An Adaptive Algorithm Switching System for Image Based Object Registration

A. Clark, R. Green

Department of Computer Science, University of Canterbury, Christchurch.

adrian.clark@hitlabnz.org
richard.green@hitlabnz.org

Abstract

Most modern day image based object registration systems typically rely on one or two algorithms to perform registration across the entire spectrum of operating conditions. While such systems may perform well in one constrained environment, such as an indoor laboratory, they fail when used in varying conditions, such as outdoors. This paper proposes an adaptive system for image based object registration in unconstrained environments. There are two main components in this system. The first is a set of image metrics such as light levels and frequency as found in a fourier transform from a scene. The second is adaptive relationships between the metrics and a set of algorithms used for object registration, such as hue thresholding, Scale Invariant Feature Transforms and Edge Detection and Segmentation algorithms.
Keywords: object registration, expert systems, hybrid algorithms

1 Introduction

Over the past twenty years, a number of different algorithms have been developed to attempt to detect an object in an image[1], and find it's location and rotation with respect from the camera. Each of these algorithms has it's own process of operation, and, as such, it's own strengths and weaknesses. While there has been some work into combining two or three algorithms together in a system([2], [3], [4]), the results often operate as a single algorithm based system with another algorithm for back up when the first algorithm's performance drops below a certain threshold. While this often works fine in controlled systems, the idea often isn't scalable to uncontrolled circumstances - no matter how fine tuned an algorithm is, it is unlikely to ever be able to perform perfect registration under all circumstances.

There are a wide variety of algorithms and procedures used to extract metrics about an image, including a large body of research which is useful towards predicting the effectiveness of algorithms. One such example is finding the fourier[5] transform of the image to determine the frequency content. Another algorithm, the Scale Invariant Feature Transform or SIFT algorithm[6], needs a feature rich image and performs poorly in an image with only very low frequency content.

The proposed adaptive system maps the image metrics to algorithms features for optimised recognition and registration. The motivation behind using an adaptive system in the object recognition field is because there are a large number of algorithms for image filtering and recognition, each with their own strengths and weaknesses. In addition to having a system change between algorithms, each algorithm can have automatic thresholding to ensure that, if it does get chosen, it will be automatically optimised based on the parameters of the scene.

The work described in this paper is the preliminary results of research into designing an adaptive system, based on how expert systems are used in the artificial intelligence domain, for performing image based registration across a wide range of environmental situations. While the work here only examines two separate algorithms and three metrics, the final system will have upwards of twelve algorithms and twenty metrics.

2 Experimental Description

As mentioned, there are two main factors to be correlated, the metrics derived from environmental factors, and how each algorithm is affected by these metrics. The system is designed to relate these two pieces of information together to make an informed decision. These two factors are described in more detail below.

2.1 Finding Metrics

Any piece of information which we can extract from a scene can be treated as a metric. However, optimal metrics for algorithm selection need some restrictions. The first is that the metrics must contain some significant data relevant to the object to be recognised. For example, a cyan pixel count in an image may not be useful unless the algorithm is specifically identifying cyan objects, which is likely to be too specific an algorithm to implement in a generic system.

The second restriction is the computational load. In a lot of computer vision systems, particularly those used to create augmented reality[7], it's often ideal for the system to be able to perform in real time or as close to it as possible. An adaptive system needs to take into account the computational efficiency of metric and algorithm calculations for real-time recognition and registration. For example, there is a need to calculate the metrics quickly enough so that they can be run every few frames to check that the algorithms are still relevant to the scene content.

Metrics are extracted from the following features in this initial experiment. The best metric or metrics to measure each feature will be investigated in further research.

2.1.1 Scene Brightness/Contrast

Obviously light is essential for computer vision systems to work, without light there's no image to track[8]. The key here is adaptability to illumination variance. Some algorithms are more susceptible to low light levels or high contrast areas such as those which would be typically found in an outdoors environment. Brightness and contrast can be determined by evaluating colour values of pixels and noise in the image.

2.1.2 Frequency Components of an Image

This metric is specific to algorithms which use edge and point information, such as the Canny Edge detector[9] and the SIFT[10] algorithm. These algorithms need high frequency components in the form of robust edge features within the image which can be reliably tracked. The frequency content of an image is calculated using a fourier transform. From this we can find the number of high frequency areas in the image. However, this can cause problems as noise is often in the high frequency range, and as such, filtering may be required to remove noise.

2.1.3 Scene Colour Components

This metric applies to colour based algorithms - images which convert an image into binary or grayscale will not be affected by the outcome of this metric. For example, a colour histogram forms the basis of a metric relevant to the optimal deployment of hue thresholding algorithms.

2.2 Choosing Algorithms

The adaptive system is key to the image registration section of the computer vision architecture. This rule based language must be able to perform in real time, as in the CLIPS[1] Expert system language.

It is expected that a large number of algorithms will be able to be categorized based on rules found by a study in to the effects of metrics which is to be conducted at a later date. However, it is probable that there may be cases where no known rules exist. In such cases, a nearest neighbour approach can be taken by the system to find the likely best algorithm.

Because the system runs in the image registration section of the Computer Vision pipeline, any useful algorithm in this section can be added, as long as a relationship is clearly defined to the set of image metrics. The following registration algorithms are included in the initial set.

2.2.1 Hue Thresholding

A computationally efficient algorithm for registration of objects of a specific colour or colour manifold, as in human skin, which is contrasts sufficiently with the background. Early computer vision systems used this due to its efficiency [11]. This algorithm can be most useful in controlled environments, such as object segmentation with a fixed background colour, where for example, people or objects are moving in front of a green screen.

2.2.2 Edge Detection

The Canny Edge Operator efficiently finds edges in an image. A good example of this is the AR-Toolkit markers[12],which have high contrast edges of black on white. Object segmentation is performed by finding all the edges which contain the object.This can be done by taking subsets of the entire edge set based on their proximity to one another. These subsets can then be compared with known edges in the object the system is attempting to locate.

2.3 Method

Relationships between metrics and algorithms were determined by analysing a library of videos captured with a digital camera. The two different

[1] http://www.ghg.net/clips/CLIPS.html

objects used perform registration were a red square and an ARToolkit marker. These two objects were chosen as they represent the main strengths of the two algorithms chosen for this test. Each object had it's photo taken in three different locations, on a cluttered desk, against a white interior wall, and on a gray cement wall outdoors on a sunny day.

Of these three locations, each object had five image samples taken, one straight on and 30 centimetres away, one at 30 degrees from the centre and 30 centimetres away, these two angles again at one metre away, and one centred photo from 2 metres away. All of the photos were taken with a ADS USB 2.0 Webcam, at 640x480 resolution which is typical of the image quality in recent Augmented Reality and real-time computer vision research. The position of the objects in each of the five separate camera locations was calculated manually and recorded, as the distances between the cameras and objects were the same across the different objects and locations.

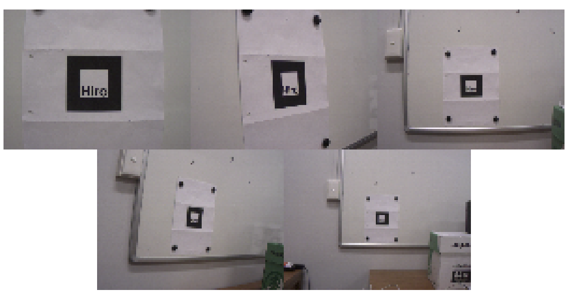

Figure 1: The ARToolkit marker used for registration for the Edge Detection Algorithm, in the five camera/marker positions, inside against a white interior wall

Of these thirty images of two objects in three locations, each metric was applied in five different values. There was the maximum, minimum, average, and the upper and lower quartiles. For example, for the brightness metric, each image was adjusted for the mean brightness set to a maximum brightness level (average=75%), a minimum brightness level (average=25%), an actual brightness level (average=50%), and half way between the actual and maximum/minimum respectively.. To alter the frequency components of the image a Gaussian/sharpen filter was applied with varying degrees of intensity.

This finally ended up with a library of 450 images of the two objects, meaning that each object had a total of 225 images. Each algorithm was run against the appropriate object for registration, the hue thresholding for the red square, edge detector for the ARToolkit marker. The locations calculated by each algorithms was compared to manually calculated location, and a error value was calculated.

3 Results

All the algorithms tested performed better when the metrics were at the average or ideal values, and became progressively worse the further from the average they became. All algorithms performed best when locating the object on a plain white interior wall, followed by the desk then the outdoors environment.

While these results were all expected as they are well known problems with computer vision techniques, the relevant and interesting results came when comparing the errors of each algorithm in the same environment and metric conditions against each other. To estimate the effect each metric had on the algorithms, the errors obtained were averaged across environment, camera location and metric intensity level. The results are summarized in Table 1:

	Hue	Edge
Brightness/Contrast	10.39%	1.86%
Hue Components	11.41%	1.63%
Frequency Components	9.80%	2.56%

Table 1: Mean Error values of Algorithms

Table 1 shows the average percentage error for the two algorithms (Hue Thresholding and Edge Detection) across the range of each metric. For example, when using the Hue Thresholding approach, the average error in registration of an object across a range of Brightness levels (ranging from very dark to very light) was 10.39%, meaning that around 90% of the contents of the area the registration calculated contained the object - either the object was partially outside the boundary or the boundary was too large or too small.

The results contained in Table 1 shows that the Edge detection algorithm proved more effective than the Hue algorithm across all results, by a considerable amount. However we can see that there is a degradation in the accuracy of Edge detection when the Frequency content of the image changes - such as the image becomes blurrier. Compared to the effects of the other metrics, this was considerably less important in the Hue Algorithm. As expected, each algorithm has it's own strengths and weaknesses based on it's design, and the weaknesses can often be more prominent in adverse environmental conditions.

These results indicate the relationships between algorithms and metrics under certain environmental conditions. An expert system implements these relationships to determine which of these factors is most prominent and choose the algorithms which has the lowest mean error when exposed to this factor.

While these are the global results across the whole range, some interesting things were noted on the performance of each algorithm in particular, and these are mentioned below.

3.1 Observations for Edge Detection Algorithm

The accuracy of the Canny Edge Detection algorithm was surprisingly high in all cases. In some cases, it was too sensitive, providing a lot of edges which weren't really that visible. For the baseline of the metrics (average image brightness, hue and detail), at the close range images (30 centimetres) often there would be a lot of objects detected which weren't the marker itself. This can often cause problems with false positives. This is the reason for the slightly higher score on the Frequency components than on the other two metrics, which scored very low.

3.2 Observations for Hue Thresholding

In the test cases the Hue Thresholding algorithm used was not very effective, and fairly susceptible to changes in both lighting and hue. However, it is important to recognise that successful registration depends both on the environment registration is taking place in, and the ability to differentiate between the known objects which are to be found. While this experiment did not contain a large object set, it is conceivable that an image set could contain two objects of similar or identical geometry with different colours. An example would be red and green square markers - in this case a gray-scale edge detector would be unable to tell the difference between between the two markers resulting in registration errors, while a hue-based method would not suffer from the same problem.

4 Conclusion

The research described in this paper establishes the foundation for creating an expert system for Image Based Object Registration. By testing the effects of a range of carefully chosen metrics against a range of equally carefully chosen algorithms, some relationships between environmental factors and algorithms were successfully established.

The results indicate that the proposed adaptive model for mapping the image metrics to algorithm features may be useful for optimised object registration in unconstrained environments.

5 Future Research

The next step in this research is to perform a more in depth experiment, testing a much larger range of algorithms, metrics, and objects to register. This will help establish some initial relationships between metrics and the algorithms.

Through further research in this area, more metrics and algorithms will be added and run through similar tests. Complex relationships between multiple metrics and algorithms will continue to be added to the knowledge database.

The expert system using this knowledge base, will be extended to optimise algorithm choice(s). There will also be further investigation into enhancing each algorithm for a better fit into this adaptive system such as automatic thresholding[12], as well as optimising image cropping areas of interest.

References

[1] J. Park, B. Jiang, and U. Neumann, "Vision-based pose computation: robust and accurate augmented reality tracking," in *Augmented Reality, 1999. (IWAR '99) Proceedings. 2nd IEEE and ACM International Workshop on*, pp. 3–12, 1999. TY - CONF.

[2] L. Vacchetti, V. Lepetit, and P. Fua, "Combining edge and texture information for real-time accurate 3d camera tracking," in *Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on*, pp. 48–57, 2004. TY - CONF.

[3] L. Masso, M. Dhome, and F. Jurie, "Contour/texture approach for visual tracking," in *Scandinavian Conference on Image Analysis*, pp. 661–668, 2003.

[4] J. Parker and B. Behm, "Use of multiple algorithms in image content searches," in *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on*, vol. 2, pp. 246–250 Vol.2, 2004. TY - CONF.

[5] N. Beaudoin and S. Beauchemin, "An accurate discrete fourier transform for image processing," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 3, pp. 935–939 vol.3, 2002. TY - CONF.

[6] I. Skrypnyk and D. Lowe, "Scene modelling, recognition and tracking with invariant image features," in *Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on*, pp. 110–119, 2004. TY - CONF.

[7] H. Kato, K. Tachibana, M. Billinghurst, and M. Grafe, "A registration method based on texture tracking using artoolkit," in *Augmented Reality Toolkit Workshop, 2003. IEEE International*, pp. 77–85, 2003. TY - CONF.

[8] M. Bajura and U. Neumann, "Dynamic registration correction in video-based augmented reality systems," *IEEE Comput. Graph. Appl.*, vol. 15, no. 5, pp. 52–60, 1995.

[9] M. Ali and D. Clausi, "Using the canny edge detector for feature extraction and enhancement of remote sensing images," in *Geoscience and Remote Sensing Symposium, 2001. IGARSS '01. IEEE 2001 International*, vol. 5, pp. 2298–2300 vol.5, 2001. TY - CONF.

[10] D. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, pp. 1150–1157 vol.2, 1999. TY - CONF.

[11] D. Koller, G. Klinker, E. Rose, D. Breen, R. Whitaker, and M. Tuceryan, "Real-time vision-based camera tracking for augmented reality applications," in *Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 87–94, Lausanne, Switzerland: ACM Press, 1997.

[12] C. B. Owen, F. Xiao, and P. Middlin, "What is the best fiducial?," in *The First IEEE International Augmented Reality Toolkit Workshop*, (Darmstadt, Germany), pp. 98–105, Sept. 2002.