

# Chapter 1

## Outline

Campbell (2016)

In this thesis we will examine the use of information and probability theory measures to track the pose, shape, and appearance of non-rigid articulated objects with multiple independently moving cameras. We demonstrate a hierarchical framework that uses the Hilbert-Schmidt independence criterion, to estimate first the shape and appearance, then the pose, then the activity of the dominant actor. At each level different features are employed, however the same kernel function and the same Hilbert-Schmidt norm are used to estimate the independence.

The most elementary level involves a single static camera tracking a single moving non-articulated, rigid object. This requires only that the appearance be tracked. Note that the appearance is defined as the combination of the gray-level intensity, the distance function, and the motion. This example is essentially TLD, which is a very robust and accurate tracker. I have made a contribution to show that my algorithm is an extension to TLD in this case. My tracker models a target as a collection of points. Each point is naturally endowed with x, y coordinates. The features for each point are the intensity at that pixel, the distance to the nearest edge, and the motion of that point from the previous frame. In each frame we find the set of points that maximises the HSIC with the reference (as well as maximising the self similarity and minimising the similarity with the negative training example). It is important to note that the HSIC fails for uniform distributions.

To track the pose we use some sort of optimisation scheme to search through potential poses. Options are: hierarchical search, PSO, LM.

1. Single camera tracking - basically done. 2. Multi camera tracking with known calibration - could be done without a prior labelling of the appearance (i.e just using

the MI scheme from previously and then creating a model on the fly.) 3. unknown calibration - investigate the additional parameters that need to be optimised. 4. activity recognition - constant alpha working (temporal synch + clustering / recognition) 5. DTW - allows unknown alpha, therefore may improve the performance.

## **1. Introduction**

- (a) Mixed reality
- (b) Pose, shape, appearance
- (c) Articulated objects
- (d) Markerless motion capture
- (e) Appearance modelling
- (f) Activity recognition
- (g) Independently moving cameras

## **2. Probability & Information theory**

- (a) Probability spaces
- (b) Random variables
- (c) Distributions
- (d) Moments
- (e) Expectation, variance, covariance
- (f) Entropy & uncertainty
- (g) Mutual information

## **3. Hilbert Schmidt independence criterion**

- (a) Euclidean spaces
- (b) Functional analysis
- (c) Hilbert spaces
- (d) Reproducing property
- (e) RKHS
- (f) Kernels
- (g) Covariance Measures

- (h) HSIC measure

#### 4. Appearance tracking

#### 5. Pose estimation

- (a) Bottom-up
- (b) Top-down: Analysis by synthesis
- (c) Calibration free - motion parameter estimation

#### 6. Activity recognition

- (a) Temporal synchronisation
- (b) Dynamic time warping

## 1.1 Information Theory

What is information theory? Where did it start? What questions does it seek to answer? What are the important concepts? – ¿ capacity of a channel to transmit information – ¿ what capacity does the system have to transmit information about data from two images?

The focus of this chapter is on the mathematical concept of information theory, which was first proposed by Claude Shannon in his seminal 1948 paper "A Mathematical Theory of Communication". Central to information theory is the concept of *channel capacity* which describes the ability of a *channel* to transmit information from a source to a destination. In particular, the noisy channel coding theorem specifies an upper bound on the rate at which information can be transmitted through a channel in the presence of noise. Notably, this upper bound depends only on the statistical characteristics of the channel.

### 1.1.1 Entropy & Uncertainty

How are entropy and uncertainty related? Make the relationship between uncertainty and variance etc quite clear. These links are going to be important later when we talk about HSIC, as it is dependent on the covariance, which is functionally related to the notions of uncertainty and therefore entropy.

## **1.2 Pose estimation**

### **1.2.1 Bottom up**

The principle difference between bottom-up and top-down is the use of the model. In BU estimation we do not know the model of the target. As we track a series of points, however, we can estimate the joint positions by computing the minimum spanning tree on the geodesic distance graph. We could do a graph cut type thing. Initially we want to find which edges are connected to which others. There is an edge between two points if the distances are preserved. We then want to find the joint positions. Similar to Steve's work.

### **1.2.2 Top down**

This is the main approach we are using. We have a known model of the target (i.e all the limb lengths and connections) that we wish to fit to the images(s) the best. For any particular pose we can evaluate the cost, then we just take the best pose. Note that we can think of the model as something that acts to preserve geodesic distances between points, i.e. it lets them move, but only in a way that preserves the shape of the target.

### **1.2.3 Calibration free**

This is currently future work. Track the pose independently in both images. Then we want to find the 3DT pose matrix that best matches the two 2D pose matrices. The principle question is how to propose any given 3DT matrix? probably hard given all the constraints etc.

# References

Campbell, J. (2016). *How to do Anything*. Dunedin, New Zealand: Non-existent Publishers.