MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28-30, 2000, The University of Tokyo, Japan

13—22

Combating Occlusion and Scene Changes for Camera Position Determination

Philip Quick and David Capson
Department of Electrical and Computer Engineering
McMaster University
Hamilton, Ontario, Canada
E-mail: quick@mcmaster.ca,capson@ece.eng.mcmaster.ca

Abstract

The determination of camera position via eigenspace methods has been applied recently to robot navigation and visual servoing. Conventional eigenspace methods are not robust to occlusion and scene changes due to their global nature. This paper describes techniques for performing robust camera position determination. Input images are divided into separate local areas, with camera position determined through eigenspace analysis for each area individually. The positions of occluded sections are ignored in determining the final camera position. The detection of occluded sections is determined by applying a threshold to the eigenspace reconstruction error of the input images. Experiments were performed determining the translational position of a camera and analyzing the magnitude of the errors caused by occlusion, the ability to detect different size occlusions and the accuracy improvement provided by the subsectioning technique.

1 Instructions

Eigenspace methods have recently become a popular technique in computer vision, notably in the areas of face and object recognition [1][2][3]. Eigenspace methods have also been used recently for the problem of determining the position of a camera relative to a scene [4][5]. This technique can be used for a variety of applications including robotic navigation and visual control applications.

A drawback of eigenspace methods is their global nature. Occlusions and small scene changes can have adverse effects on the ability to determine the camera's position accurately. Larger occlusions or those that differ greatly from the original scene contents have proportionally greater effects on accuracy.

To avoid the effects of occlusion for determining camera position, the input image can be divided into sections with separate eigenspaces. By detecting those sections containing occlusions, the remaining nonoccluded sections can be used to robustly determine position. A similar technique has been used in face recognition, whereby separate eigenspaces were constructed for the eyes, nose and mouth regions [6].

Occlusions can be detected by measuring the eigenspace reconstruction error of an image section. Occluded sections will no longer lie primairly within the eigenspace, resulting in a larger reconstruction error and thus detectable via a threshold. The ability to differentiate between occluded and nonoccluded scenes depends upon the size and appearance of the occlusion relative to the background. Smaller image sections are more apt to differentiate between occlusions because of the larger relative size of the occlusion, hence a larger deviation from the eigenspace.

A different approach to solving occlusion for camera position determination is described in [5] and [7]. Our local based approach gives the benefit of also determining in which region the occlusions occur. This idea is similar to the eigenspace time inspection technique described in [4].

This paper shows experimental results demonstrating the effects on accuracy of occlusion on positional accuracy, the ability to detect occlusion via eigenspace reconstruction as well as the effectiveness of the image subsectioning technique for dealing with occlusion. The experiments were performed with a camera mounted on an XY table to simulate the movement of a mobile robot.

2 Method

2.1 Basic Method

Defining a motion range for a camera, whether it is translational, rotational or a combination of both defines a visual subspace (eigenspace) consisting of the images from the camera within this range. Images from nearby camera positions will in general be highly correlated, consequently such visual subspaces are highly compressible via the Karhunen-Loeve Transform (KLT).

The KLT produces a small set of vectors that captures most of the variance of the eigenspace, allowing a low dimensional representation of images from the visual subspace. If x[1] represents a concatenated image vector drawn from the visual subspace and E represents a matrix consisting of the first n KLT vectors of the eigenspace, the low dimensional representation y of x (projection vector) is formed by:

$$y = Ex \tag{1}$$

$$E^T = [e_1 \dots e_n] \tag{2}$$

By storing the projection vectors for a set of images equally spaced throughout the camera's movement range, the current position of the camera can be ascertained by performing a nearest neighbor search of this set and the current image's projection vector, a technique first proposed by Nayar [2].

This approach limits the accuracy to the distance between the images in the stored set. The projections can be interpolated to create points for intermediate positions between the actual images.

The problem with occlusion is the fact that occlusions will alter the value of the projections of the occluded image resulting in positional errors when performing the nearest neighbor match. The extent to which an occlusion affects positional accuracy depends on the size and the difference in the appearance of occlusion and the background. Increasing the size of the occlusion or the difference from the original background results in larger changes to the resultant projection vector and consequently larger errors with the nearest neighbor matching. The location within an image can also alter the impact of an occlusion on accuracy.

2.2 Occlusion Correction

An input image can easily be divided into subimages and separate eigenspace calculations performed with each of them. In this manner, if one subimage contains an occlusion, that positional information can be disregarded and the positional information of the other subimages used. For our experiments we averaged the remaining subimage positions to determine the overall position.

Potentially the image could be divided into many small subimages, so with even numerous occlusions there would still be unoccluded subimages. The problem with this scheme is that small images have the problem of increasing ambiguity, where images from nearby positions have very similar appearances resulting in large errors. For our experiments, the subimages were large enough for this not to be a problem. An alternative to subimages would be the use of additional cameras aligned in different directions, and their whole images taking upon the subimage role. These separate images could be subdivided as well.

Small scene changes can also be dealt with in this manner, such as those encountered by a mobile robot in its workspace over time.

2.3 Occlusion Detection

To determine which sections of the image are occluded and thus should be disregarded can be accomplished using the reconstruction of the new image x from the KLT vectors:

$$x' = \sum_{i=1}^{n} (x^T e_i) e_i \tag{3}$$

Occluded images will no longer lie within the original eigenspace basis, thus the Euclidean error between the original image and the reconstructed image will in general be larger than that of nonoccluded images. Pentland [1] used this concept to determine whether an image contained a face or not. An alternative measure to detect occlusion or scene change is to measure the distance between the image's projection and that of its nearest neighbor match[4].

$$||x - x'|| \tag{4}$$

This reconstruction error will increase with occlusion size and the deviance of occlusion's appearance from the original background. Using smaller image sections will increase the reconstruction error since the occlusion will be a proportionally larger part of the image.

For simplicity we assumed a Gaussian distribution based on the reconstruction error statistics of a set of random images and experimented with several threshold levels, whereby an image with a reconstruction error surpassing the threshold was determined to be occluded.

Our experiments determining the translational position of the camera showed that occlusions large enough to impair accuracy were easily differentiable from nonoccluded images.

3 Experimental Results

To illustrate the effects of occlusion on positional accuracy, the detection of occlusions and robust accuracy despite occlusion, a base experiment was set up using an XY table. The camera's movement range was a 40 cm by 40 cm square within our vision

[1]We assume the image x has already had the mean image of the subspace subtracted

Occlusion Size (Pixels)	Occlusion Size (Percentage)	Location A Error (mm)	Location B Error (mm)
0	0%	1.62	1.62
10 by 10	0.13%	1.82	1.62
20 by 20	0.52%	2.52	1.68
40 by 40	2.08%	3.90	1.77
60 by 60	4.74%	5.70	1.88

Table 1: Average Error Vs. Occlusion Size

laboratory. Figure 1 shows images from the corners of the movement range. The stored projection vectors consisted of images separated by 2.5 cm in each direction interpolated to provide projections 1 mm apart. The images were 320 by 240 pixels.

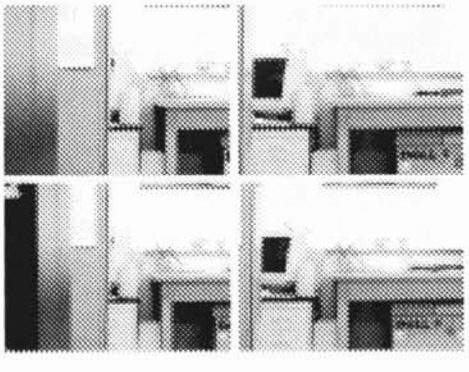

Figure 1: Corner Images of Camera Range

A set of 300 test images was acquired spaced randomly throughout the camera's range. For testing purposes, the average lateral position error of this set was used in the rest of the experiments for assessing the effects of occlusion on accuracy. Forward accuracy is often considerably poorer and can be avoided by using the lateral information of two offset cameras.

3.1 Effect of Occlusion on Accuracy

To investigate the relationship between occlusion size and accuracy, the test image set was embedded with different size squares with zero intensity as a quantifiable artificial occlusion. Zero intensity was chosen so the occlusion would have a relatively large difference from the original background and thus have a comparatively large effect for its size. Two different locations were used to show how some areas of the image respond differently to a particular occlusion. Table 1 shows the average lateral error for the nonoccluded case and errors for both locations with different size occlusions.

Interestingly, location A (square centered row 30, column 30 of the image) is much more sensitive to the occlusion than location B (square centered row 210, column 30) with significant error as the occlusion becomes an appreciable portion of the image. Location B shows little increase even with the largest occlusion.

3.2 Occlusion Detection

To illustrate the feasibility of detecting occlusions using reconstruction error thresholds, the artificially occluded test images from the previous section were utilized. To show the relationship between occlusion size and reconstruction error, the reconstruction error was recorded for the nonoccluded set of the entire images, as well as the 5, 10 and 20 pixel occlusion sets, with the square in location A. For visualization purposes the mean and variance of each set was calculated and Figure 2 shows a plot of each occlusion's statistical information with a Gaussian distribution.

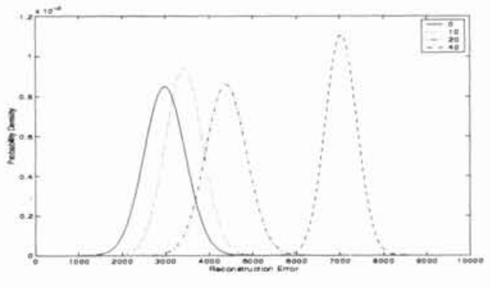

Figure 2: Gaussian Distributions, Full Image

In this case there is clearly no overlap of reconstruction error between the nonoccluded images and the 40 pixel square, but significant overlap with the smaller occlusion squares.

Smaller image subdivisions ensure the occlusion will be larger proportionally and easier to detect. Figure 3 shows the same reconstruction error data, except the images consists of the top left quarter of the entire image, with separate eigenspace analysis performed for that subsection. As expected, using a smaller image subsection will make detecting the same occlusion easier, as the overlap of the probability between occluded and nonoccluded sets is significantly less. Thus for this example, detecting the 20 pixel occlusion can be accomplished with little chance of rejecting nonoccluded images.

Certainly an occlusion could lie on the boundary of two image subsections and thus be more difficult to detect. A possible solution would be to allow overlapping subsections, increasing the chances of detecting such an occlusion.

3.3 Threshold Selection

To experiment with setting a threshold, the same testing images were used as the previous settings.

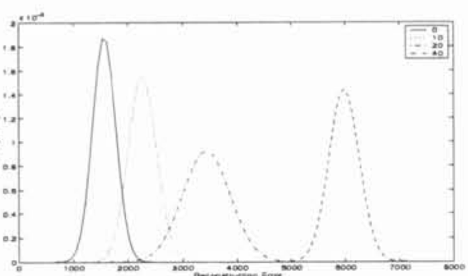

Figure 3: Gaussian Distributions, Quarter Image

Gaussian Threshold	0 Pixels	10 Pixels	20 Pixels	40 Pixels
10%	18%	51.3%	98%	100%
5%	12.7%	42%	97%	100%
1%	5.3%	28 %	91.7%	100 %

Table 2: Gaussian Threshold Vs. Occlusion Size

Assuming a Gaussian distribution for the nonoccluded reconstruction error, several thresholds for reconstruction error were set by determining the point of the top 1%, 5% and 10% of the nonoccluded reconstruction error. Using these thresholds, the occluded and nonoccluded test sets were classified based on these thresholds. Table 2 shows the percentage of images rejected as being above the respective occlusion thresholds. The images were the entire images.

Using the entire image to detect occlusion in this instance, the 40 by 40 square is easily detectable. The smallest occlusion is virtually impossible to detect, although it should be noted from the previous sections, it has very little effect on accuracy.

Table 3 shows the same experiment except using the upper left quarter image subsection. There is marked improvement, showing the increased accuracy with a smaller image size.

3.4 Final Experiment

As a test of the overall technique, the original test set was modified to include the occlusions at both locations A and B. These test images were separated into four subsections. Separate eigenspace analysis was performed for each of the four subsections. For each test image, the position was determined separately for each quarter image. A Gaussian based threshold of 1% was applied to each subsection image. The final position was determined as the average of those subsections below the thresholds. Table 4 shows the average error using this technique versus using the entire occluded image. The error is similar to that of the unoccluded full size image.

Gaussian Threshold	0 Pixels	10 Pixels	20 Pixels	40 Pixels
10%	16.7%	99%	100%	100%
5%	13.3%	97%	100%	100%
1%	3.67%	90 %	100%	100 %

Table 3: Gaussian Threshold Vs. Occlusion Size

Occlusion Size	Entire Image	Subsectioning Technique
20 by 20	2.89	1.72

Table 4: Error: Full Image Vs. Subsectioning

4 Conclusions

The subsectioning technique was shown to be an effective method of dealing with occlusions large enough to cause significant error. Ongoing work includes developing a more sophisticated method of detecting occlusion as well as devising a technique that determines position from the combined nonoccluded sections, rather than determining position from each separately.

References

[1] M. Turk and A. Pentland. Eigenspaces for Recognition. Journal of Cognitve Neuroscience, 3(1), pp. 71-86, 1991.

[2] H. Murase and S. Nayar. Visual Learning and Recognition of 3D Objects from Appearance. Int. Journal of Computer Vision, 14(1), pp. 5-24, 1995.

[3] B. Moghaddam and A. Pentland. Probabilistic Visual Learning for Object Representation. IEEE Trans. PAMI 19(7), pp. 696-710, 1997.

[4] S. Nayar, S. Nene, and H. Murase. Subspace Methods for Robot Vision. IEEE Trans. Robotics and Automation 12(5), pp. 750-758, 1996.

[5] M. Jogan and A. Leonardis. Robust Localization of Spinning Images. IEEE Workshop on Visual Surveillance, 2000.

[6] B. Moghaddam and A. Pentland. Face Recognition Using View-Based and Modular Eigenspaces. Automatic Systems for the Identification and Inspection of Humans, SPIE vol. 2277, 1994.

[7] A. Leonardis and H. Bischof. Dealing with Occlusions in the Eigenspace Approach. Proceedings of CVPR 96, pp. 453-458, 1996.