A Statistical Consistency Check for the Space Carving Algorithm.

A. Broadhurst and R. Cipolla
Dept. of Engineering, Univ. of Cambridge, Cambridge, CB2 1PZ
{aeb29 | cipolla}@eng.cam.ac.uk

Abstract

This paper investigates the use of the Space Carving algorithm with outdoor image sequences, using a lambertian lighting model. A new consistency function is proposed that uses a statistical comparison instead of the voxel centroid sampling that was initially proposed. This is important when there is more detail in the images than can be stored in a voxel representation. The new function is evaluated using synthetic data and real image sequences.

1 Introduction

Many different techniques have been applied to the problem of reconstructing three-dimensional shape from image sequences. These techniques all work well for constrained problems, such as small camera baseline [6, 1, 2, 8], or smooth curved objects [4]. Recently voxel based algorithms [5, 9] have been demonstrated which can reconstruct very complex shapes, but at the cost of large memory requirements. One of these algorithms is called Space Carving [5] (see section 2.1 for a description).

In this paper the effectiveness of the Space Carving algorithm will be tested on difficult outdoor images. These images have been chosen to show off both the benefits, and difficulties of Space Carving. The voxel based representation means that the algorithm can reconstruct complex 3-D shapes, that would not be possible with other approaches, however it does not use all the image information that is available. This means that the algorithm can be susceptible to noise, which is especially true of images taken outdoors in uncontrolled lighting conditions, and with uncalibrated images.

Many of the difficulties with Space Carving are due to the choice of the consistency check criterion which is used to decide whether to keep or discard a voxel. These voxels usually project to an area in the image which is larger than a pixel, so it is important to use all the image information, not just the information about the pixel closest to the centre of the voxel. In this paper we will use a statistical consistency check and we will show that this performs better than the existing method of sampling the centre of the voxel.

2 Background to the Space Carving algorithm

2.1 Background

The "theory of shape by space carving" by Kutulakos and Seitz [5] is a provably correct theory for the reconstruction of three-dimensional shape from multiple images. This section gives a brief summary of the work, but the reader is advised to refer to [5] for more detail.

The space carving algorithm starts with a volume of space that is larger than the object being reconstructed. At each iteration the volume is carved away until the resulting shape is consistent with the input images. A shape is said to be *Shape Photo-Consistent* if all points within the volume are consistent with the input images. A three-dimensional point within the volume is *Point Photo-Consistent* with an image point, if its colour could have resulted from the radiance of the 3D point. This is valid for lighting models which are locally computable (such as the Lambertian model). When more complex models are used, then a three-dimensional volume is *Shape-Radiance Photo-Consistent* with an image point if the colour of the image point could have resulted from the shape of the volume under the current lighting model.

The algorithm is supplied with an initial volume (represented by an array of voxels) and a *Consistency Check Criterion*. The purpose of the consistency check is to decide whether there exists a radiance value which could be assigned to a point in space so that it is consistent with the input images. At each iteration the consistency check is computed for an exterior voxel, and if consistent, is assigned the radiance value, otherwise it is removed from the model. This process is repeated until no more voxels can be removed, and what remains is a three-dimensional shape called the *Photo Hull*. Kutulakos and Seitz proved that if this shape was found, then it must be *Shape Photo-Consistent* with the input images.

In practice, a lambertian lighting model is used. This means that the consistency check function only needs to compare the RGB values. A voxel is accepted if the RGB variance is less than a threshold. The projected RGB values are obtained by projecting a voxel into an image, and sampling the image at the voxels centre. In this paper it will be shown that this is an oversimplification and that a statistical comparison produces better results.

2.2 Analysis

In this section we will test the space carving [5] algorithm using two very difficult image sequences. The first sequence (see figure 1) consists of seven uncalibrated images of the fountain in Great Court, Trinity College, Cambridge . This sequence is difficult to reconstruct because the fountain has a very complex shape with large amounts of occlusion. The second sequence of Great Court (see figure 3) is difficult to reconstruct because it is a concave shape. The occluding contour gives very little information about the structure of the scene and is hence a good sequence for evaluating the consistency function.

The images were captured using a Fuji-700 digital camera, and about 20-30 point correspondences were entered manually. The images were projectively calibrated using [3], and were then upgraded to a metric calibration using [7]. The final solution was optimised using a bundle adjustment, and the reprojection error was 0.5 pixels.

The space carving algorithm [5] has two parameters that need to be chosen. The first is the size of the voxel array, and the second is the threshold for the consistency criterion. This experiment will show that it is not always possible to choose a value for the threshold that finds the correct shape, as the algorithm has a tendency to create holes in the model.

Figure 4 shows a number of different reconstructions for the Great Court image sequence, with different values of the threshold. Notice how the windows have been removed. These errors result because the resolution in the image is higher than that of the voxel representation. The Space Carving algorithm is only provably correct if the images can be re-generated exactly from the voxel representation. By correctly sampling the voxels, the algorithm can take into account the different amounts of detail in the different images.

Figure 1: This image sequence shows the fountain in Great Court, Trinity College, Cambridge. This sequence has some very complex structures, and would be difficult to match with edges or surfaces. The hardest problem with this sequence is separating the fountain from the courtyard behind the fountain. The fountain has been manually segmented using The Gimp.

Figure 2: Results from the existing Space Carving algorithm using different threshold settings. The voxel array size was 128^3, and the thresholds were 48,32,24,16 (of 255) respectively. Notice how much of the shape was obtained from the occluding contour (left).

Figure 3: This image sequence was taken in Great Court, Trinity College, Cambridge. The basic structure of the sequence is fairly simple, but it is a difficult sequence to reconstruct using Space Carving as very little information is obtained from the occluding contour. This sequence will be used to demonstrate how well the algorithm is able to carve concave shapes.

Figure 4: Results from the existing Space Carving algorithm using different threshold settings. The voxel array size was 128^3, and the thresholds were 64,48,32,24,16 (of 255) respectively. Notice that it is not possible to chose a value of the threshold that finds the correct shape and preserves the windows.

3 The new statistical consistency function

3.1 How to estimate the statistics of a projected voxel

The problem with the existing consistency check criterion is that it does not correctly sample the image. This can been seen in figure 5. When a voxel projects to an area larger than a pixel, the information from all the pixels needs to be considered, not just the pixel at the centre.

The simplest approach is to smooth each of the images with a fixed NxN mask, but this is not the correct way of solving the problem as voxels project to different sizes, depending on how close they are to the viewer. A better way of sampling a voxel is to project its exterior boundary into each of the images, and determine which pixels are inside the projected shape.

The difficulty with projecting a complete voxel cube is that at the time of performing the consistency check, we do not know whether the voxel neighbours are solid or empty. We only know information about voxels which are closer to the viewer than the voxel being tested. It would hence be incorrect to project the side faces of the voxels as we would not necessarily know if they are valid or not. For this reason the front faces of all voxels are projected into the images, and this information is used to compute the statistics.

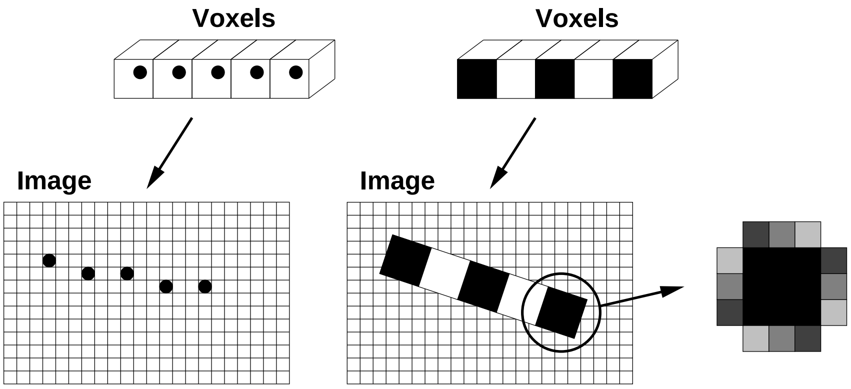


Figure 5: (a) The old centroid sampling method, (b) The new statistical technique.

3.2 The statistical consistency check

The problem that we have identified, is that there may be more information in the images than in the voxel representation, so we need some way of modelling this difference. Each voxel is coloured with a unique RGB value which is projected into an image. If the image contains more information than the voxel array, then this is modelled as Gaussian noise.

The consistency check is based on the F-statistic where the data is modelled as a fixed value plus noise. A voxel is only removed if there is sufficient evidence to suggest that the image samples could not have had the same mean, or in other words, they were not the reprojections of the same voxel. This is evaluated by computing the ratio of the between class variance, to the within class variance. The correct variance ratio for a confidence interval can be computed from the number of degrees of freedom in the two samples using the F-statistic.

This method has been suggested by Seitz and Dyer [9] as a possible extension to the Voxel Colouring algorithm, but it has not been analysed and has not been applied to the Space Carving Algorithm. In this paper we show that this is an important improvement to the Space Carving algorithm, especially when difficult image sequences are used.

3.3 A new method for determining the visibility of a voxel

When implementing the plane sweep algorithm, it is necessary to query each voxel and determine whether it is visible in each of the images. A voxel is visible in an image if there is a clear line of sight from the camera to the voxel. Or, in other words, all voxels closer to the camera have been carved away. This could be considered as an image based constraint. If there was a voxel closer to the camera, then it would have already been rendered into the image, as the algorithm works away from the camera centres. This means that to check visibility we only have to look in the image and see if the projected area is empty (similar to a Z buffer).

Algorithm 1 This algorithm implements a plane sweep in the x-direction without the need to perform ray casting in the voxel array. The visibility tests are all performed in the image plane.

```
for each iteration of the sweep plane: x = x_min to x_max do
  for each voxel in plane x do
    for each camera (which is behind the sweep plane) do
      Project the front face of the voxel into the image.
      if the region has not already been rendered then
        Compute the RGB statistics for the projected region
      end if
    end for
    Compare the statistics (if visible)
    if the voxel satisfies the statistical consistency function then
      Assign the voxel with the mean RGB value.
      Paint the front face of the voxel. (The same region as above.)
    else
      Remove the voxel from the model.
    end if
  end for
  for each voxel in plane x do
    Paint all the side faces
  end for
end for
```

4 Qualitative Results

The modified algorithm was applied to the same two image sequences that were demonstrated in section 2.2, and the results are shown in figures 6 and 7. The important thing to notice in figure 6 is that the windows have been correctly reconstructed, which was not the case in figure 4.

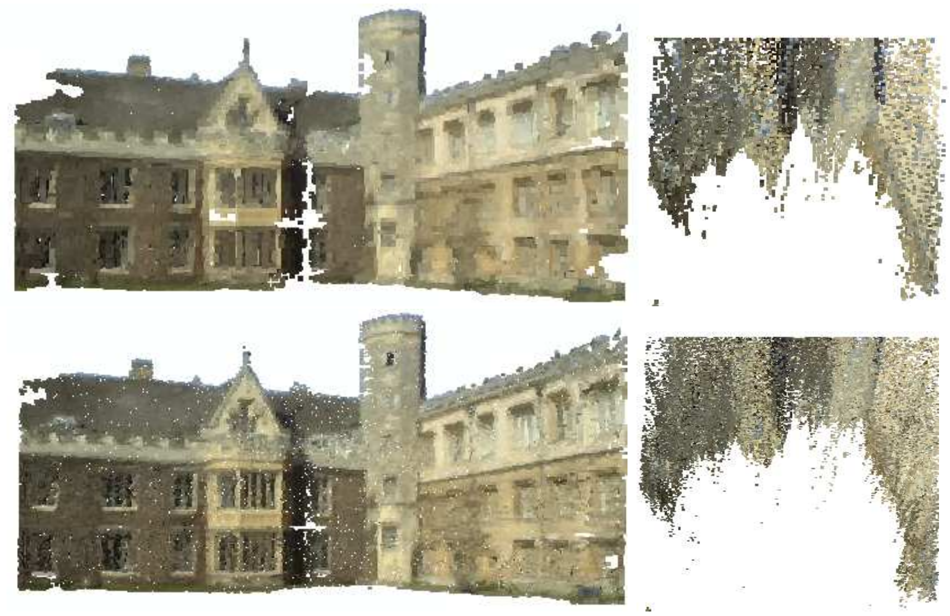


Figure 6: Space carving using the statistical consistency function. The top two images show a reconstruction of the Great Court image sequence with 128^3 voxels. The important thing to notice is that the windows have been correctly reconstructed, in contrast to figure 4. The lower two images use 256^3 voxels.

Figure 7: Space carving using the statistical consistency function. Compare these results with figure 2. This reconstruction uses 128^3 voxels.

5 Quantitative Evaluation

A synthetic sequence of images was used to test the statistical consistency function against the centroid sampling method used by Kutulakos and Seitz [5]. The test sequence (see figure 8) consists of a hollow unit cube with textured images on the back three faces. This configuration was chosen because it has a large hollow volume that has to be carved away, and the exterior boundary gives no information about the internal shape. All the voxels will have to be carved using the consistency function.

The graphs in figure 9 show the performance of the two algorithms with and without adding image noise. The *correct solid* plot shows the percentage of the voxels which are coloured as a fraction of those that should have been coloured. Likewise the *correct empty* plot shows the percentage of voxels that have been carved away, of those that should have been carved away. The *incorrect solid* plot shows the percentage of voxels that were not carved away from the region that should have been empty.

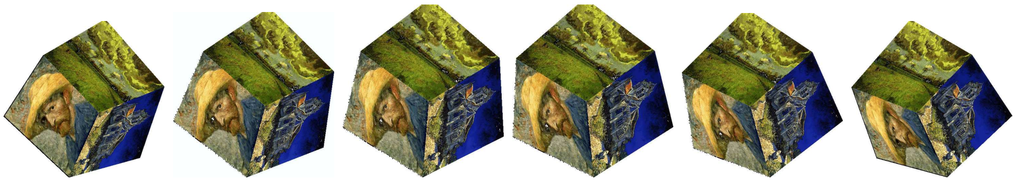
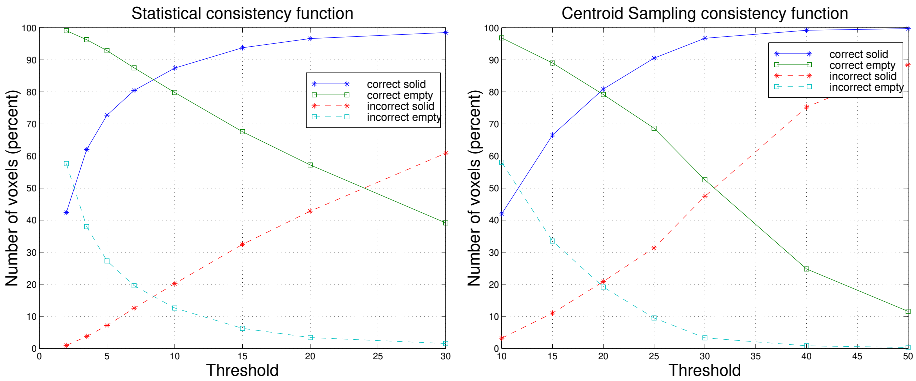


Figure 8: This is a synthetic image sequence of a cube textured with three well known paintings by Van Gogh (Self Portrait, Little Grey Church). The cube is rotated about a single axis. The important thing to notice about this sequence, is that the occluding contour does not give any information about the interior shape. All the interior voxels must be carved using the consistency function.

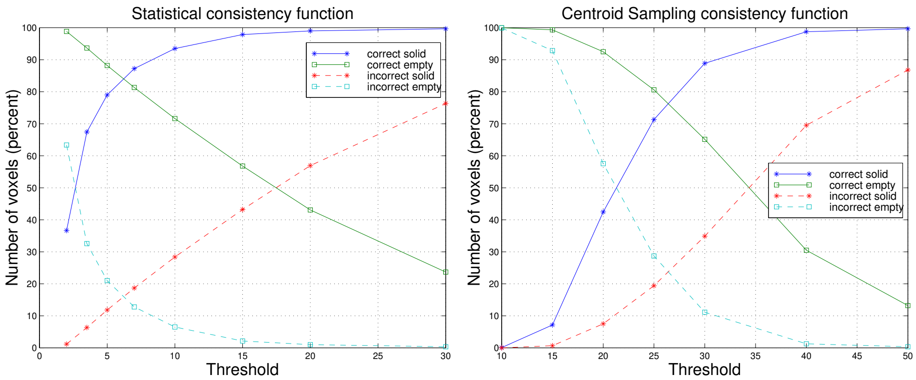
6 Discussion

The statistical method of comparing projected voxels is very important when dealing with outdoor image sequences. Outdoor sequences suffer from the difficulties of uncontrolled lighting and calibration errors which can be avoided in the lab. The statistical method has been shown to perform better under noise than the centroid sampling method which was previously used. In particular the statistical method copes better with voxel aliasing which was the cause of the windows to be removed in figure 4. The size of the window panes was smaller than the size of the voxels.

As a more general comment, the space carving algorithm performs very well on objects which have complex occluding contours, such as the Trinity fountain. (A quick initialisation is obtained by removing all voxels that project to the *background* in any of the images.) The algorithm performs less well on concave shapes where the shape is completely determined by the consistency function, and it is hence important to choose the best consistency function to achieve this task.



(a) These graphs were generated from synthetic data, with no added noise. Compare the intersection of the two *correct* curves which occur at 85% and 80%. Notice how the statistical method performs slightly better due to better handling of voxel aliasing.



(b) These two graphs show the effect of corrupting the synthetic images with Gaussian noise (sigma=10%). The centroid sampling method performs considerably worse than the statistical method in this case. Notice how the two *correct* curves cross at 85% and 75%. (not 85% and 80% as they did in figure 9(a)).

Figure 9: These graphs compare the performance of the statistical and centroid sampling consistency functions. Synthetic test data was used so that there would be no calibration errors, image noise, or illumination effects (see figure 8).

7 Future work

There is a possible extension to this algorithm which would save a large amount of computer memory. It was observed in section 3.3 that the image plane could be used to perform the visibility test. This image only needed to store the visibility (1-bit) but it could be used to store the RGB value of the voxel as well. This would mean that only 1-bit would need to be stored in the voxel array. This bit would store whether or not the voxel had been carved away. For high resolution voxel arrays this would save a large amount memory, but it would complicate the rendering process.

8 Conclusion

In this paper the Space Carving [5] algorithm has been tested using difficult outdoor image sequences. It has been shown that Space Carving can produce reasonable reconstructions from a relatively small number of uncalibrated images. A statistical consistency function has been used to replace the centroid sampling method used by Kutulakos and Seitz. A quantitative evaluation using synthetic images has shown that the statistical method is more robust to noise, which is important when dealing with uncalibrated images taken in uncontrolled environments.

Acknowledgements

This research is funded by Trinity College Cambridge, the Cambridge Commonwealth Trust and the Overseas Research Students (ORS) Award Scheme.

References

[1] S.T. Barnard. A stochastic approach to stereo vision. In *Proc. 5th National Conf. on AI*, pages 676–680, 1986.

[2] P.N. Belhumeur and D. Mumford. A Bayesian treatment of the stereo correspondance problem using half occluded regions. *Proc. Conf. Computer Vision and Pattern Recognition*, pages 506–512, June 1992.

[3] A. Broadhurst and R. Cipolla. Calibration of Image Sequences for Model Visualisation. In *Proc. Conf. Computer Vision and Pattern Recognition*, volume I, pages 100–105, 1999.

[4] R. Cipolla and A. Blake. Surface shape from the deformation of apparent contours. *Int. Journal of Computer Vision*, 9(2):83–112, 1992.

[5] K.N. Kutulakos and S. M. Seitz. A Theory of Shape by Space Carving. In *Proc. 7th Int. Conf. on Computer Vision*, pages 307–314, Kerkyra, Greece, September 1999.

[6] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, vol.194:283–287, 1976.

[7] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *Proc. 6th Int. Conf. on Computer Vision*, pages 90–95, 1998.

[8] S. Roy and I.J. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *Proc. 6th Int. Conf. on Computer Vision*, pages 492–499, 1998.

[9] S.M. Seitz and C.M. Dyer. Photorealistic scene reconstruction by voxel coloring. *Int. Journal of Computer Vision*, 35(2):1067–1073, 1999.