

# Registration, Mapping and Context: Computer Vision for Record and Reuse

**Tony Pridmore**

Image Processing and Interpretation Research  
School of Computer Science & IT  
University of Nottingham  
tpp@cs.nott.ac.uk

**Steven Mills**

Image Processing and Interpretation Research  
School of Computer Science & IT  
University of Nottingham  
smx@cs.nott.ac.uk

## ABSTRACT

Cameras figure highly in several Equator experiences, making the incorporation of computer vision both a natural extension to current practice and a practical possibility. In this paper we consider three key areas of record and reuse: registering data gathered in the real world with a corresponding virtual environment, mapping information from one to the other, and, perhaps the area of greatest potential for computer vision, adding contextual information to sensed data. In each case the potential offered by computer vision is outlined and some proposals are made.

## Keywords

Computer vision, image interpretation, context, registration, mapping

## INTRODUCTION

Record and reuse requires that people's activities are monitored in the environment of interest, stored in some form that can be easily searched and reviewed, and then re-presented in a novel context or format. Several technologies have already been used in Equator projects to record people's activities. These have all been essentially point identification tools, such as GPS or RFID tags. While these technologies are effective and reliable they do not provide an overall context to the measurements made, and require an individual sensor for each piece of information to be recorded.

A camera, whether fixed or mobile, can record a view of a large area of space, but this information requires visual interpretation, which makes the information hard to search and review. In order to successfully use video information a variety of computer vision techniques might be applied. In this paper we consider some of the ways that computer vision might help in record and reuse. These are the registration of cameras and models, exploitation of the mapping between images and models that registration supplies, and the extraction of contextual cues from images and image sequences.

## REGISTERING CAMERAS AND MODELS

When implementing a computer vision system the first step is often to calibrate the cameras that are to be used. This can be done in either of two ways.

In off-line calibration [1], a specially constructed target object is used. Planar grids, created by printing black squares onto a white background, are commonly employed, though three-dimensional objects can give improved performance in some situations. Calibration objects need not be engineered to a high degree of accuracy, though the relative positions of features on the target should be measured as accurately as possible. Figure 1 shows the "calibration hat" used at Nottingham. This is constructed from laser-printed grids mounted onto a rigid cardboard box and has proved sufficiently accurate to calibrate a stereo pair of cameras used to recover the 3D shape of a person's head.



**Figure 1.** A surprisingly practical calibration target.

Given an image of the calibration target, key features (usually the corners of squares or the centers of circles) are extracted and matched to a model of the target object. Any of a number of numerical methods may then be applied to determine the camera parameters and geometry needed to generate the perceived pattern of features from the known model. Two groups of camera parameters are recovered; intrinsic and extrinsic. The intrinsic parameters describe the internal composition of the camera(s), such as focal length.

Extrinsic parameters describe the position and orientation of the camera(s) with respect to a three-dimensional, real-world coordinate system defined by the calibration target.

In off-line, or self-calibrating, systems [2] similar data is obtained without the use of a dedicated target. Here a small number of the most visually obvious features are extracted from and matched between stereo pairs of images of the real world acquired from a moving camera. These small numbers of matched feature pairs are used to hypothesise intrinsic and/or extrinsic parameters. Tests are then performed to determine whether or not these parameters account for other image measurements. If they do the parameters are accepted, if not an alternative feature selection is made. The process repeats until an acceptable calibration is identified.

In a record and reuse scenario participants carry cameras through a real environment for which a corresponding virtual environment (or model) exists. Techniques developed by the vision community to calibrate cameras might therefore be used to register images captured by those cameras with the model. This would allow recovery of the position and orientation of each camera with respect to the model as well as the mapping of image data onto the model and vice versa.

A large number of camera calibration techniques have been reported, given the present literature we would make a number of suggestions/observations:

- Intrinsic camera parameters should, if possible, be computed off-line, using a purpose-built target. Self-calibration mechanisms are good at recovering extrinsic parameters, but not so accurate on intrinsics. This would mean that the cameras used in experiences should ideally be fixed focus.
- Calibrating fixed cameras is now a standard task in computer vision. To place a camera above a city square or street, for example, and recover its pose relative to large-scale features of the scene should be achievable to a high degree of accuracy.
- Calibrating moving cameras is a more difficult task and the degree of success expected depends on how smooth (and so predictable) the motion of the camera is. Steadycams and those traveling on vehicles or along rigged wires may produce image sequences from which extrinsic parameters can be recovered fairly reliably. Image sequences with a large random component in their motion (such as those from a lapel mounted wearable camera) are likely to be problematic: at best calibration/registration is likely to be achieved only intermittently.

If camera calibration were to be seen as a valuable registration tool in and record and reuse package, we would suggest that extrinsic parameters be extracted using so-

called self-calibration. There are a number of reasons for this:

- Although constraints are imposed on the feature sets that can be used (for example, often only so many can be coplanar) these can easily be made clear to inexperienced users. It seems reasonable to expect a novice user to be able to interactively select appropriate model features with a minimum of instruction/guidance.
- The method is general; features can be selected from a large or small area of the model depending upon the images expected. It is however, always better to use features that are more or less evenly distributed across the image.
- The hypothesise-and-test structure of self-calibration systems means that measurements are provided of how well the output parameters fit the input image data. From this information it is possible to identify cases where the calibration has failed, and to take remedial action.
- Self-calibration also implicitly performs motion segmentation on the input feature set. Features whose movement is not accounted for by the recovered camera motion must be moving independently. This raises the possibility of identifying moving objects on the fly.

Although it is hard to make predictions regarding the possible use of camera calibration techniques to register images with models in a record and replay scenario without hand-on experience of the image sequences to be processed, there would seem to be some potential here. GPS and other sensor technologies can provide estimates of position with respect to a model, though the accuracy of positional estimates varies with local conditions and orientation information is hard to obtain. In the right circumstances, vision-based camera calibration methods can provide highly accurate positional and orientation estimates. These might be used in place of other sensor in some situations or alongside them in others. One possible scenario is the use of GPS data to initialize a camera calibration method, which might then provide a consistency check on the initial estimate and/or produce a more accurate, refined measurement.

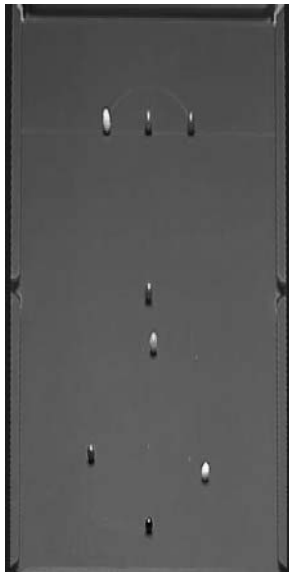
## **MAPPING, REPROJECTION AND SEGMENTATION**

Once the pose of a camera relative to some model has been recovered, model components can be overlaid on images (as seen in the ARToolkit [3]) and vice versa. Images may be mapped onto models to improve the perceived quality of the model or reprojected to provide new views of the image data. Figure 2, for example, shows interim results of a project aimed at linking real video of a snooker match with a 3D model of the game. The standard TV view in Figure 2a is reprojected to show a plan view in Figure 2b. These

techniques are now fairly standard, but their inclusion in a record and reuse toolkit could be of value.



a. Original view



b. Reprojected view

**Figure 2.** Reprojecting a TV view of a snooker game to provide a plan view.

A possibility that is perhaps less well explored is the use of the model to segment the image data. Given a registered camera and model it should be possible to:

- Search for model features that are obscured in the image(s), effectively highlighting objects and/or people appearing at key locations.
- Search for features that appear in the image but not in the model, these could then be segmented out and mapped onto the model, providing a more informed transfer of data from the real to the virtual environment.

The EQUATOR experiences appear to have generated large numbers of images, but to treat each one as an indivisible unit. Image segmentation provides opportunities to select from this huge repository of data. Image segmentation is a difficult task, not least because there is no clear definition

of what constitutes a “good” segmentation [4]. The availability of a registered three-dimensional model, however, raises the possibility of providing a clear definition of what is required of image segmentation by record and reuse applications. It may therefore be possible to produce model-based image segmentation tools that are both principled and useful.

#### ADDING CONTEXT

Images typically contain information about a sizeable area of the physical world. Image segmentation is one way of extracting this information; in most segmentation applications regions that are in some sense important are distinguished from those which are not. This approach can usefully be extended: segmentation and other image interpretation methods could be used to separate key features, objects and events from those which, though secondary, provide useful contextual information and those which are completely irrelevant.

#### Recognising Signs and Markers

Image segmentation could be used to identify predetermined world features, such as signs. Though this would be similar in some ways to the techniques used during image/model registration it should be noted that there are important differences. The features detected here would be larger and contain much more semantic information than the simple geometric primitives used in camera calibration. Some possible applications of these techniques include:

- Wall posters, street name and other road signs could be identified. Those containing text could be processed further and an attempt made to extract and recognise the characters they contain [5].
- Significant objects, such as lampposts, might be recognised and used to provide contextual information such as how easy it is/was to move to a well-lit area.
- Given a view of a larger area, it may be possible to identify the relative positions of groups of features. For example we could locate the runner in a chase game and determine how many people are in his/her vicinity.
- There is a sizeable literature on image classification. While whole-image techniques [e.g. 6] could be used to provide context they are restricted to identifying the broad nature of the surroundings (urban, rural, indoor, outdoor), which in a record and replay scenario are likely to be known beforehand.

#### Motion Context from Optical Flow

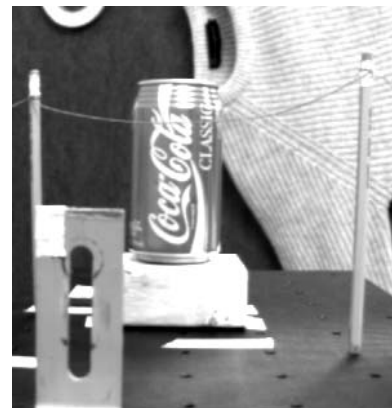
A central topic in computer vision for the last three decades has been the identification and description of motion. Two

broad forms of motion analysis and recovery should be considered:

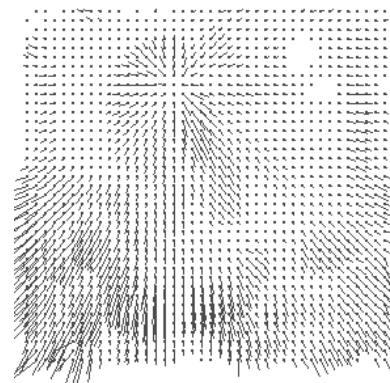
- Tracking, in which specific objects or features are identified in each frame of an image sequence and their motion across the image plane described. Tracking produces accurate motion estimates, but only for the particular objects being tracked. Figure 3 shows a group of straight lines, arising from a bus, being tracked [7] through an image sequence.
- Optic flow, which recovers the apparent motion of a local brightness/colour pattern at each pixel. Optic flow produces a dense, but often less accurate motion description in the form of a vector field. Figure 4 shows the optic flow field extracted [8] from a community standard image sequence in which a camera translates slowly towards a collection of objects.



**Figure 3.** Tracking a bus through a traffic scene [7].



a.



b.

**Figure 4.** The optic flow field obtained [8] from the community standard NASA “Coke Sequence”. One image from the sequence is shown in a., the flow field in b.

Tracking is useful for following objects that have been identified as interesting, but is harder to apply to a general image sequence. Optical flow, by contrast, can be used to give a general description of the motion from any sequence of images. For example, it would be possible to compute optical flow from a crowded city scene in order to follow patterns in crowd behaviour. Those patterns could be used as contextual information; one might track a particular individual across a city square, annotating a description of his/her motion with summaries of the general motion seen in the square extracted from an optic flow field.

Motion information could also provide information as to the actions of the camera operator. Several techniques exist for recovering the motion of a camera relative to the scene. These have been applied to identifying camera motions such as pan or zoom. The identification of these actions provides a description of the way in which the camera operator, as an intelligent observer, views the scene.

## CONCLUSION

Images and video sequences provide a rich source of information about the world, and a detailed record of events. The problem faced in record and reuse is accessing this information in order to create a searchable and reusable record of these events. In this paper we have outlined a few ways in which computer vision techniques may be applied to this domain.

The problem of registering the real and virtual worlds could be solved through the application of existing camera calibration techniques. Given the often dynamic environments explored in Equator off-line calibration of the intrinsic parameters, followed by extrinsic 'self-calibration' would seem to be a viable option in many circumstances.

Once the real and virtual worlds are registered, further information could be recovered from the mapping between them. Discrepancies between the model and the world could be used to identify objects in the world that are not modeled. These could be used to update the model, or to identify transient objects which may be of interest. The model could also be used to assist in image segmentation.

Finally, context could be added to images and videos by extracting visual information. The possibilities discussed here are the identification of objects that have particular interest, such as signs, and the extraction of motion to describe the movement within the scene, and of the camera through the scene.

## REFERENCES

1. Tsai, R., A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras, *IEEE Journal of Robotics and Automation RA-3*, 4 323-344 (1987).
2. Maybank, S. and O.D. Faugeras, A theory of self-calibration of a moving camera, *International Journal of Computer Vision* 8, 123-151 (1992).
3. Kato, H., M. Billinghurst, I. Poupyrev, K. Imamoto, K. Tachibana, Virtual Object Manipulation on a Table-Top AR Environment, *Proceedings of ISAR 2000*, Oct 5th-6th (2000).
4. Istiadi, O., S. Mills and T. Pridmore, A framework for evaluating image segmentation algorithms, *Proc ICPR 2004*, under review.
5. Clark, P. and M. Mirmehdi. Recognising text in real scenes. *International Journal on Document Analysis and Recognition*, 4(4):243-257, August 2002
6. Vailaya, A., A. Jain and H. Zhang, On image classification: city images vs landscapes, *Pattern Recognition* 31 (12) 1921-35 (1998).
7. Mills, S., T. Pridmore and M. Hills, Tracking through a Hough space with an extended kalman filter, *Proc British Machine Vision Conference* (2003).
8. Fu, S., T. Pridmore and S. Mills, Epipolar flow, in preparation.