth images. The real ground truth was detected by hand, while the synthetic using the shadow and shadowless rendered images. The overlap shadow regions are presented in the fifth and sixth columns. In the same figure, it is possible to make a qualitative evaluation of the proposed method to obtain synthetic shadow ground truth images against the real shadow cast by the tree. In the synthetic overlap images it is also possible to see the self-shadow of the tree. This effect also appears in the real tree, but the detection of these small shaded regions by hand is a considerably difficult task.

Following the strategy presented, we generated a synthetic dataset for seven consecutive days, from October 27th to November 2nd, 2011. We add several objects to the scenario to define an urban infrastructure environment with building structures, a street and a big electricity tower. Figure 6 shows a sequence of images representing the shadow move-

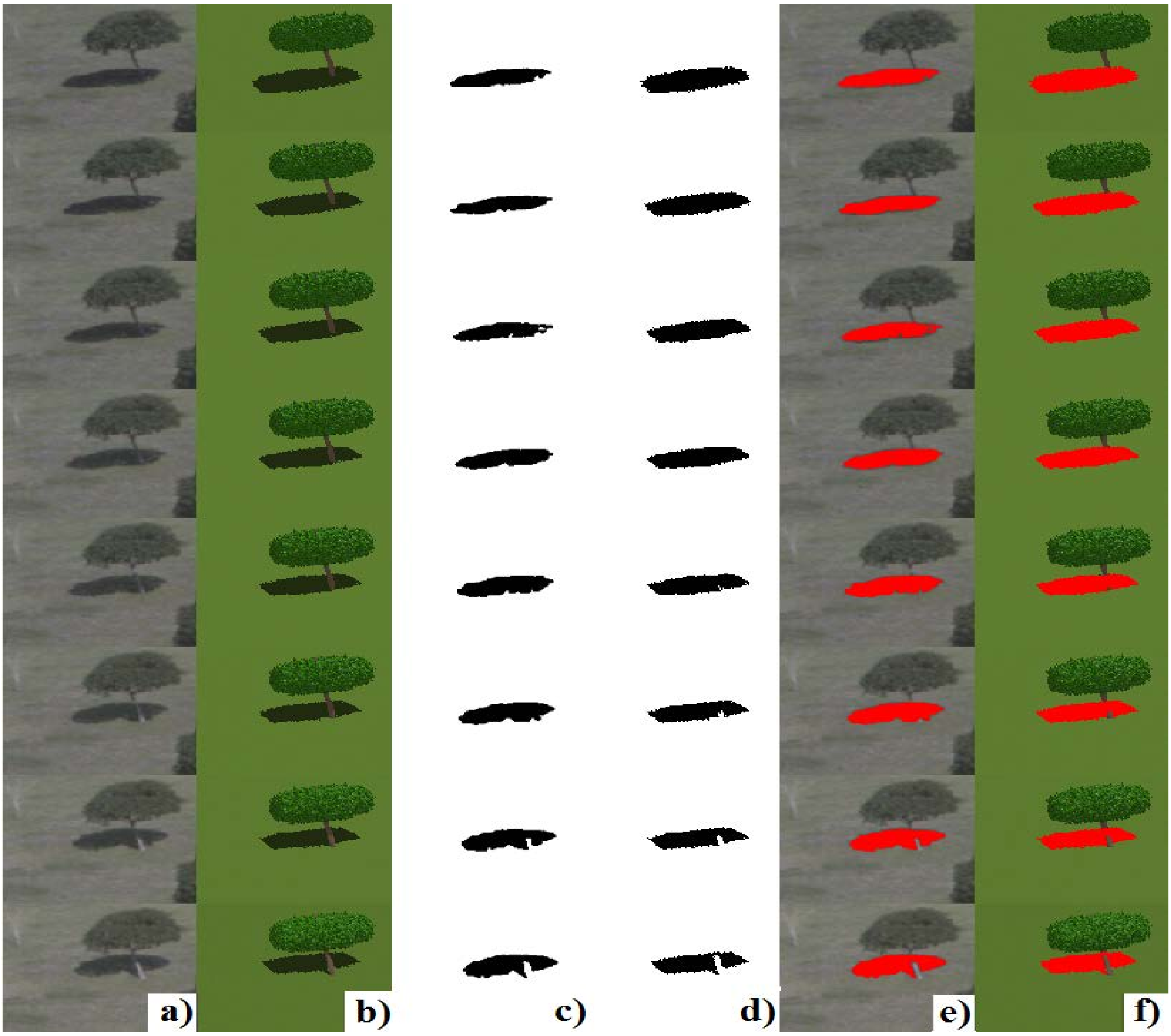


Figure 5: Real and synthetic inputs and their corresponding shadow ground truth and overlap images. a) Real image, b) simulated counterpart, c) shadow cast by the real tree, d) shadows in the simulated image, e) the shaded area is highlighted in the real image, f) the shades are highlighted in the simulated image.

ment throughout the all day. The first and second rows are images rendered with the sun position information of October 27th and November 2nd of 2011 respectively. This synthetic dataset simulates the sun position each 10 seconds during seven consecutive days.

5. CONCLUSION

In this article, we have presented a ground truth images to study the problem of detecting shadows cast by static objects in outdoors environments where by far the main light source is the sun. The strategy is sound, as it allows to overcome the time expenses and inaccuracies associated with the process of getting ground truth images by hand in complex scenarios. This dataset is important since the evaluation and comparison of algorithms to detect automatically shadows cast by static objects require ground truth data. We have made available several dataset of synthetic ground truth data at `http://imagenes.cicataqro.ipn.mx/shadows/`. We hope that the ability to develop, train and evaluate on our synthetic dataset will spur further progress on this field.

6. ACKOWLEDGEMENTS

This research was partially supported with a grant from IPN-SIP under grant contract 20121642

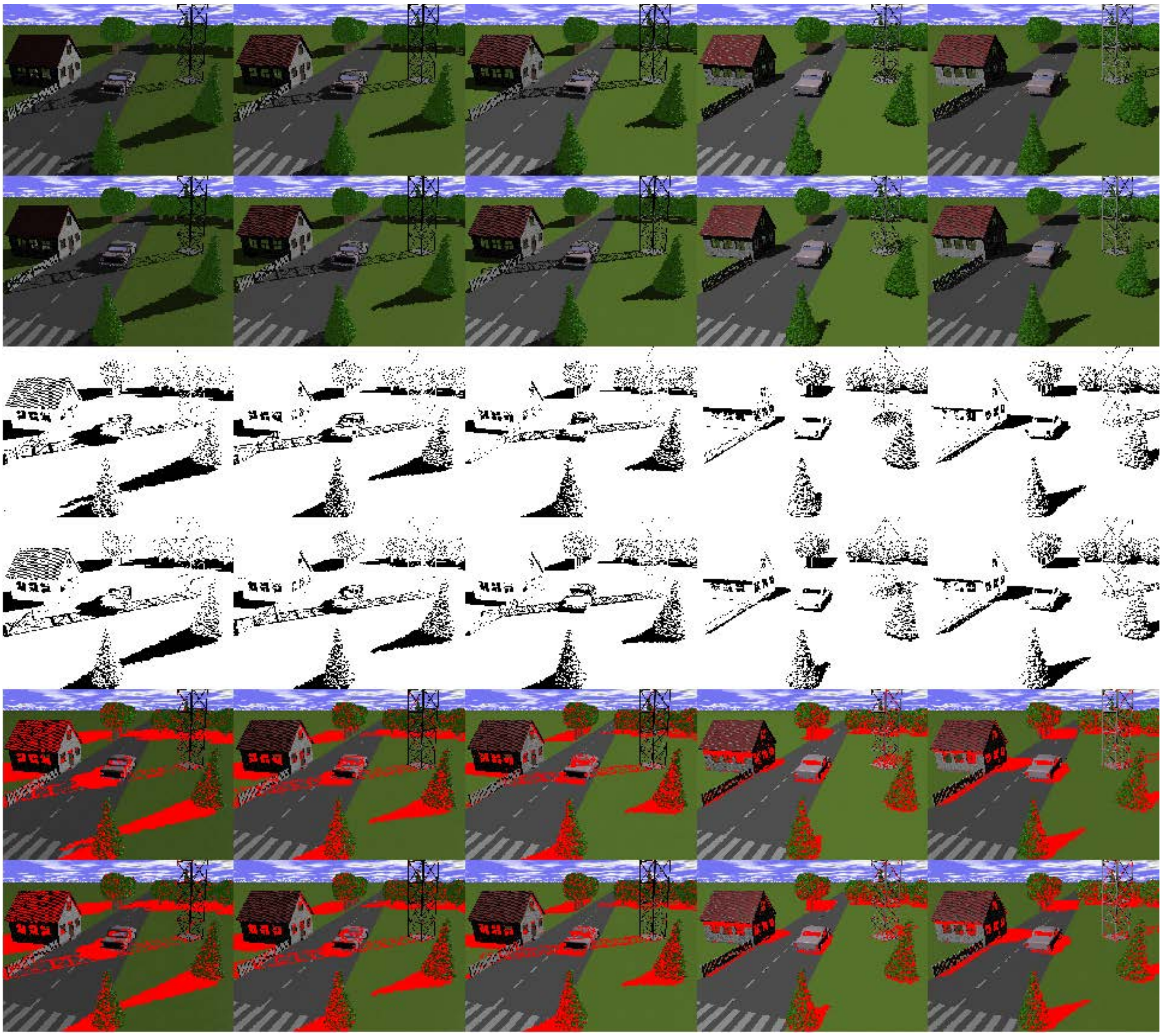


Figure 6: Synthetic ground truth dataset of shadows cast by static object in outdoors. Each column corresponds to the same scenario but highlighting a different characteristic. The first and second rows show the simulated scenario, the third and fourth one the shadows ground truth, the fifth and sixth one are highlight the shadows in the simulated.

7. REFERENCES

[1] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting objects, shadows and ghosts in video streams by exploiting color and motion information. In *International Conference on Image Analysis and Processing*, pages 360–365. IEEE, 2001.

[2] G. Foresti. Object recognition and tracking for remote video surveillance. *Transactions on Circuits and Systems for Video Technology*, 9(7):1045–1062, 1999.

[3] J. Hsieh, W. Hu, C. Chang, and Y. Chen. Shadow elimination for effective moving object detection by gaussian shadow modeling. *Image and Vision Computing*, 21(6):505–516, 2003.

[4] N. Martel-Brisson and A. Zaccarin. Moving cast shadow detection from a gaussian mixture shadow model. In *Computer Vision and Pattern Recognition*, volume 2, pages 643–648. IEEE, 2005.

[5] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi. Illumination normalization with time-dependent intrinsic images for video surveillance. *Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1336–1347, 2004.

[6] P. of Vision Raytracer Pty Ltd. Persistence of vision. `http://www.povray.org`, March 2012.

[7] A. Prati, I. Mikic, C. Grana, and M. Trivedi. Shadow detection algorithms for traffic flow analysis: A

comparative study. In *Intelligent Transportation Systems*, pages 340–345. IEEE, 2001.

[8] A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara. Detecting moving shadows: Algorithms and evaluation. *Transactions on Pattern Analysis and Machine Intelligence*, 25(7):918–923, 2003.

[9] I. Reda and A. Andreas. Solar position algorithm for solar radiation applications. *Solar energy*, 76(5):577–589, 2004.

[10] V. Roy. Sun Position. `http://www.mathworks.com/matlabcentral/fileexchange/4605`, March 2012.

[11] E. Salvador, A. Cavallaro, and T. Ebrahimi. Cast shadow segmentation using invariant color features. *Computer Vision and Image Understanding*, 95(2):238–259, 2004.

[12] J. Stander, R. Mech, and J. Ostermann. Detection of moving cast shadows for object segmentation. *Transactions on Multimedia*, 1(1):65–76, 1999.

[13] M. Trivedi, I. Mikic, and G. Kogut. Distributed video networks for incident detection and management. In *Intelligent Transportation Systems*, pages 155–160. IEEE, 2000.

[14] D. Xu, X. Li, Z. Liu, and Y. Yuan. Cast shadow detection in video segmentation. *Pattern Recognition Letters*, 26(1):91–99, 2005.