Machine Vision and Applications (2008) 19:217–222
DOI 10.1007/s00138-007-0094-y

ORIGINAL PAPER

An occlusion metric for selecting robust camera configurations

Xing Chen · James Davis

Received: 5 January 2006 / Revised: 17 February 2007 / Accepted: 9 May 2007 / Published online: 25 July 2007

Abstract Vision based tracking systems for surveillance and motion capture rely on a set of cameras to sense the environment. The exact placement or configuration of these cameras can have a profound affect on the quality of tracking which is achievable. Although several factors contribute, occlusion due to moving objects within the scene itself is often the dominant source of tracking error. This work introduces a configuration quality metric based on the likelihood of dynamic occlusion. Since the exact geometry of occluders can not be known a priori, we use a probabilistic model of occlusion. This model is extensively evaluated experimentally using hundreds of different camera configurations and found to correlate very closely with the actual probability of feature occlusion.

1 Introduction

In designing a vision-based tracking system it is important to define a metric to measure the "quality" of a given camera configuration. Such a quality measure has several applications. By combining it with an optimization process we can automate camera placement in complex environments. Further, dynamically changing arrangements require some metric to guide the automatic choice of best configuration. For example, a multi-target tracking system with many pan-tilt cameras might dynamically focus different subsets of cameras on each target. A metric is need to guide this process.

Some applications require dynamic reconfiguration due to bandwidth or processor power limitations. For example, consider the tracking system shown in Fig. 1 with dozens of cameras. Due to bandwidth constraints only a subset of the cameras can be active. The figure shows three possible subsets. Which subset is best? A quality metric allows us to choose the camera configuration that enables the best tracking performance.

In a motion capture or surveillance system, multiple cameras observe a target moving around in a working volume. Features on the target are identified in each image. Triangulation or disparity can be used to compute each feature's 3D position. In such a system, performance degradation can come from two major sources: *low resolution* which results in poor feature identification; and *occlusion* which results in failure to see the feature. Occlusion is often the more important of the two. When not enough cameras see a feature, it is difficult or impossible to calculate its 3D position.

Occlusion may be due to either dynamic or static objects in the scene. Static occlusion is commonly caused by limited viewing angle, walls, and other known obstacles. Dynamic occlusion is caused by the unknown motion of the target itself.

A quality metric for placing cameras should reflect the impact of all relevant factors. A great deal of prior work exists which addresses image resolution and the impact of static geometry, but no metric yet exists for evaluating the impact of dynamically changing geometry.

This work addresses that need by introducing a quality metric that accounts for dynamic feature occlusion. The metric itself is the primary contribution of this work. In addition,

Authors X. Chen and J. Davis were in Computer Graphics Lab at Stanford University at time of research.

X. Chen
NVIDIA Corporation, Santa Clara, CA, USA
e-mail: xcchen@ieee.org

J. Davis (✉)
University of California - Santa Cruz, Santa Cruz, CA, USA
e-mail: davis@cs.ucsc.edu

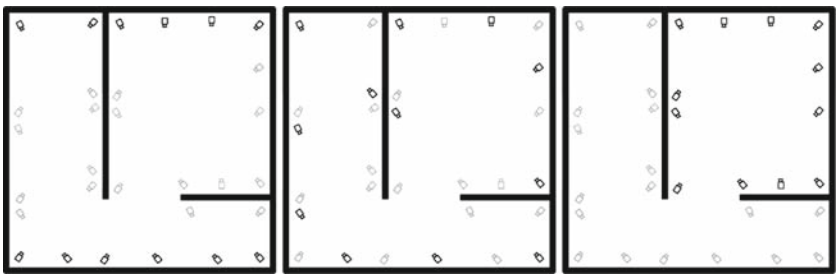

Fig. 1 A tracking system shown with three possible configurations of active cameras. A quality metric is necessary to dynamically reconfigure the system subject to real-time information and constraints

the metric's predictive power is evaluated experimentally on many actual camera configurations, and shown to have high correlation with actual occlusion probabilities.

2 Related work

The effects of static occlusion on camera placement are well studied. For example, the camera placement problem can be regarded as an extension to the well-known art-gallery problem [8,11]. Both problems have the goal of covering a space using a minimum number of cameras and the solutions are greatly affected by the visibility relationship between the sensor and target space. However, the art-gallery problem focuses on finding the theoretical lower-bounds on the number of guards under *known* geometry. Automatic sensor planning has also been investigated in the area of robotic vision [15–18,21], motion planning [4,5,20], image-based modeling [10], laser scanning [12–14], and measuring BRDFs [7]. The target domain in all these cases is a static object, and the task is to find a viewpoint or a minimum number of viewpoints that exposes the features of interest on the target as much as possible. Unfortunately, nearly all work in these areas assumes that scene geometry can be evaluated deterministically. This paper introduces a quality metric that explicitly accounts for the non-deterministic nature of dynamic occlusion.

Image resolution has also been proposed as a metric for placing multiple cameras. Olague et al. [10] approximated the projective transformation of a camera using Taylor expansion, and used a scalar function of the covariance matrix as the uncertainty measure. Wu et al. [19] proposed a computational technique to estimate the 3D uncertainty volume by fitting an ellipsoid to intersection of projected error pyramids These methods consider limited image resolution as the only cause of 3D uncertainty. However, occlusion is frequently present in feature-based motion tracking systems and is often the dominant source of error.

Some researchers have accounted for multiple sources of error. As one example, Cowan and Kovesi consider resolution, focus, static occlusion, and field of view in designing a quality metric [3]. However none of the combined metrics account for the effects of dynamic occlusion.

A complete quality metric for evaluating camera configurations would include many factors, including static occlusion, image resolution, and dynamic occlusion. This work augments previous research by introducing a metric which predicts the impact of dynamic occlusion.

3 Occlusion metric

Occlusion occurs when a target point is not visible from a camera. This occlusion may be caused by either static or dynamic objects in the scene, as shown in Fig. 2. The challenge to computing the error caused by dynamic occlusion is that we do not know exactly where the target or occluder will be at any time. Without knowledge of target and occluder location it is impossible to arrange cameras such that the target is guaranteed to be visible.

We address the dynamic occlusion problem by first assuming perfect knowledge and then making several approximations that allow a probabilistic model to be developed. Assume first that we know exactly the geometric model of a

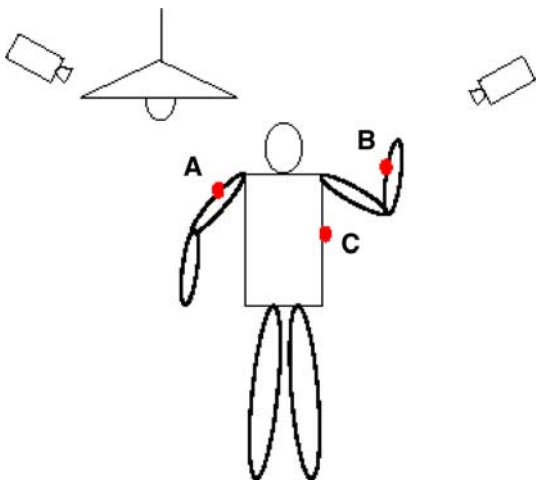

Fig. 2 Both static and dynamic occlusion can affect the probability of observing features. In this example static occlusion by the lamp exists between the *left camera* and point *A*. This sort of occlusion can be evaluated deterministically. Dynamic occlusion exists between the *right camera* and points *B* and *C*. Since this occlusion depends on the time varying pose of the person, this occlusion is evaluated probabilistically

target object and the path that the object takes during a tracking session. We can evaluate the error caused by dynamic occlusion precisely, because we can simulate the exact geometry and count exactly how many feature points are occluded from each camera viewpoint at any time. Of course, this method is not very useful for designing a real tracking system. In reality, the path that a target takes could vary greatly and camera configurations optimized for one specific path may be poor for other paths. We seek to find a camera configuration that best avoids occlusion for all possible paths.

Given a camera configuration, $\mathbb{C}$, we would like to predict whether features are likely to be observed by an adequate number of cameras. Since the number of observing cameras will vary depending on feature location and occluder position, we express the likelihood of observation as a probability function. For real camera configurations, this function can be empirically measured. For example, Fig. 3 shows the measured probability that point features will be observed by N cameras. This data represents the probabilities for one possible camera configuration, during a single motion capture session. An ideal predictor would allow the probability distribution function, $\mathcal{P}_{\text{oc}}(\mathbb{C})$, to be estimated for any given camera configuration without physical experimentation.

A quality metric is a mapping from a camera configuration to a single scalar value. Since most tracking systems rely on observing a feature from more than one viewpoint, we define our metric as the probability that at least two cameras observe a feature. Letting $Q(\mathbb{C})$ represent our quality metric, and maxcam be the total number of cameras, we have:

$$Q(\mathbb{C}) = \sum_{2}^{\text{maxcam}} \mathcal{P}_{\text{oc}}(\mathbb{C}). \tag{1}$$

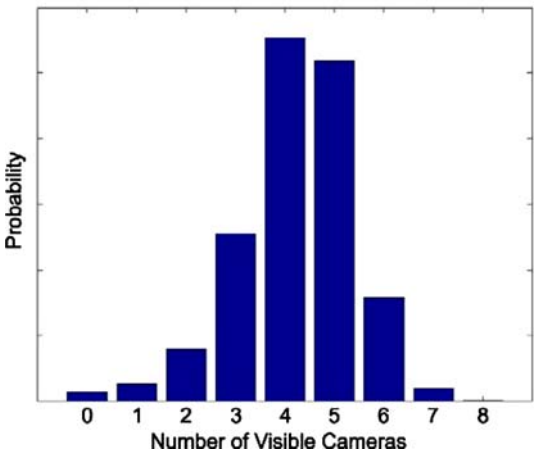

Fig. 3 The likelihood that target features are occluded can be represented as a probability distribution function over the number of cameras from which the feature is visible. This plot shows the probability of observing a feature during an actual motion capture session. Note that it is most likely that about half of the cameras see a feature, but that occasionally only a few or nearly all cameras can observe the feature

Since occlusion characteristics may vary over the target space (the entire working volume of interest), we define a spatially dependant quality metric. We make the assumption that the probability of occlusion at each position is independent. Thus by sampling the target space and aggregating the per-point metric, the mean probability of occlusion can be computed. For example if the target space is a room, then we calculate the probability of occlusion at n sample points in the room, and aggregate these values. Further, it is straightforward to incorporate knowledge of where features are more likely to be located. The sampled points can merely be drawn from a non-uniform feature density function, δ_{f}.

$$\mathcal{P}_{\text{oc}}(\mathbb{C}) \approx \sum_{\mathbf{P}\in\delta_{\text{f}}} \mathcal{P}'_{\text{oc}}(\mathbb{C}, \mathbf{P})/n \tag{2}$$

Where $\mathbf{P}$ represent a particular point in the scene drawn from distribution δ_{f}, and n is the number of samples drawn.

In a complete tracking system, the distribution δ_{f} will be time varying due to precisely the factors that require a quality metric. For example, suppose that we are certain our target is currently in the left half of the room. A good quality metric would prefer camera configurations that more carefully observe this region. In this case, δ_{f} should have zero probability in the right half of the room, so that no samples are used in the irrelevant region. In this work we focus on the quality metric itself, rather than a complete tracking system. Thus in the evaluation presented later, we simply set δ_{f} to be a distribution that uniformly covers space.

The probability of occlusion is clearly dependant on the precise nature of occluders. Although we may not know exactly where the "occluders" will be, we may have some idea as to how likely they are at certain positions and orientations. Given a density function for occluder positions, δ_{o}, we can generate possible occluders by drawing samples from the distribution. Figure 4 shows an example configuration of three cameras and a particular sampled occluder. In this case two cameras can see the target feature. For each possible occluder, we calculate how many cameras can observe the target point, $\mathbf{P}$. The result for each occluder are aggregated

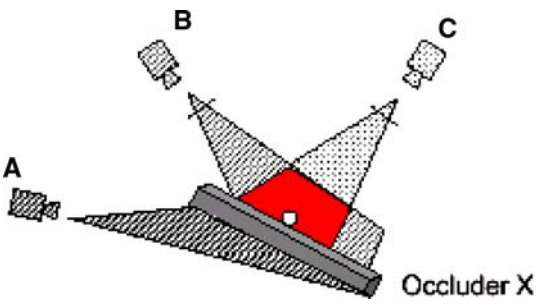

Fig. 4 A randomly sampled occluder will prevent a given feature point from being seen by some cameras. In this case cameras *B* and *C* observe the feature

into a single estimate of occlusion characteristics. Drawing m samples from the space of all possible occluders gives us a distribution describing the expected probability of occlusion at a point.

$$\mathcal{P}_{\text{oc}}(\mathbb{C}, \mathbf{P}) \approx \sum_{\mathbf{O} \in \delta_{\text{o}}} \mathcal{P}''_{\text{oc}}(\mathbb{C}, \mathbf{P}, \mathbf{O})/m \tag{3}$$

Many occluder distribution models are possible. A feature can be occluded either directly by the object it is attached to, or by another object from a distance. In the most general case, the visibility of a point depends on the position, orientation and size of the occluder.

Although it would be possible to sample occluders of all possible sizes, at all possible positions, in all possible orientations, the size of this space would lead to an inefficient implementation. In this work we observe that the worst case occluders can be drawn from a much smaller sample space. Furthermore, the evaluation presented in the next section shows that the simplification of sampling only the worst case occluders still leads to a metric that closely correlates with actual measured data.

The worst case occlusion occurs when the occluder is very near the point, as shown in Fig. 5. In this location a whole hemisphere is occluded. To obtain a conservative estimate of occlusion, we need only simulate this worst case occluder pose. When the occluder is in this position, very near the target point, the size of the occluder does not matter. Therefore, to generate a conservative estimate of occlusion at point **P**, we define the occluder density function, δ_{o}, to include only planar occluders through **P**, allowing variation in orientation, but omitting size and position. As a further simplification the distribution of the planar occluders is defined to be uniform.

Combining Eqs. 1–3, the quality of a given camera configuration can be written as the probability that at least two cameras see a feature point. We evaluate this value by sampling over all possible feature locations and all possible occluder positions.

$$Q(\mathbb{C}) \approx \sum_{2}^{\text{maxcam}} \sum_{\mathbf{P} \in \delta_{\text{f}}} \sum_{\mathbf{O} \in \delta_{\text{o}}} \frac{\mathcal{P}''_{\text{oc}}(\mathbb{C}, \mathbf{P}, \mathbf{O})}{m \cdot n} \tag{4}$$

Static occlusion is caused by static objects in the scene that are known a priori and time-invariant. Due to these occlusions, a point **P** is not visible from certain cameras. These occlusions can be easily included in the above formulation: we simply need to perform a visibility test to see if any static object in the scene is between point **P** and the optical center of each camera. If **P** is not visible, the given camera is marked as occluded, regardless of the orientation of the random planar occluder.

The use of sampling also allows us to easily encode further application-specific knowledge into the quality metric. For example, one might know a priori that certain parts of the target spaces require higher robustness. This can be incorporated by increasing the number of samples drawn from those regions.

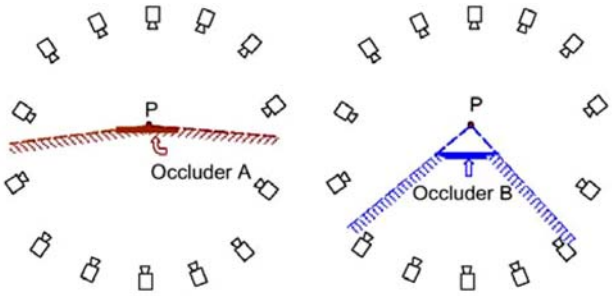

Fig. 5 The size and orientation of an occluder determine the shape of the occluded region. A conservative estimate of occluded region can be made by assuming the occluder is very close to point **P**. In this case an entire hemisphere of camera locations will not be able to see point **P**

4 Evaluation

Although we have formulated of our quality metric based on physical principles, we have made some simplifying assumptions: that the probability of occlusion is independent across spatial locations, that occluders can be modeled conservatively, and that occluder orientations are uniformly distributed. Therefore we would like to verify that the simplified model adequately predicts reality. We extensively evaluated our metric by comparing predicted probabilities of occlusion to actual measured probabilities for 256 different camera configurations. We found that the proposed metric correlates well with experimental data.

Each of the 256 unique camera configurations was constructed by placing between 1 and 8 cameras on a circular ring surrounding the working volume. The ring around the target area measured roughly 4×4 m, with motion observable in a 2×2 m area in the center. Spacing between the cameras varied from nearly uniform to heavily weighted to one side. Camera height varied between approximately 1 and 3 m from the floor. Figure 6 shows an example of three of the configurations. Because of the wide variance in how cameras were placed, it is expected that some of the configurations would result in relatively little occlusion, while some would result in frequent occlusion. Cameras were calibrated using standard methods [1,2,6].

For each configuration, we attached a bright LED on various parts of a human body such as the head, shoulder, knee, etc. In each configuration, five trials were run, each with a different LED position in order to minimize accidental bias based on body attachment location. We asked the person to move around in the working volume for approximately 20 s in each trial, and sampled the cameras at 30 Hz. The motion

Fig. 6 Three camera placement scenarios out of the set of 256 conditions tested

consisted of waving arms, turning, walking in circles, and similar actions meant to have suitable complexity that marker occlusion could be expected. However markers were not explicitly covered, such as by placing a hand over a marker.

By processing the video streams from the various cameras, we counted at each sampled time instant, how many cameras could see the bright LED feature point. Aggregating the number of "visible" cameras over all frames and all trials (with the LED placed on different body parts) yields a measured distribution function for a given camera configuration.

In our implementation data is collected and stored to disk in real time. All evaluation was conducted offline in MATLAB. The metric we suggest is computationally quite simple and we are confident that it could trivially be included in a real-time system.

Figure 3 shows the measured observation probability function for a single camera configuration. Due to space constraints it would be impossible to include numerical data for all 256 experimental conditions. Instead we aggregate all experiments into a single plot which shows the correlation between our metric and measured data.

A good model should predict a distribution that is similar to what is obtained from the experiment. In order to see the "similarity" quantitatively, we use a common regression analysis technique [9]. In order to perform the regression analysis, we use the metric defined in Eq. (4). We compute an occlusion probability for each of the 256 camera configurations both from the experimentally measured data and through our metric.

The experimental and predicted probabilities are plotted against each other in Fig. 7. A perfect model would generate a straight line from (0,0) to (1,1), showing perfect correlation. The prediction from our model is quite good, with a correlation between experiment and prediction of 0.97, quite close to 1.0. Moreover, the correlation between different sessions in the experimental data itself is 0.98, which means that there is variance in the measured data itself. Thus a correlation of 0.98 is the upper bound of how well the experimental data can be predicted. We also computed the mean error in predicted probability across all data points as 0.06 with a standard deviation of 0.05. From the graph we can see that

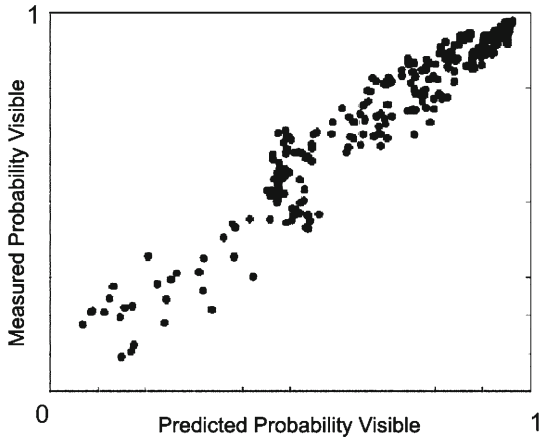

Fig. 7 The predicted probability of observing features is plotted against the measured probability for each of 256 different camera configurations. Each *dot* represents a unique camera configuration. The predicted and measured data have a correlation of 0.97

the predictions improve as we obtain more desirable camera configurations. Given that, we conclude that the simplified occlusion model used in our quality metric predicts the actual occlusion behavior quite well.

The collected data is shown in another form in Fig. 8. In this case, the predicted and experimentally acquired probabilities of occlusion are plotted against the number of cameras in the configuration. Again we see that the range of predicted probabilities are well matched by empirical observation. In addition, this plot graphically illustrates that the more cameras there are in a system, the less often occlusion occurs. For example, when only four cameras are available, using the best configuration tested, feature points have a 90% probability

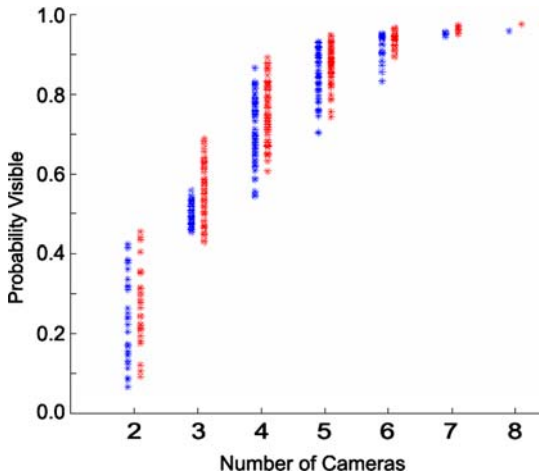

Fig. 8 The number of cameras in a configuration is plotted against the predicted (*left bar*) and experimentally determined (*right bar*) probability of seeing features. The upper envelope of this curve represents the configuration with the best quality. As the number of cameras increases, so does the probability of observing features

of being visible. Using four cameras and a poorly chosen configuration, features would have only a 60% probability of being visible.

The upper envelope of this plot gives some indication to the number of cameras required for a given robustness. If it is desired that feature points be visible 95% of the time, then at least six cameras are required to cover the working volume.

It should be noted that the plot shown in Fig. 8 contains only data from the 256 configurations generated during our experimentation. Both higher and lower quality configurations are possible. For example, our eight camera configuration had cameras even distributed around the working volume. It is certainly possible to create a bad configuration from eight cameras, for example by placing them all right next to each other, even though that possibility is not represented in this plot. Similarly it may be true that alternate configurations would produce higher quality.

5 Discussion

This work has introduced a quality metric for camera configurations based on the probability of dynamic occlusion. The metric was extensively evaluated by empirical comparison with many actual camera configurations, and found to have high correlation to actual probabilities of occlusion.

One surprising aspect of the results presented here is the simplicity of the proposed metric. It assumes a uniform distribution of feature measurement locations. It makes a worst case assumption about the size, shape, and location of occluders; modeling them as always large and very close to the feature in question. The orientation of occluders is modeled as uniform. Despite these simplifications, we have found that the metric models reality surprisingly well. The correlation between prediction and measured performance is very close to the upper bound of possible predictive performance. Therefore, we conclude that our simplifications were in fact justified, a more complex model of occlusion is not necessary.

We are currently investigating the use of this metric to produce a camera configuration selector. In large tracking systems with hundreds of cameras, it is infeasible to have all cameras active at all times. The metric allows us to choose the optimum subset of cameras which meet both our resource and robustness requirements.

References

1. Azarbayejani, A., Pentland, A.: Real-time self-calibrating stereo person tracking using 3-D shape estimation from blob features. In: Proceedings of the 13th International Conference on Pattern Recognition, vol. 3, pp. 627–32 (1996)
2. Chen, X., Davis, J.: Wide area camera calibration using virtual calibration objects. IEEE Comput. Vis. Pattern Recogn. (2000)
3. Cowan, C.K., Kovesi, P.D.: Automatic sensor placement from vision task requirements. IEEE Trans. Pattern Anal. Mach. Intell. **10**, 407–416 (1988)
4. Fleishman, S., Cohen-Or, D., Lischinski, D.: Automatic camera placement for image-based modeling. In: Proceedings of Seventh Pacific Conference on Computer Graphics and Applications, Los Alamitos, CA, USA, IEEE Comput. Soc. (1999)
5. Gonzalez-Banos, H., Latombe, J.-C.: Navigation strategies for exploring indoor environments. Int. J. Robot. Res. **21**, 829–848 (2002)
6. Heikkila, J., Silven, O.: A four-step camera calibration procedure with implicit image correction. IEEE Comput. Vis. Pattern Recogn. 1106–1112 (1997)
7. Lensch, H., Lang, J., Sa, A., Seidel, H.-P.: Planned sampling of spatially varying brdfs. Comput. Graph. Forum (Eurographics) 22 (2003)
8. Marengoni, M., Draper, B., Hanson, A., Sitaraman, R.: Placing observers to cover a polyhedral terrain in polynomial time. In: IEEE Workshop on Applications of Computer Vision (WACV), Sarasoto, FL (1996)
9. Neter, J., Wasserman, W., Kutner, M.H.: Applied Linear Statistical Models, 4th edn. Irwin (1996)
10. Olague, G., Mohr, R., Venkatesh, S., Lovell, B.C.: Optimal camera placement to obtain accurate 3D point positions. In: Proceedings of Fourteenth International Conference on Pattern Recognition, Los Alamitos, CA, USA (1998)
11. O'rourke, J.: Art Gallery Theorems and Algorithms. Oxford University Press, Oxford (1987)
12. Pito, R.: A solution to the next best view problem for automated surface acquisition. IEEE Trans. Pattern Anal. Mach. Intell. **21**, 1016–1030 (1999)
13. Reed, M., Allen, P.: Constraint-based sensor planning for scene modeling. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1460–1467 (2000)
14. Scott, W., Roth, G., Rivest, J.-F.: View planning with a registration constraint. In: Third International Conference on Recent Advances in 3D Imaging and Modeling (3DIM) (2001)
15. Sedas-Gersey, S.: Algorithms for automatic sensor placement to acquire complete and accurate information. Ph.D. Dissertation (CMU-RI-TR-93-31), The Robotics Institute and Department of Achitecture, Carnegie Mellon University, Pittsburgh, PA (1993)
16. Tarabanis, K.A., Allen, P.K., Tsai, R.Y.: A survey of sensor planning in computer vision. IEEE Trans. Robot. Automat. **11**, 86–104 (1995)
17. Tarabanis, K.A., Tsai, R.Y., Allen, P.K.: The Mvp sensor planning system for robotic vision tasks. IEEE Trans. Robot. Automat. **11**, 72–85 (1995)
18. Triggs, B., Laugier, C.: Automatic camera placement for robot vision tasks. In: Proceedings of 1995 IEEE International Conference on Robotics and Automation. Ieee, New York, NY, USA (1995)
19. Wu, J.J., Sharma, R., Huang, T.S.: Analysis of uncertainty bounds due to quantization for three-dimensional position estimation using multiple cameras. Optical Eng. **37**, 280–92 (1998)
20. Ye, Y., Tsotsos, J.: Sensor planning for 3d object search. Comput. Vis. Image Understand. **73**, 145–168 (1999)
21. Yi, S., Haralick, R.M., Shapiro, L.G.: Automatic sensor and light source positioning for machine vision. In: Proceedings of 10th International Conference on Pattern Recognition, IEEE Comput. Soc. Press Los Alamitos, CA, USA (1990)