Design and calibration of multi-camera systems for 3D computer vision: Supplementary material

Tom Botterill, Matthew Signal, Steven Mills, Richard Green

October 10, 2015

This document summarises the recommendations from our paper "Design and calibration of multi-camera systems for 3D computer vision: lessons learnt from two case studies". Section 1 provides a checklist for designing multi-camera systems, and Section 2 summarises trade-offs required in multi-camera system design.

1 Camera setup checklist

This checklist summarises our recommendations for the design of multi-camera systems.

1.1 Imaging enclosure design

Imaging enclosures allow lighting levels to be controlled, and provide a uniform background. This makes segmentation and feature detection simpler and more robust.

Structure Rigid so that cameras stay in calibration.

Background material Matte, uniform colour, different colour to subject, free of artefacts that create shadows that could be confused with the subject (seams).

Background construction Rigid mounting and backing enables background subtraction to be used, and avoids shadow artefacts because of wrinkling.

Size Large enough to surround view of subject from all camera viewpoints.

Ambient lighting Brushes can exclude most ambient light or sunlight.

Occlusions Keep cables, backdrop material, etc. tidied away so they can't move and occlude cameras or lights.

1.2 Camera positioning and lens selection

A software model of the camera, lens, background and subjects (at a range of dimensions) can be used to ensure that the subject is always sufficiently within the field-of-view.

Camera and lens field-of-view View all of subject (or as much as possible) from all cameras for all likely subject dimensions. It is much simpler to model objects which are completely within view.

Baselines Large enough for accurate 3D reconstruction, small enough for robust matching.

Positioning Subject is surrounded by background when imaged. Not surrounded by similar-looking/similar-coloured subjects (this increases the robustness of segmentation, model registration, and correspondence between views).

Object sizes Field measurements or preliminary studies.

Maintainability Labelling both ends of every cable is useful for identifying and fixing faults.

1.3 Lighting design

It is challenging to get even illumination for 3D scenes, especially where the camera's field of view or depth of field are large, where shadows and occlusions are common, or where object's are shiny and show specular reflections. Optimising light positions within a computer model is an effective way to design effective lighting configurations that provide even illumination.

1.3.1 Light sources

Source LEDs are often ideal but sometimes other sources are more suitable (see Jahr, 2006). Commercial machine vision lights are convenient, but are often have a narrow beam angle compared with bare LEDs.

Heat sinks LEDs generally require heat sinks to avoid overheating.

Power For moving scenes, more light=shorter shutter times=less motion blur. Alternatively, cameras with larger sensors, or monochrome (rather than RGB) cameras require less light to give the same image quality.

Power supply LEDs require "constant current" power supplies. LED power supplies running off AC may require additional capacitors to smooth high frequency variation.

1.3.2 Light arrangement

Aim to minimise variation (range of light levels) across region of interest. If the range is too high, cameras can't simultaneously image the brightest parts of the scene (where the sensor is saturated) and the darkest (where details are lost in noise).

Number of sources Increasing the number of sources provides robustness against shadows.

Positioning In general, widely-spaced lights, and lights positioned further from the subject, give more even illumination.

1.4 Cameras

1.4.1 Camera data acquisition

Camera interface Camera manufacturers provide APIs and example programs for grabbing images from cameras onto a PC. Robot Operating System (ROS) has support for integrating machine vision cameras into robot systems[1].

Bayer-to-colour conversion (demosaicking) On PC (potentially higher CPU consumption) vs. on camera (higher bandwidth to PC required).

Camera configuration Shutter times need to be long enough to capture enough light so that none of the image is too dark (signal-to-noise too low), while avoiding saturation (data loss). See "trade-offs". Auto-exposure and auto white balance cause image changes that may make registering views more challenging. Auto focus may affect calibration. Gain (image scaling) is best applied in software (if necessary), so saturation can be avoided.

Synchronisation Hardware synchronisation with trigger pulses is widely used and reliable.

Compression Compression artefacts in JPEG or MPEG images/video can affect computer vision, particularly if colour information is required, or features are tracked or matched between frames. Lossless PNG compression may reduce the required bandwidth to disk, at the expense of increased processing requirements.

Disks Higher bandwidth-to-disk is possible with solid state disks, rather than magnetic hard drives.

Reliability Checks for missing or duplicate frames are worthwhile if variable write speeds and CPU availability limit the software's ability to keep up with the camera.

1.4.2 Lens focus

The subject should be in focus throughout the range of positions where it might appear. This means the lens needs to be in focus at a range of depths for each location in the image.

Aperture Trade-off between depth of field (getting the entire scene in focus) and amount of light (shutter time necessary to get a high quality image).

Automated focus Lens can be adjusted to maximise a focus measure displayed in a GUI (e.g. the sum of squared differences between neighbouring pixels).

[1] `http://wiki.ros.org/pointgrey_camera_driver`

Focus object Objects designed to have sharp edges at a range of depths (e.g. a 3D mesh of wire) are useful for verifying images are in focus throughout the range of depths required.

Manual inspection Manual inspection generally requires zooming in to inspect all areas of the image, to see that edges are not blurred.

1.5 Camera calibration

Method Zhang's method is the most widely used method (e.g. in many libraries/toolboxes). Methods requiring only a single view or a small number of views may give inaccurate results.

Origin Coordinate frame can be chosen to suit application. Choosing an origin that is not a camera position provides robustness to one camera moving.

Target Large enough to fill field of view at closest region of interest.

Border A wide white border improves detector performance.

Orientation Some targets (e.g. checkerboard) may be detected in different orientations in different views. Patterns with unambiguous orientation are best.

Positioning Capture target images throughout the range of depths and field of view where the subject might appear.

Number of target images Capturing hundreds of images of the target in each pair of cameras reduces errors from calibration, for example from 8% of the depth for ten views to 1% for one hundred views.

Camera mounting Secure mountings with multiple points of attachment per camera reduces the chance of a loss of calibration.

Camera protection Physical protection (bars or cages) around cameras, cables and computers also reduces the chance of hardware failure or calibration loss.

Capturing images Capturing calibration images every time data is collected allows calibration changes to be detected. Cameras can be calibrated retrospectively if necessary.

2 Camera setup trade-offs

This section summarises the trade-offs required when designing multi-camera systems.

Camera/lens field-of-view

Narrow	$\Longleftrightarrow$	**Wide**
✓ Higher resolution on subject		× Lower resolution on subject
✓ Simpler lighting design		× Challenging to get even illumination
× Subject may be partially visible		✓ More of subject in view

Stereo baseline

Short	$\Longleftrightarrow$	**Long**
× Lower depth resolution		✓ Higher depth resolution, if localisation accuracy doesn't change
✓ Similar appearance between views		× More appearance change and occlusions between views
✓ More overlap between views		× More toe-in required to get good overlap between views (i.e. cameras point inwards)
✓ Localisation accuracy is more consistent between views		✓ More compact visual hull if voxel carving

Shutter time

Short	⟺	Long
✓ Less motion blur		× More motion blur
× Relatively higher imaging noise		✓ Relatively lower imaging noise
× Brighter lighting required (or strobe, or larger sensor)		✓ Less intense lighting required

Aperture

Small	⟺	Large
✓ Images in focus over greater depth of field		× Smaller depth of field
× Brighter lighting required		✓ Less light required

Framerate

Low	⟺	High
× Features move far between views, slow robust matching needed		✓ Features move a short distance frame-to-frame (small search area for tracking)
✓ Lower data rate and storage requirements (simple data capture)		× Higher data rate and storage requirements (especially with many high-resolution cameras)
✓ More CPU time available per frame		✓ After tracking, doesn't necessarily need more CPU time (e.g. if using a subset of keyframes)
× Potentially higher latency for robot control		✓ Potentially lower latency for robot control
✓ Less motion blur		× More motion blur
× Potential for missing data between frames (fast phenomena)		✓ Less missing data