Limiting the Search Range of Correlation Stereo Using Silhouettes

Geoffrey Egnal, Max Mintz and Kostas Daniilidis
GRASP Laboratory
University of Pennsylvania
Email: {gegnal,mintz,kostas}@grasp.cis.upenn.edu

Abstract

We present a new approach to combine two approaches to three-dimensional reconstruction: silhouette-based and correspondence-based approaches. The two approaches have complementary costs and benefits. Silhouette-based approaches deliver volumetric descriptions which often have very few outliers, but they cannot reconstruct concave surfaces. Correspondence-based approaches give surface descriptions with sub-pixel accuracy, but their search range either allows outliers or falls short of the correct match. We show that a combination of the two can deliver fine-grained accuracy with few outliers. Our specific implementation uses the silhouette reconstruction as prior data to center and bound a stereo search process. We explore the different performance characteristics of the combination and its two component methods qualitatively and quantitatively using real imagery.

1 Introduction

Silhouette and correspondence reconstruction approaches have both been successfully used for visual reconstruction. Despite their success, there is still a need for more accurate reconstruction. Aside from inherent limitations to the accuracy of the methods, some situations present different scenery and allow different camera configurations than is typically found for the two approaches. For instance, a system might only have 4-5 cameras, but which can be placed far apart. Moreover, some images in a system may have high texture while others have low texture. In some images, it may be easier than others to segment the foreground against the background. For these reasons, we present a new method that combines the two approaches. We show that the combination gives a more accurate reconstruction and operates under a larger range of camera configurations than either approach.

The first approach, based on correspondence, searches for matching points in two or more images. Using the resulting disparity information between the images, the method triangulates to build a 3D surface model of the scene. Correspondence approaches operate under two basic scenarios: uncontrolled environments and controlled environments. In uncontrolled environments, a single moving platform usually holds all cameras. Since the cameras are closely spaced relative to the scenery, correspondence approaches work well. In controlled environments, systems use correspondence to achieve highly accurate surface reconstructions. Some of these environments only permit a few camera locations; others are highly textured. In such cases, correspondence seems an ideal choice. Even though correspondence can produce high resolution reconstructions, often to sub-pixel accuracy, the approach does have problems. Among them, if the search range is too small, the correspondence search will not find the correct match, but if the range is too large, the matching process may find a similar match far away from the actual match.

The second approach uses silhouettes to build a volumetric model of objects of interest. Silhouettes from multiple cameras can each be projected out to the real world, forming a conic which bounds the possible volume of the object of interest. The intersection of these conics is a 'visual hull' around the actual 3D shape of the object of interest. The silhouette approach generally operates in controlled environments, where multiple cameras can be placed at large separations. When the objects of interest are easily segmented against the background and when there are many permissible camera positions, silhouette approaches are ideal. A benefit of silhouette approaches is that they exhibit few outliers because segmentation algorithms usually produce contiguous silhouettes. However, the silhouette approach does not provide insight into concave surfaces because concave surfaces are invisible to silhouettes. This fault ultimately limits the accuracy of the silhouette approach, especially when the system has few cameras.

We present a method combining the correspondence-based and silhouette-based approaches. The method uses the silhouette model from at least one camera as a starting point for the correspondence process for each pixel. The prior model limits the correspondence search and prevents outliers, but also keeps the search range large enough so that it covers the true match. In this report, we demonstrate that the combination method capitalizes on its two components' strengths to produce a high-resolution surface model with few outliers.

Much research has focused on the two reconstruction approaches. For stereo correspondence, one can find general discussions from a number of sources (e.g.,[3, 8, 9, 23]). Because stereo systems have historically been mounted on robotic platforms, the research into incorporating visual data from stable, remote cameras has been limited. In a somewhat similar approach to silhouettes, many stereo systems employ the edges around an object of interest to anchor the disparity map [18, 17]. Although this approach technically uses silhouettes, the cameras are too closely spaced to build an accurate 3D model for purposes other than view synthesis. Also, the silhouette edges are generally not consistent between views, limiting the utility of this approach. One interesting method that relates volumetric approaches with correspondence is [24], in which a stereo system searches a 3D support space for the best match. Another related approach, voxel coloring [20, 10], where multiple cameras use color consistency to carve 3D volumes, conceptually resembles the silhouette-based approaches in that it searches inward from a bounded volume, but algorithmically lies closer to correspondence approaches because it checks the similarity (color consistency) of a given pixel across multiple images.

A rich body of work describes silhouette research [4, 2, 13]. Much of the research can be divided into two groups: volumetric representations [6, 19, 21, 22, 5] and image based representations [15]. Other researchers have attempted to prove the limits of silhouette-based reconstruction [11, 12].

There has been a small body of work that combines the silhouette-based and correspondence-based approaches. Cross and Zisserman [7] and later, Mulayim and Atalay [16], have used a voxelized silhouette reconstruction to initiate a voxel-coloring search. Our approach differs in three ways: (i) we use image-based silhouettes rather than volumetric methods, which decreases computational time and space requirements, (ii) our stereo algorithm explicitly deals with holes in the silhouette, rather than only searching within the region of the silhouette, and (iii) the study provides an empirical evaluation from a generalized setting, analyzing reasons why the combination method works better than either component. Szeliski has suggested using the visual hull as an inequality for optical flow reconstruction [22], but has not since published results on the topic. We present a combination method that does not merely truncate stereo disparities outside the silhouette visual hull after the stereo results have been generated; our method uses the silhouette data to limit the search range during the search process.

In light of previous work, this report's main contribution is to (i) present a new method to that uses silhouette-based data as a prior for the stereo search process, (ii) demonstrate practical examples of this combination, and (iii) quantitatively and qualitatively show the improvement of the combination over either approach individually.

2 Implementation

To show the effectiveness of the combination approach, we compare the combination with its component methods. The two test cases represent the range of application for the combination. In one test, we simulate a restrictive situation which only allows a few cameras and where segmentation is difficult for a subset of those cameras. In the second test, we simulate a less restrictive situation where we can easily place cameras, and those cameras can easily segment the objects of interest against the background. More specifically, the first uses three cameras, one overhead camera placed orthogonally to two stereo cameras. In this case, the overhead camera is the only silhouette provider. The second case uses four cameras: an overhead camera, and a side camera, both placed orthogonally to two stereo cameras. In this case, all four cameras provide silhouette data for the stereo pair to use as prior information.

In both cases, the combination method relies on three pieces of prior information. We assume that the cameras are calibrated, and that the depth of the background is known beforehand. The prior background depth is given to the algorithm in the form of a hand-calculated disparity map. Further, we assume that the maximum dimensions of the subject's position are known. These assumptions are typical in many practical environments.

2.1 Three-Camera Implementation

When the system receives a single silhouette, in this case from an overhead camera, the combination method has seven steps:

1. Take the silhouette of the overhead image.
2. Find the 'front' and 'back' edges for the silhouette relative to the two camera centers.
3. Using the overhead calibration information, project the front and back silhouettes to the world frame.
4. Using the rectified stereo camera projection matrices, project the points from the front and back edges of the silhouette conic to a front and back pair of left and right rectified images.
5. Based on the horizontal difference between the projections of a given point, find the disparity for the front and back images.
6. The disparity search range extends from the back edge of the conic to the front edge of the silhouette conic, as well as a small area near the background (see Figure 1).

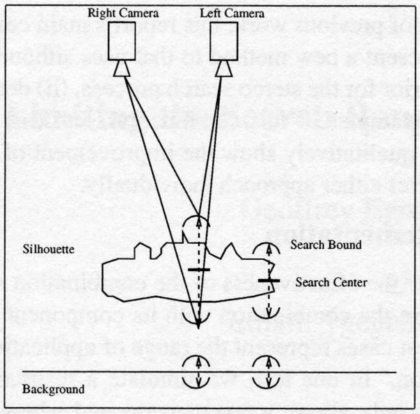

Figure 1: Detail of Combination Approach. The overhead image gives an estimate of the stereo search range and center. The figure displays a humanoid silhouette in front of a planar background. The dark line denotes the search center, and the dotted lines with arced endpoints denote the search bounds for both the frontal silhouette search and the background search.

7. Run stereo on the two horizontal images using these disparity intervals.

To accomplish the first step, finding the overhead silhouette, we use a simple background subtraction method, based on a single Gaussian distribution [14]. We take the mean and standard deviation of six background images. After subtracting the new image from this background, we threshold the results to find an initial foreground object. We then apply an erosion and a dilation morphological operation with a 5x5 kernel to sharpen the foreground object.

The second step, finding the 'front' and 'back' edges of the silhouette, requires calibration information from all three cameras. First, we project the two camera centers of the stereo pair onto the image plane of the overhead camera. Then, the algorithm starts from each point in the silhouette and traces two rays from the point on the silhouette towards each projected camera center. If both rays encounter other silhouette points before meeting the projected projection centers, then the point should vanish because it is fully-occluded. Otherwise, the point is visible to at least one camera and should remain on the front of the conic of silhouette points. Similarly, the farthest point from the centers of projection, along the same rays, lies along the 'back' edge of the conic.

The third step, projecting the front and back edges of the silhouette down to the world frame, uses the geometry of the overhead camera and knowledge of the maximal dimensions of the object of interest. The algorithm loops through each possible height value of the object, and applies the following formula:

$$P_w = M^{-1}m + \frac{Height - c_2}{M_2^{-1} p_{Silh}} M^{-1} p_{Silh})$$

where P_w is the final world point, M and m are components of the overhead projection matrix (see below), c_2 is the y coordinate of the center of projection of the overhead camera in world coordinates, M_2 is the second row of M, and p_{Silh} is an image point on the silhouette. The calibration process provides the relevant information, and the formula is derived by inverting the standard perspective transformation:

$$p_{Silh} \propto [\, M_{3x3} \quad | -m_{3x1} \,]\, P_w.$$

In the fifth step, we determine the stereo search interval. One of the problems with the overhead silhouette is that the object may have concavities or holes that are invisible to the overhead camera. For this reason, we cannot fully trust the overhead data, and so we search for a match from the back to the front of the silhouette, and then in a small interval near the background. Since the background and silhouette data have errors, we expand the search range by 3 pixels in each direction.

In our study, we employ a basic scanline-search stereo algorithm [3, 9]. We rectify the input images using Ayache's rectification algorithm [1]. Following rectification, the matcher calculates 7 by 7 windowed similarity values for integral shifts. Match goodness is measured using the modified normalized cross-correlation (MNCC) similarity metric:

$$\text{MNCC}(X, Y) \triangleq \frac{2\,\text{Cov}(X, Y)}{\text{Var}(X) + \text{Var}(Y)}.$$

We use the search range and center given by the prior silhouette data. The disparity is the shift of each peak correlation value. Subpixel peak localization is made via interpolation with a quadratic polynomial.

2.2 Four-Camera Implementation

As the second case includes more silhouettes, it requires a more-complex silhouette approach. Our silhouette reconstruction implementation is based on [15]. After finding the silhouette in each image with the same method as the three-camera case, we choose a reference image and project a ray out to the real world from each pixel in the silhouette on this reference image. In our tests, we choose the left stereo image as the reference image. We project quantized points on a segment of this ray to each of the other images. We trace the ray outwards, and label the front at the first point at which all projected points lie in the silhouettes of the other images. Presumably, this is the outermost surface of the volume we are modeling. We keep tracing the ray until no projected point lies in a silhouette. At this point, the ray has exited the volume of interest, and we call the first such point the 'back' edge. Since we need to compare the silhouette results with stereo surface results, we only reconstruct the surface and not the more typical volume model. To make the comparison simple, we record the surface as a disparity surface, calculated by projecting each world point to the two horizontally rectified images.

In the four-camera case, the combination method is essentially the same as the three-camera case. However, in the four-camera case, the stereo begins with a more refined, prior silhouette surface. In essence, this approach uses steps three through seven of the three-camera silhouette approach.

3 Results

3.1 Results for Three Cameras

The first test suite consists of a doll standing behind a pole, but in front of a board (see Figure 2). The doll is the subject of interest that we are trying to model. This test demonstrates the combination method in a more restrictive setting with only three cameras. Also, the overhead camera can segment the subject, but the stereo cameras cannot. The example is realistic in that the floor is often easier to segment than the horizontal background.

The images exhibit traits that are difficult for both correspondence and silhouette algorithms. Without prior background color information, the subject cannot be segmented horizontally. Also, the stereo search range is difficult to determine beforehand because the subject's location is unknown. With a large stereo search range, the low texture of the background will induce false matches. With a short search range, the disparity of the doll's chest and head might fall outside the disparity search range.

We compare the combination results against a standard stereo algorithm and a pure silhouette algorithm. The only difference between the combination method and the standard stereo is that the combination uses the overhead silhouette data to center and bound its search and only searches in areas where the silhouette cone gives a disparity estimate. Because the background is known at the time of design, even the standard stereo algorithm can center its search more accurately than about zero disparity. We set the search center to a positive shift of 26, corresponding to the average depth of the background.

The pure silhouette method uses all three images to build a disparity model. In contrast, the combination method only uses the overhead silhouette. To make the comparison more conservative against the combination method in this test, we generate the silhouettes by hand for the pure silhouette method. We quantitatively test all three algorithms at multiple search ranges.

Figure 3 demonstrates qualitative evidence that the combination works well. The disparity maps in the top row of the figure compare the three methods at a stereo search range of ± 20 pixels. The disparity map given by the combination method appears accurate to the eye. Comparing the results with the standard stereo results, we can see that the silhouette data helps the combination method. The chest

First Test

Second Test

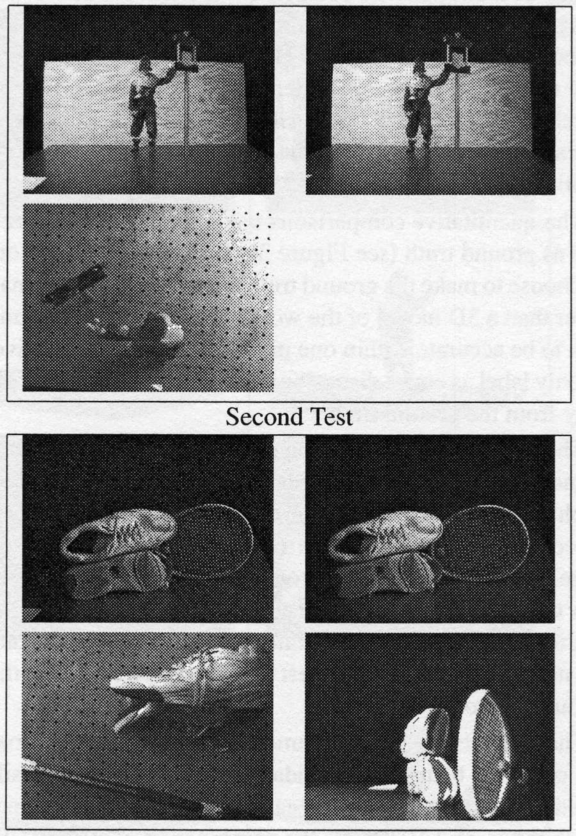

Figure 2: The Test Imagery. The first test consists of a doll figure behind a pole. The top row shows the left and right stereo pair, while the second row has the overhead image. At bottom is the second test, which consists of a pair of sneakers in front of a racquet. The top row shows the left and right stereo pair, while the overhead image and side image are at bottom.

and head are out of range for the standard stereo, but match well for the combination method. One should not compare the disparity maps in the half-occlusion region because all such matches are false by definition. Although the pure silhouette disparity map appears truthful, the values are actually incorrect by a few pixels in most cases.

At the second row of Figure 3, we display the reconstructed scene from a different vantage point. Since the pure silhouette method only reconstructs the doll figure, we only display the results in the area of the doll figure. The floor gives context and comes from ground truth disparity information. The holes in all of the images are due to local image expansion, and not to any matching error. These holes are easily filled using interpolation. We do not use any outlier elimination or occlusion detection to clean up the reconstruction.

Notice the face of the reconstructed doll. The standard stereo is unable to match pixels in this region, while the combination method performs well. The pure silhouette approach does not produce outliers, but the reconstruction lies along a conic surface. The error is noticeable in areas such as the doll's foot and head. These two surfaces are concave relative to the three cameras. Since the combination searches from the front of the silhouette to the background, concave surfaces fall within the search range and the combination is able to correctly reconstruct these areas.

The quantitative comparisons use a hand-built disparity map as ground truth (see Figure 3). For ease of generation, we choose to make the ground truth an integer disparity map rather than a 3D model of the world. We believe the ground truth to be accurate within one pixel, but to be conservative, we only label as errors disparities that are more than 2 pixels away from the ground truth.

The quantitative comparison compares results in a hand-segmented area around the doll figure (see Figure 4). The graph displays the percentage of possible matches found correctly as a function of the maximum allowable search range. The segmented area around the doll is a smaller region than that in which either the standard stereo or combination algorithm would operate, but produces an accurate statistic for the object of interest that is comparable with the silhouette algorithm.

The quantitative results demonstrate that the combination performs better than standard stereo and the pure silhouette method at all tested search ranges. At a maximal search range of ± 10 pixels, the combination achieves 92% accuracy, while the standard stereo has 0% accuracy for the object of interest. The doll does not stand in the standard stereo system's range. However, at a search range of ± 25 pixels, the standard stereo results rise to 90% accuracy, while the combination still achieves 92% accuracy. As the search range increases, the standard stereo system can find the correct match more easily and come close to the combination method's accuracy. As predicted, higher search ranges hurt the standard stereo system, allowing it to find false matches; at a search range of ± 80 pixels, the standard stereo drops to 88% acuracy. The combination does not change its behavior as the maximal allowed search range increases because the estimated disparity range does not increase to exploit this allowance. This smaller range decreases the probability that the search will fixate on a false local maximum. The pure silhouette approach consistently achieves 19.9% accuracy because many of the doll's features lie on concave surfaces relative to the three silhouette cameras.

Although the standard stereo always searches within the maximum range, the combination approach often does not exploit the larger search range, even when permitted to do so. When allowed to search up to ± 40 pixels, the combination method keeps itself within a maximum search range of ± 27 pixels, searching an average of only ± 8.85 pixels. Not only does the smaller search range save computation, but it increases accuracy.

3.2 Results for Four Cameras

The second test suite consists of a pair of sneakers in front of a racquet (see Figure 2). The sneakers are the subject of interest that we are trying to model. We use four cameras: one overhead, one to the side, and two frontal cameras. This test demonstrates the combination method in a less restrictive setting with more cameras placed at farther separations and where all cameras can segment the subject against the background.

The images exhibit traits that are difficult for both correspondence and silhouette algorithms. The shoes have concavities which the silhouette algorithms cannot pick up. Both the toe of the lower shoe and the shoe's tongue fall in a concave region relative to the silhouettes. The low texture around the sole of the lower shoe also causes correspondence difficulties.

Qualitatively, one can again see that the combination is useful. The standard stereo produces many outliers around the edges of the shoe, while the silhouette approach does not accurately reconstruct the toe of the lower shoe. As seen by the shoelaces in front of the toe of the lower shoe, the entire surface reconstructs at the same depth as the toe of the top shoe. The combination provides an accurate reconstruction with fewer outliers than standard stereo.

The quantitative results are similar to those of the three-camera example (see Figure 4). The graph displays the successful match percentage in the area of the shoes. The silhouette approach cannot find the exact disparity (within 2 pixels) and so only has a 38% hit rate. The combination consistently achieves above 71%, while the standard stereo always underperforms the combination. At a small search range of 10 pixels, the standard stereo cannot find the shoe and only matches 17% of the shoes' surface. Although coming close to the combination method at a range of ± 20 pixels, the standard stereo suffers with larger search ranges, finding only 67% of possible matches at a range of ± 90. The larger ranges increase the probability that the standard stereo may choose a false local maximum as the optimal match. As before, the combination does not increase its search range when allowed to do so. While the standard stereo searches a range of ± 60 pixels, the combination approach averages a search range of ± 11.76 pixels in the silhouette body and ± 3 pixels around the background.

3.3 Grouped Results

In order to understand how the combination compares to the simple stereo algorithm, we have grouped the errors for the algorithms into four categories. The first category, 'Out of Bounds', calculates when the stereo or the combination search interval did not cover the correct match. The second category, 'Near an Edge', calculates when an error occured

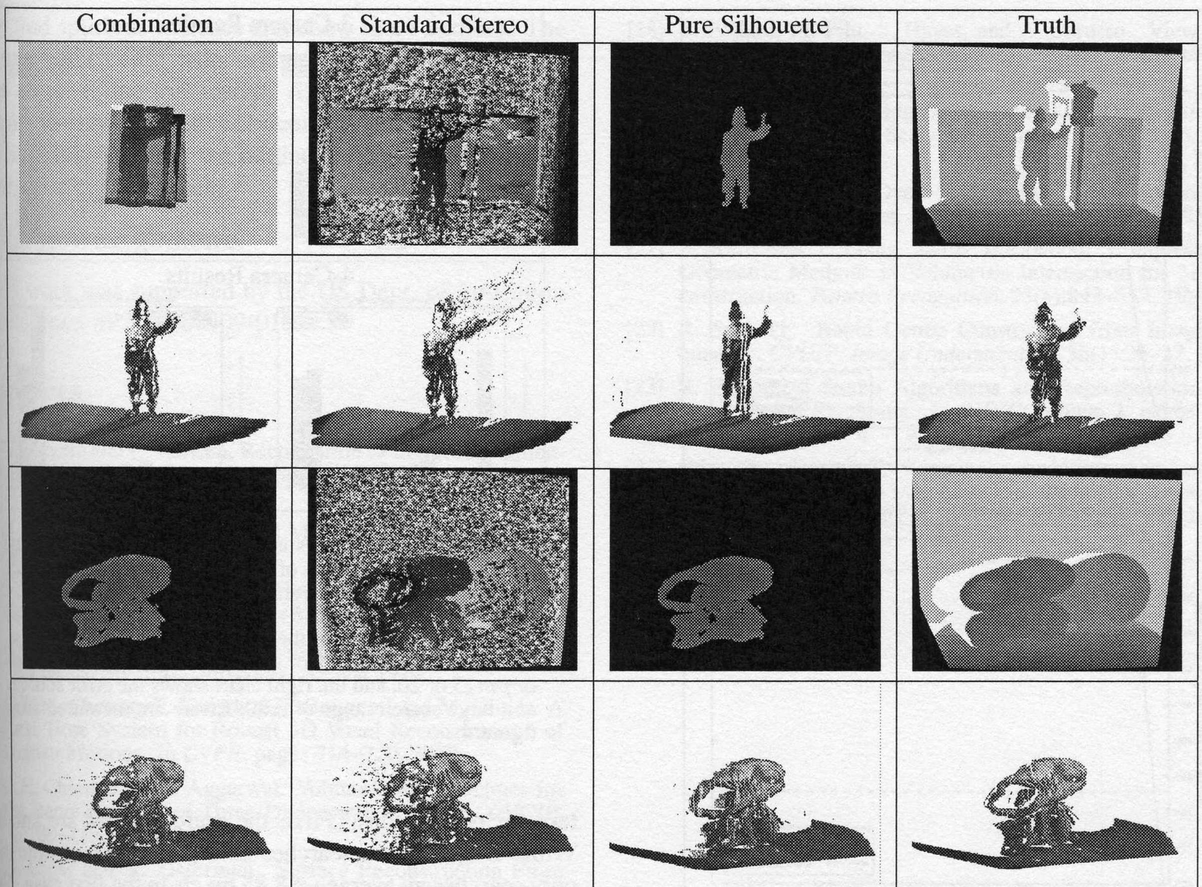

Figure 3: The Qualitative Results. The figure displays results for both the three- and four-camera cases. Within each display, the disparity results are at top. From right to left are the combination method, the standard stereo, the pure silhouette algorithm, and the ground truth disparity images. The bottom row displays the reconstructed points generated from the disparity image in the area of the object of interest. A floor has been added for context based on ground truth disparity information. All stereo results are generated with a maximal search range of ±20 pixels. The white pixels in the ground truth indicate half-occluded regions.

near an edge. One reason edges might cause error is that both the combination and the simple stereo use windows, and windows average the foreground and background surfaces and distort the matching results. The third category, 'Occlusions', labels matches in half-occluded areas as erroneous because by definition, there can be no disparity in a half-occluded region. The last category, 'Other', represents all other stereo error. This category includes errors due to low-texture, repetitive patterns, and lighting variation.

The grouped-error charts (see Figure 5) show why the combination improves over the simple stereo algorithm. At low search ranges, the simple stereo suffers many Out of Bounds errors. The combination avoids these errors using silhouette data. At higher search ranges, the simple stereo algorithm avoids Out of Bounds errors, but finds many false matches in the Other category. Since the combination avoids these errors, we may surmise that the simple stereo's larger search range allows a false, yet similar, match to enter the stereo algorithm's consideration.

4 Discussion

4.1 Algorithm Results

The primary driver behind the performance difference between standard stereo and the combination approach is that a correctly centered search range covers the matching region and allows the stereo system to find the right match. This behavior is clearly exhibited in Figure 4, where the standard stereo's small search range cannot find the correct match. A secondary driver is that a smaller search range decreases the probability that the combination chooses an incorrect local maximum. The correlation profile might have several local maxima, one or none of which may be correct. With a larger search range, the probability that the standard stereo chooses an incorrect local maxima increases. One can see this behavior in both quantitative results when the standard stereo performance decreases as the search range increases from ±25 to ±100 pixels. A tertiary driver is that the smaller search range limits the extent of bad matches, eliminating extreme outliers. Eliminating

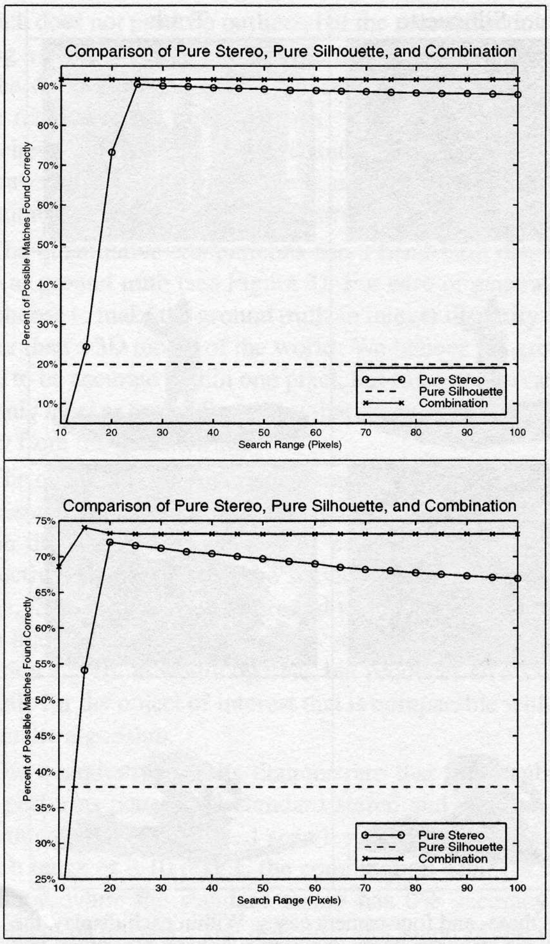

Figure 4: The Quantitative Results. The top plot displays the results for the doll figure, while the bottom plot displays the results for the shoes. The plots show the percentage of correct matches in the area of the object of interest as a function of the maximal search range (±10 to ±100). We measure a correct match as within 2 pixels of the ground truth (see Figure 3). Note that the combination method does not always expand its search range to fill the permissible search range.

even a small number of extreme outliers has a disproportionately beneficial effect on the qualitative visual appearance of the results.

The driver behind the poor performance of the pure silhouette method is that there are only three or four cameras, which is not enough to reduce the visual hull to within 2 disparity pixels of the truth. One advantage for the combination method is its ability to work well with fewer cameras. Concavities cause noticeable visual distortion for the pure silhouette method, such as the doll's right foot or the bottom shoe's toe. These errors are not due to segmentation error, but rather to the limitations of the silhouette process itself. One sees this because the first test case uses hand-segmentation.

The combination is faster than the standard stereo system for two reasons. First, the search ranges are smaller. While simple stereo searches ±40 pixels, the combination only searches an average ±8.85 pixels in the first case. Second, the overhead data limits any search of the stereo cameras to the areas where the silhouette's conic would project onto the image plane. This can save processing roughly 90% of the image in the first case and 40% of the image in the second case.

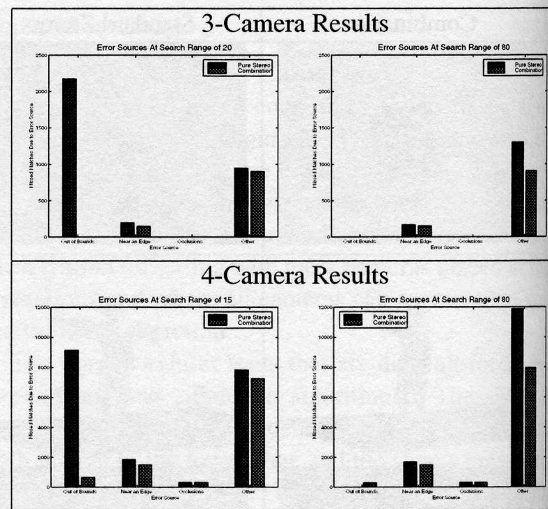

Figure 5: Error Sources at Various Search Ranges. The top row displays sources of error for the doll results, while the bottom row displays error sources for the shoe results. Each bar chart shows the number of missed matches due to each error source in the area of the object of interest. The left chart shows error sources at a small search range or *pm*25 or 20, and the right chart shows the error sources at a large search range of ±80. Errors are measured as in figure 4.

4.2 Summary

We have presented a method to combine silhouette-based and correspondence-based reconstruction approaches. Using practical examples with real-world assumptions of prior knowledge, we show that the combination delivers more accurate results than standard stereo or silhouette-based approaches. The stereo is able to complement the weaknesses of the silhouette-based approach and find small concavities and small features. The silhouette data helps the stereo in two ways: making sure the search range covers the correct match and limiting the search range to reduce the potential for matching far away from the truth. The silhouette data allows stereo to operate with large uncertainty about the potential search volume it may encounter; so, the system designer does not have to tune the system precisely to anticipate the subject's location.

We predict that other adverse conditions for stereo would further demonstrate the need for the combination approach. As the stereo baseline grows, the search range would ordinarily need to expand. At these larger ranges, the stereo would find spurious matches and show more outliers. The silhouette data could help center and keep the disparity ranges closer to the true match. Also, as the subject material starts to show repetitive structure, a larger search range will noticeably increase the outliers in the system. Again, the combination could help.

5 Acknowledgements

This work was supported by the US Dept. of Education under contract #P200A80409-01.

References

[1] N. Ayache and C. Hansen. Rectification of Images for Binocular and Trinocular Stereovision. In *ICPR*, volume 1, pages 11–16, 1988.

[2] H. Baker. Three Dimensional Modelling. In *IJCAI*, pages 649–655, 1977.

[3] S.T. Barnard and M.A. Fischler. Computational Stereo. *Computing Surveys*, 14(4):553–572, 1982.

[4] B.G. Baumgart. *Geometric Modelling for Computer Vision*. PhD thesis, Stanford University, August 1974.

[5] G.K.M. Cheung, T. Kanade, J.Y. Bouguet, and M. Holler. A Real Time System for Robust 3D Voxel Reconstruction of Human Motions. In *CVPR*, pages 714–720, 2000.

[6] C.H. Chien and J.K. Aggarwal. Volume/Surface Octrees for the Representation of Three-Dimensional Objects. *CVGIP*, 36:100–113, 1986.

[7] G. Cross and A. Zisserman. Surface Reconstruction From Multiple Views Using Apparent Contours and Surface Texture. In A. Leonardis, F. Solina, and R. Bajcsy, editors, *NATO Advanced Research Workshop on Confluence of Computer Vision and Computer Graphics, Ljubljana, Slovenia*, pages 25–47, 2000.

[8] U.R. Dhond and J.K. Aggarwal. Structure From Stereo – A Review. *Trans. Systems, Man and Cybernetics*, 19(6):1489–1510, 1989.

[9] O. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, Cambridge, Massachusetts, 1993.

[10] K.N. Kutulakos and S.M. Seitz. A Theory of Shape by Space Carving. *IJCV*, 38(3):199–218, 2000.

[11] A. Laurentini. The Visual Hull Concept for Silhouette-Based Image Understanding. *PAMI*, 16(2):150–162, 1994.

[12] A. Laurentini. How Far 3D Shapes Can Be Understood from 2D Silhouettes. *PAMI*, 17(2):188–195, 1995.

[13] W.N. Martin and J.K. Aggarwal. Volumetric Descriptions of Objects from Multiple Views. *PAMI*, 5(2):150–158, 1983.

[14] F.C.M. Martins, B.R. Nickerson, V. Bostrom, and R. Hazra. Implementation of a Real-Time Foreground/ Background Segmentation System on the Intel Architecture. In *IEEE ICCV99 Frame Rate Workshop*, 1999.

[15] W. Matusik, C. Buehler, R. Raskar, S.J. Gortler, and L. McMillan. Image-Based Visual Hulls. In *SIGGRAPH*, pages 297–306, 2000.

[16] A.Y. Mulayim and V. Atalay. Multi-Baseline Stereo Correction For Silhouette-Based 3D Model Reconstruction From Multiple Images. In *Three Dimensional Image Capture and Applications IV*, volume 4298, pages 143–154. SPIE, 2001.

[17] J. Mulligan, V. Isler, and K. Daniilidis. Performance Evaluation of Stereo for Tele-presence. In *ICCV*, 2001.

[18] S. Pollard, M. Pilu, S. Hayes, and A. Lorusso. View Synthesis by Trinocular Edge Matching and Transfer. In *BMVC*, 1998.

[19] M. Potmesil. Generating Octree Models of 3D Objects from Their Silhouettes in a Sequence of Images. *CVGIP*, 40:1–29, 1987.

[20] S.M. Seitz and C.R. Dyer. Photorealistic Scene Reconstruction by Voxel Coloring. In *CVPR*, pages 1067–1073, 1997.

[21] P. Srinivasan, P. Liang, and S. Hackwood. Computational Geometric Methods in Volumetric Intersection for 3D Reconstruction. *Pattern Recognition*, 23(8):843–857, 1990.

[22] R. Szeliski. Rapid Octree Construction from Image Sequences. *CVGIP: Image Understanding*, 58(1):23–32, 1992.

[23] R. Szeliski. Stereo Algorithms and Representations for Image-Based Rendering. In *BMVC*, volume 2, pages 314–328, 1999.

[24] C.L. Zitnick and T. Kanade. A Volumetric Iterative Approach to Stereo Matching and Occlusion Detection. Robotics Institute CMU-RI-TR-98-30, CMU, Pittsburgh, PA, 1998.