FAST: Fast Adaptive Silhouette Area based Template Matching

Daniel Mohr
dmoh@tu-clausthal.de

Gabriel Zachmann
zach@tu-clausthal.de

Clausthal University, Germany

Clausthal University, Germany

Template matching is a well-proven approach in the area of articulated object tracking. Matching accuracy and computation time of template matching are essential and yet often conflicting goals.

In this paper, we present a novel, adaptive template matching approach based on the silhouette area of the articulated object. With our approach, the ratio between accuracy and speed simply is a modifiable parameter, and, even at high accuracy, it is still faster than a state-of-the-art approach. We approximate the silhouette area by a small set of axis-aligned rectangles. Utilizing the integral image, we can thus compare a silhouette with an input image at an arbitrary position independently of the resolution of the input image. In addition, our rectangle covering yields a very memory efficient representation of templates.

Furthermore, we present a new method to build a template hierarchy optimized for our rectangular representation of template silhouettes.

With the template hierarchy, the complexity of our matching method for n templates is $O(\log n)$ and independent of the input resolution. For example, a set of 3000 templates can be matched in 2.3 ms.

Overall, our novel methods are an important contribution to a complete system for tracking articulated objects.

Tracking an articulated object is a challenging task, especially, if the configuration space of the object has many degrees of freedom (DOF), e.g., the human hand has about 26 DOF. Most tracking approaches require the object to be in a predefined state. If this is not desired or impractical, one has to search the whole configuration space at initialization time. Such a global search, of course, only needs to find a coarse object state, which typically consists of two parts. The first part is a similarity measure between the object in each configuration and the observed object (e.g. images from a camera). The second part is an algorithm to combine the best matching configurations, detect and eliminate false positives, potentially reduce the search space, and estimate the final object state. The focus of this paper is the first part of such a global search method.

Tracking articulated objects, there is a large number of object configurations that have to be compared with an input image. Therefore, this comparison should be extremely fast. We propose a novel method for very fast approximate area silhouette comparison between model templates and input images. For one comparison, Stenger *et al.* [3] achieved a computation time proportional to the contour length of the template silhouette. We propose a new method, which reduces the computation time to be *independent* of the contour length and image resolution. It only depends on the desired accuracy of the template representation. This accuracy is a freely adjustable parameter in our approach. To achieve this, we first approximate all template silhouettes by axis-aligned rectangles, which is done in a preprocessing step. In the online phase, we compute

the integral image [2, 4] of the segmented input image. With this, the joint probability of a rectangle to match an image region can be computed by four lookups in the integral image. Moreover, we present an algorithm to build a template hierarchy that can compare a large set of templates in sublinear time.

Our *main contributions* are:

1. An algorithm that computes a representation of arbitrary shapes by a small set of axis-aligned rectangles with adjustable accuracy. This results in a resolution-independent, very memory efficient shape representation.

2. An algorithm to compare an object silhouette in $O(1)$. In contrast the algorithm proposed by [3] needs $O(\text{contour length})$.

3. We propose an algorithm to cluster templates hierarchically guided by their mutually overlapping areas. Our method builds on the recently developed batch neural gas clustering algorithm [1], which yields better results than more classical algorithms. As distance measure between the templates, we use the silhouette area overlap of regions of rectangular foreground/background regions. This hierarchy further reduces the matching complexity for n templates from $O(n)$ to $O(\log n)$.

Figure 1 gives an overview of our approach. In the preprocessing step, we generate the templates using an artificial hand models. Next, we extract the silhouettes and approximate them by axis-aligned rectangles. During the tracking, we compute the segmentation based on skin-color and compute the integral image of the segmentation. This integral image and the rectangles are then used to compute the joint probability between an input image and the templates.

[1] Marie Cottrell, Barbara Hammer, Alexander Hasenfuß, and Thomas Villmann. Batch neural gas. In *5th Workshop On Self-Organizing Maps*, 2005.

[2] Franklin C. Crow. Summed-area tables for texture mapping. In *SIGGRAPH: Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, volume 18, 1984.

[3] Bjorn Stenger, Arasanathan Thayananthan, Philip H. S. Torr, and Roberto Cipolla. Model-based hand tracking using a hierarchical bayesian filter. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 28, pages 1372–1384, 2006.

[4] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–511–I–518, 2001.
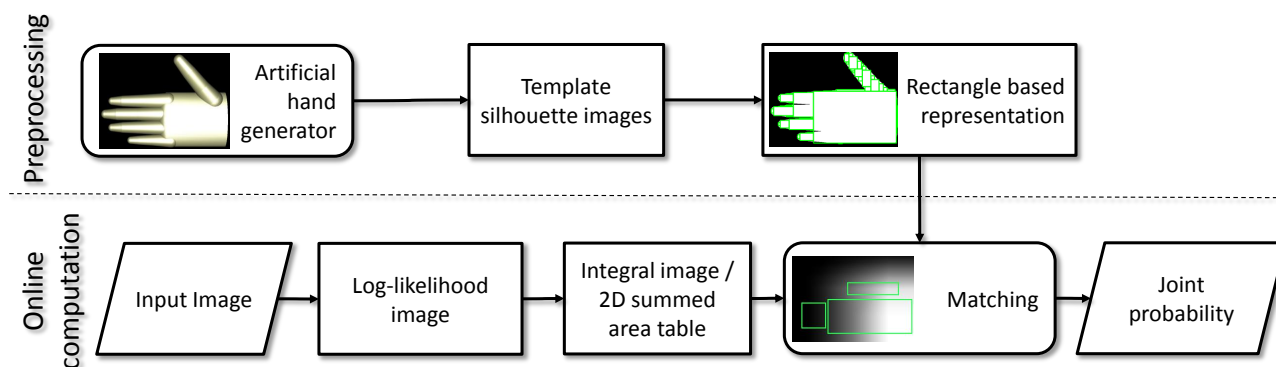
Figure 1: Overview of our approach using rectangle sets to approximate a silhouette. This speeds up the matching by a factor 5–30 compared to the approach proposed by Stenger *et al.* [3].