# Robust Automatic Tracking of Skin-Colored Objects with Level Set based Occlusion Handling

Liang-Guo Zhang[1,2,3], Xilin Chen[1,2], Chunli Wang[4], Wen Gao[1,2,3]

[1] Key Lab of Intell. Inf. Processing, Chinese Academy of Sciences (CAS), China
[2] Institute of Computing Technology, CAS, China
[3] Graduate University of the CAS, China
[4] School of IST, Dalian Maritime University, China
lgzhang@jdl.ac.cn

## 1 Introduction

The automatic tracking of skin-colored objects in videos is an important research topic. For example, the high performance tracking will definitely improve the gestural human computer interaction. Since our tracking work aims for the future visual gesture and sign language recognition (VGSLR), and especially SLR on medium or large vocabulary, some coarse tracking (e.g., only extracting the location or/and rough geometry information of the targets) is NOT sufficient. Thus, during tracking, we want to extract (/preserve) more rich information of the target such as silhouette (i.e., complete region within the target's boundary) for the future feature analysis. And we call it the *complete object region tracking*.

In this paper, we propose a hybrid tracking framework, which consistently combines the blob based temporal data association and the level set based spatial occlusion handling. Blob based temporal data association can provide rapid and accurate tracking when there's no occlusion; while level set based occlusion handling can guarantee the occluded shape recovery as precisely as possible. To the best of our knowledge, this is the first work on visual tracking of a varying number of skin-colored objects accurately to obtain complete region even during occlusion.

## 2 Methods

Our methods for tracking of the *complete regions of skin-colored objects* include (1) adaptive skin-colored blobs extraction, (2) blob-based temporal data association, (3) on-line shape prior learning, and (4) level set based spatial occlusion handling. A flowchart of our system is shown in Fig. 1.

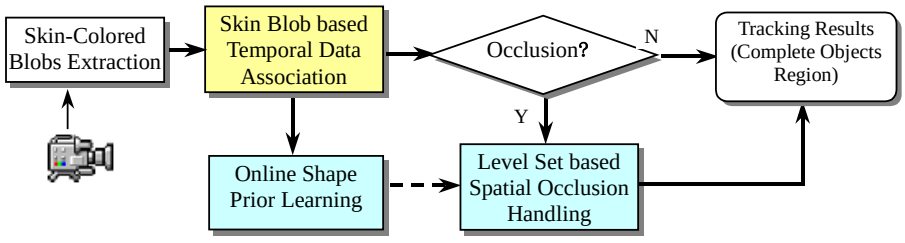


**Fig. 1.** System overview

To extract the skin-colored blobs, we filter the frame to get the skin mask by histogram based Bayesian classification of skin color. To associate each blob with the

existing or new tracks, inspired by Argyros' work [1], we enhance the original association mode. Further, different from [1], if occlusion occurs, we apply the process of ellipse fitting on the shape-recovered region after occlusion handling, which will guide the hypothesis update more accurately and robustly in a consistent manner. To obtain the shape prior, we propose an approach to online shape learning within the incremental updated subspace, where each target's shape is described as its "contour map" after distance transforming. To detect the occlusion, we just propose a simple hierarchical decision scheme by considering the overlap relationship among two or more blobs, i.e., from the bounding area level to the "actual" blob area level determined by data association before. For the shape recovery, we apply level set based active contour method. We borrow the idea of the level set evolution without re-initialization [2] and extend it by integrating the learned online shape prior to minimize the energy for occluded part of the tracked object. When occlusion occurs, we recover the occluded shape within the specific area, i.e., the bounding box area determined by the tracked object's elliptic area of the previous frame scaled up by a constant factor. After the occluded shape recovery, the corresponding hypothesis of the object will be updated.

Further, we introduce the spatio-temporal interaction constraint into the data association module to track one signer's face and hands, which demonstrate the direct use of our tracking approach in the context of the VGSLR task.

## 3 Results

All our experiments were performed on a PC with Intel Pentium-IV 3.2GHz CPU, 512MB RAM. Our own videos are captured by Sony DCR-PC120E in 25fps, where the image size is 320x240. We have implemented all algorithms with non-optimized C++ code, and tested it on several video sequences. First, we compare our tracking performance with that of Argyros's work [1] on the paper's test video (totally 3720 frames). Then, we experiment the method on the captured sequences "taiji" (totally 2929 frames) and "pui-sign" (totally 2595 frames) with complex motion patterns. Representative results of the tracked frames are provided in Fig. 2. Also, we test our approach on several sequences to track signer's face and hands for the VGSLR task. Please visit http://www.jdl.ac.cn/user/lgzhang/Research/VisTrack/gw2009.asp for the complete videos of tracking results.

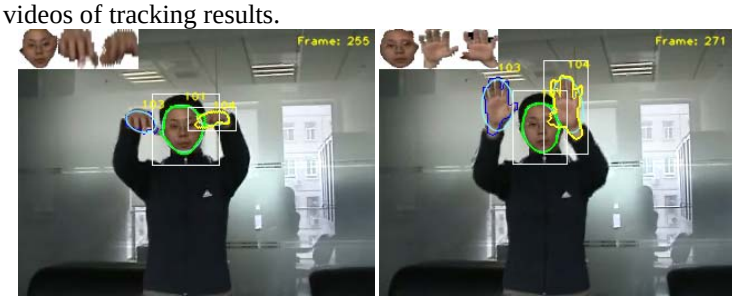


**Fig. 2**. Typical tracking results on sequence "taiji"

## References

1. Argyros, A.A., Lourakis, M.I.A. Real-time tracking of multiple skin-colored objects with a possibly moving camera. In: Proc. *ECCV'04*, pp.368-379, 2004
2. Li, C., Xu, C., Gui, C. Fox, M.D. Level set evolution without re-initialization: a new variational formulation. In: Proc. *CVPR'05*, vol.1, pp. 430-436, 2005