# Face Tracking using a Combination of Colour and Pattern Matching Based on Particle Filter

Matilde Gonzalez

Université Paul Sabatier
118 Route de Narbonne
F-31062 TOULOUSE
CEDEX 9

gonzalez@irit.fr

Christophe Collet

Université Paul Sabatier
118 Route de Narbonne
F-31062 TOULOUSE
CEDEX 9

collet@irit.fr

## ABSTRACT

Abstract. Robust real-time face tracking is an important and challenging task in computer vision applications. In this paper, we propose a novel particle filter algorithm to robustly track faces. Particle observations are computed by considering cue and appearance feature. Cue feature is used to identify skin regions, e.i. face and hands, while appearance is used to directly label targets. Normalized Cross Correlation (NCC) between an image template and particle samples is computed to robustly find the face among other skin regions. In other words, the image template is registered to a frame using particle filter to perform the optimization. Real-time is achieved by using integral images to compute image features. Evaluation results show the advantages and limitation of our approach.

## Keywords
Face Tracking, Particule filter, Real-time, Robust tracking, Occlusions, Color and appearance particles model

## 1 INTRODUCTION

Face tracking is a necessary step in many computer vision applications such as gesture recognition, human computer interaction, surveillance systems and sign language analysis [Gia09a][Gre05a][Mit07a]. The presence of noise, occlusions, fast dynamic changes and background complexity make face tracking a hard task. We focus our research in the domain of sign language analysis, and more specificaly sign language corpora automatic annotation. Sign languages are visuo-gestural languages used by deaf community as natural mean of communication. They are studied by linguists and computer scientists mainly through videos corpora of persons (signers) in spontaneous expressions or dialogs. One activity of these research consists in annotating the videos with relevant informations like basically, the glosses associated to signs (words in spoken language associated to the meaning of each sign). In order to facilitate and speed up this task we propose computer vision algorithm to automaticaly track pertinent component of the signer, e.g. hands and head. In the present paper we focus on

head tracking which is quite straightforward knowing that there is only one person to track in the video. But sign language expression leads the hands to move very fast, in a rather erratic way and very often to occlude the face. Head tracking makes it easier to track hands in the same time and it allows to segment the hand region of that of the head [Gon11a]. In this context, we don't necessarily care about the head pose or the facial features tracking. The solutions like Active Appearance Model (AAM) with 2D video [Zho10a][Pia10a] or video associated with depth camera [Smo14a], or landmark tracker [Uri15a], are not suitable and they do not handle well occlusions. More complex methods can handle occlusions but needs to be initialised by hand and are resource intensive [Zha13a].

Many tracking algorithms have been proposed to deal with these problems. Deterministic methods are based on a similarity cost function between a template and the current image incorporating, then, *a priori* information [Bra98a][Bir98a][Hag02a]. On the other hand stochastic methods are based on a dynamic model of the system. In the case of linear-Gaussian model, a Kalman filter estimates the posterior probability density function [Ste01a][Jan02a][Kir02a]. For non-Linear or non-Gaussian multi-modal distributions, the particle filter algorithm [Isa98a] has become very popular since it solves the limitation imposed by Kalman Filter.

Particle filter tracking algorithms usually use contours, colour features and appearance models

[Num03a][Mic04a][Gia09a][You10a]. Colour based algorithms have the inconvenient that same model could be used to represent different objects, e.g. skin blobs represent head and hands. Other solutions propose the fusion of several cues [Rad06a][Zha07a]. Particle observations are computed using a linear combination of various features, e.g. colour geometrical features. However it is becomes dependent on the coefficients used in the linear combination.

In this paper, we propose to combine *a priori* information with a dynamic model of the system. The proposed method includes global appearance information, an image template, and colour feature while performing particle filtering. The contributions of our work lie mainly in the fact that (i) multiple features are directly integrated on the computation of particle weights instead of a linear combination of the observation likelihood, (ii) geometric information is implicitly considered by the proposed model while computing colour cue likelihood and (iii) the registration of the template model represents an approach on pattern matching that can be described as image registration with particle filter optimization.

The remainder of this paper is organized as follows. Section 2 describes particle filter principle. Section 3 details the proposed model to combine colour and shape. Section 4 presents the proposed observation model to take into account multiple features. Section 5 shows the evaluation performed and last Section 6 presents our main conclusions.

## 2 PARTICLE FILTER

Visual tracking intends to estimates the state of the system that changes over time by using a sequence of noisy measurements. Bayes filter compute the posterior probability density function $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ of the current state $\mathbf{x}_t$ conditioned on all observations $\mathbf{z}_{1:t} = \mathbf{z}_1 \ldots \mathbf{z}_t$ with $\mathbf{z}_t$ to the observation vector obtained at time $t$. For a first-order Markov process, i.e. the state $\mathbf{x}_t$ depends only on $\mathbf{x}_{t-1}$, the probability density function $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ can be obtained in two stages: prediction and update. It is derived as

$$p(\mathbf{x}_t \mid \mathbf{z}_{1:t}) = k \cdot p(\mathbf{z}_t \mid \mathbf{x}_t) \cdot p(\mathbf{x}_t \mid \mathbf{z}_{1:t-1}) \qquad (1)$$

$$p(\mathbf{x}_t \mid \mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) \cdot p(\mathbf{x}_{t-1} \mid \mathbf{z}_{1:t-1}) \, dx_{t-1} \qquad (2)$$

where $k$ corresponds to a normalization term independent of $\mathbf{x}_t$. Eq.1 represents the update stage where the posterior probability density is computed using the observation likelihood $p(\mathbf{z}_t|\mathbf{x}_t)$ and the temporal prior distribution, $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$, over $\mathbf{x}_t$ given past

observations. Eq.2 corresponds to the prediction stage where the prior distribution for $t + 1$ is estimated by the convolution of the posterior distribution $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$ and the transition probability distribution $p(\mathbf{x}_t|\mathbf{x}_{t-1})$, i.e. the dynamic model of the system.

Particle filter (PF) [Bla98a] presents a good solution framework for tracking stochastic movements. It sequentially estimates, using random sampling to approximate the optimal solution, the states $\mathbf{x}_t$ of the system by implementing a recursive Bayesian filter by Monte Carlo simulations. The posterior probability density $p(\mathbf{x}_t|\mathbf{z}_t)$ of the current state $\mathbf{x}_t$ is approximated by a weighted particle sample set, $s_t^n, \pi_t^n {}_{n=1}^N$. PF maintain multiple hypothesis, i.e. each particle is a hypothetical state of the object, weighted by a discrete sampling probability $\pi_t^n \propto p(\mathbf{z}_t \mid \mathbf{x}_t = \mathbf{s}_t^n)$. Particle weights correspond to the observation generated by the hypothetical state and reflects the image feature relevance associated to each particle, see Section 4. The state $\mathbf{x}_t$ is finally estimated using the particle set and the associated weights.

The algorithm consists essentially of the following steps:

**input** : The particle set at time $t - 1$ : $\{s_{t-1}^n, \pi_{t-1}^n\}_{n=1}^N$

**output** : The expectation result at time $t$ : $E[\mathbf{x}_t]$

1. **Resample** $N$ particles from the set $\{\mathbf{s}_{t-1}^n, \pi_{t-1}^n\}_{n=1}^N$ to $\{s'{}_t^n, \frac{1}{N}\}_{n=1}^N$

2. **Propagate** each particle using the dynamic model $\mathbf{s}_t^n \sim p(\mathbf{x}_t \mid \mathbf{x}_{t-1} = \mathbf{s}'{}_{t-1}^n)$ to obtain $\{\mathbf{s}_t^n, \frac{1}{N}\}_{n=1}^N$

3. **Weight** particles with the image feature $\mathbf{z}_t$ as $\pi_t^n \propto p(\mathbf{z}_t \mid \mathbf{x}_t = \mathbf{s}_t^n)$ and normalize so that $\sum_{n=1}^N \pi_t^n = 1$

4. **Estimate** the tracking result of the object at time $t$ by $E[\mathbf{x}_t] = \sum_{n=1}^N \pi_t^n \mathbf{s}_t^n$

In this work each hypothetical state $\mathbf{s}_t^n = \mathbf{x}_t^n, \mathbf{y}_t^n$ corresponds to a state propagated through a first order autoregressive process model, $\mathbf{x}_t = \mathbf{x}_{t-1} + \eta$, where $\eta$ is a zero-mean Gaussian random variable and $\mathbf{x}_t^n, \mathbf{y}_t^n$ are the the coordinates of particle $n$ at time $t$.

## 3 OBJECT MODEL

Many methods combine shape and colour information [Ima98a][Mic04a][You10a] to robustly track face and hands. Face is modelled as a rectangular skin blob. These methods handle skin region occlusions by merging and separating blobs. However when hands fully occludes another skin object, one blob may be lost. Other methods use more complex shapes, but they
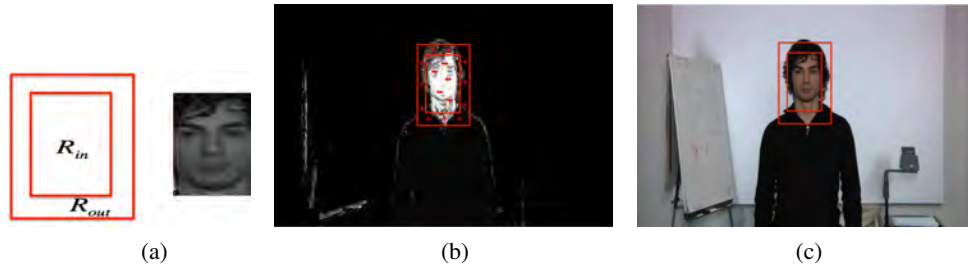
(a)                          (b)                          (c)

**Figure 1:** (a) Illustrates the proposed face model. (b) shows the rectangular model lying over the skin probability map when $\rho_{R_T}$ is maximal and (c) represents the best matching position for the template registration.

are computationally expensive and time-consuming [Bir98a]. Gianni *et al.* [Gia09a] model each skin body part as a cloud of points where each particle corresponds to a pixel. Occlusions between similarly coloured objects are handled using the exclusion principle [Due00a]. This method penalizes the particles when the objects are close, so that filters do not track the same skin object. The lack of shape information make this algorithm unstable since face and hands filters can be exchanged during the tracking.

In order to address these problems, we propose to use an image template of the subject in addition to shaped model, i.e. a rectangle $R_T$ divided in two regions of equal area (Figure 1(a)). $R_{int}$ and $R_{ext}$ define the sign of the pixel colour probability. Thus the weighted sum of skin probabilities inside $R_T$,

$$\rho_{R_T} = \sum_{\forall(x,y)\in R} R_T(x,y) \cdot p(c(x,y)|skin) \qquad (3)$$

is minimal when most of the pixels with high probability are inside $R_{int}$, Figure 1(b). This representation of the face seems to be a fair trade-off between robustness and speed. In addition, considering an image template of the face, updated up to time, make the face tracking more robust to occlusions between similarly coloured objects. Objects are directly labelled using a similarity

measure to avoid any exchange between the face filter and any other similarly coloured objects present in the frame, e.g. hands, other people, etc.

A face detection technique using Haar-like features [Vio02a] is used in this work to initialize the model size and the face template. This technique has shown robustness against illumination changes, scale and variation on facial expression for frontal faces. However as soon as the face is fully or partially occluded, detection tends to fail.

The skin model used to compute the skin probability map is built by using the pixels belonging to the face. First we use Kovac *et al* [Kov03a] explicitly defined model in the RGB colour space to extract a rough skin sample region from face. Then we transform sample pixels into the $YC_bC_r$ colour space and we use them to estimate the mean vector $\mu_S$ and the covariance matrix $\Sigma_S$ Eq.4, of the bivariate normal distribution $C_bC_r$ (Figure 2) that will be used to generate the skin probability map in the next frames.

$$\mu_S = \begin{bmatrix} \mu_{C_b} \\ \mu_{C_r} \end{bmatrix}, \qquad \Sigma_S = \begin{bmatrix} \sigma_{C_b}^2 & \sigma_{C_bC_r} \\ \sigma_{C_bC_r} & \sigma_{C_r}^2 \end{bmatrix} \qquad (4)$$

## 4 OBSERVATION MODEL

The entire set of visible features can be used as observation to compute weights. However, it is wiser to select few features that characterize the target among other objects in the frame. We propose to use multiple cues; colour, shape and appearance. Face model gives two measurements for weights computation. Firstly the weighted sum of skin probabilities $\rho$ inside $R_T$ and secondly the $NCC$ between the template and the image. The smallest $\rho_t^n$ and the largest $NCC_t^n$ leads to the largest likelihood function between the object model and the observation at the hypothetical state $n$.

### 4.1 Colour and Shape Features

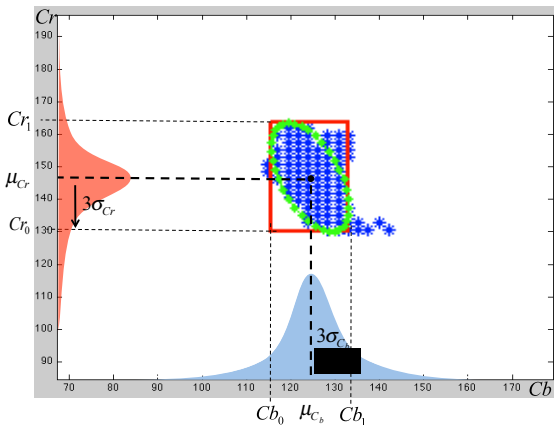The first observation measurement takes into account a rectangular shaped skin blob. First a specific model is



**Figure 2:** Bivariate normal distribution $C_bC_r$

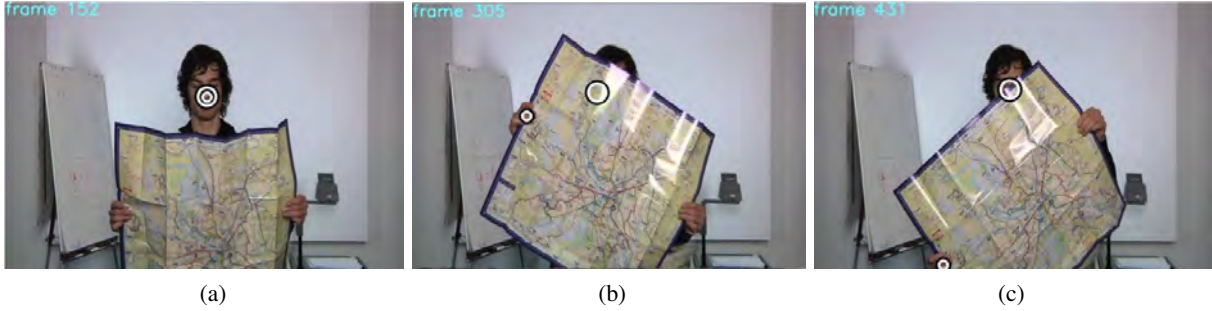<div style="text-align:center">(a)          (b)          (c)</div>

**Figure 3:** Tracking results before occlusion (a), during occlusion (b) and after occlusion (c). Big circles represent tracking results of our approach and small circles results of a single feature tracking.

built using the pixels samples from the model initialization step. Colorspace $YC_bC_r$ is used to compute the skin probability map $\mathbf{S}$ by only considering the chrominance components and avoid illumination changes influence. Let $\mathbf{c}(x,y)$ be the colour vector of the pixel at the coordinates $(x,y)$ and $p(\mathbf{c}(x,y)|skin)$ the probability of $\mathbf{c}(x,y)$ to belong to the skin colour class. Thus the first measure $\rho_t^n$ for a particle sample $\mathbf{s}_t^n = x,y$ is expressed as

$$\rho_t^n = \sum_{(x',y')\in R} f_s(x+x',y+y')p(\mathbf{c}(x+x',y+y')|skin) \quad (5)$$

$$f_s(x,y) = \begin{cases} -1 & \text{if } (x,y) \in R_{int} \\ 1 & \text{if } (x,y) \in R_{ext} \end{cases} \quad (6)$$

In order to speed up the algorithm and achieve real time, $\rho_t^n$ is computed using integral images [Vio02a] of the skin probability map. An integral image is an intermediate image representation allowing fast rectangular feature computation. Let $S_{(x,y)}$ be the pixel intensity in the skin probability map $S$ at the coordinates $(x,y)$. The value of the integral image $II$ at $(x,y)$ corresponds to the sum of $S_{(x,y)}$ and all pixels above and to the left. It is expressed as

$$II_{(x,y)} = \sum_{i=0}^{x}\sum_{j=0}^{y} S_{(i,j)} \quad (7)$$

Using this representation any rectangular region can be easily computed performing basic mathematical operations. Let $R_i$ be a rectangle defined by $(x_1,y_1)$ and $(x_2,y_2)$, the sum of the pixels inside the rectangle is computed using Eq.8 and $\rho_t^n$ is easily computed for each particle, Eq.9.

$$R_i = II_{(x_2,y_2)} + II_{(x_1,y_1)} - II_{(x_2,y_1)} - II_{(x_1,y_2)} \quad (8)$$

$$\rho_t^n = R_{ext} - R_{int} \quad (9)$$

This measurement implicitly considers geometric information since the best hypothetical state (particle) correspond to a maximum of skin pixels inside $R_{int}$.

## 4.2 Appearance Feature

The second measurement introduces spacial distribution information that is not considered before. This allows to determine where the face is when several skin objects are in the frame. Several measurements can be used to determine similarity between two objects. In this work, we use the Normalized Cross Correlation (*NCC*) for each hypothetical state $(x,y)$. It is defined as

$$NCC_t^n(x,y) = \frac{\sum_{x',y'} T(x',y')I(x+x',y+y')}{\sqrt{\sum_{x',y'} T(x',y')^2 \sum_{x',y'} I(x+x',y+y')^2}} \quad (10)$$

where $I$ represents the image and $T$ the face template. When *NCC* is closer to 1 the best correspondence between $I$ and $T$ is achieved.

## 4.3 Multiple Features Observation

Using colour cue to track face and hands has shown good results in a controlled environment. However additional work is required to label each object, e.g. anatomical models. On the other hand tracking by matching a template might be time-consuming depending on the way of optimization. In this paper we propose to combine both features directly in particle filter weight computation. This has as advantage that the skin object is directly labelled by matching the template and no further work is required to distinguish face from hands and vice versa. In addition using the *NCC* in particle filter is explained, in other words, as an image registration with particle filter optimization which is not time consuming, also integral image representation is used to speed up the algorithm.

Considering colour and appearance measurements, particle weights are defined as

$$\pi_t^n = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(k-\rho_t^n/NCC_t^n)}{2\sigma^2}} \quad (11)$$

where the parameter $\sigma$ ensures the effectiveness and diversity of particle resampling and $k$ corresponds to a normalization term.

**Figure 4:** Tracking results of our method (big circle) and tracking results considering a single feature (small circles).



**Figure 5:** Tracking results of our method (big circle) and tracking results considering a single cue (small circles).

## 5 EXPERIMENTAL RESULTS

In this section we present the conducted experiments in order to show the robustness of our tracker. A data set containing difficult tracking cases is composed of 4500 frames. It shows several occlusions between the head and the hands, another person and other objects. Firstly we compare our contribution against a model considering only colour cue. Secondly we implement our head tracker in the tracking system proposed in [Gia09a] to show the improvements that our tracker offers to the system.

In the first evaluation framework three experiments are conducted to show difficult tracking cases. The first scenario (Exp. 1) shows a subject holding a city map. The subject fully occludes the face with the map. Figure 3 shows tracking results for the first scenario. Before the first occlusion both methods show similar results. When the head is completely occluded by the map both methods try to reach another skin coloured

| Method | Exp. 1 | Exp. 2 | Exp. 3 |
|---|---|---|---|
| Single feature approach | 32.9 | 74.1 | 34.3 |
| Proposed method | 86.89 | 96.73 | 76.6 |

**Table 1:** GTR(%) for three experiments. Exp. 1: the subject fully occludes the face with a city map. Exp. 2: the subject passes his hands over the head in several directions and various speeds. Exp. 3: other person passes in front of the subject.

region. When the head is visible again our method returns to track the head while the other method stays in a local maximum.

The second scenario (Exp. 2) shows the same subject passing his hands over the face. The proposed method gives good results as long as a part of the face is still visible. For example when both hands fully occludes the face, the results is influenced by the low similarity measurement. The lack of shape information make the other algorithm less robust and track the hands instead the face. Figure 4 shows tracking results of series of occlusion.

The third scenario (Exp. 3) shows a person walking through the scene and fully occluding the head of the subject. The proposed tracker follows the subject face whereas the colour based tracks the other person (Figure 5). This case show the robustness of the tracking by adding an image template against similar objects tracking.

Table 1 shows quantitative results of Good Tracking Rate (GTR) for the three experiments described above. The low rate of the colour based method is due to the attachment of the filter to local maximum during or after occlusions. Our filter achieves better results since the local maximum are avoid by considering subject information in addition to colour and shape.
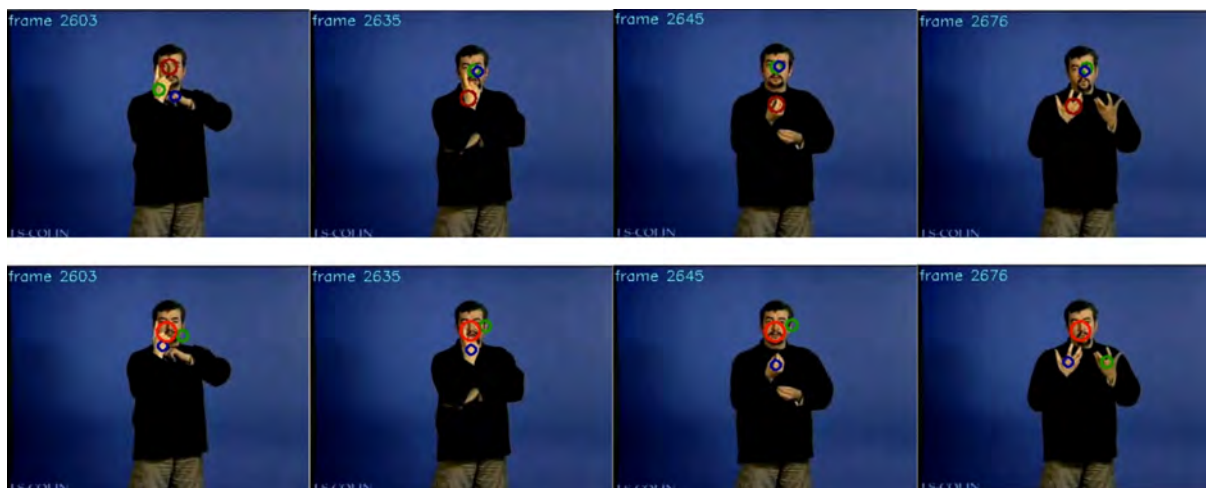
**Figure 6:** Tracking results of Gianni *et al.* [Gia09a] (top line) and our method (bottom line). Red circle stands for head trackers, blue and green ones for hands.

In the second evaluation framework we have implemented our head tracker in the system proposed by Gianni *et al.* in [Gia09a]. They models each body part (head and both hands) as a cloud of points and allows filters to interact between them to avoid tracking the same skin object. The same principle is considered in our case but we replace the filter that tracks the head with the proposed method. We conducted the evaluation using a corpus where native signers perform deaf sign language, LS-COLIN corpus [Bra01a], to evaluate the robustness of the proposed approach against occlusions and high dynamics within a real context.

The evaluation has been performed using videos of size 320x240 and 720x576. The number of particles for Gianni *et al.* tracker depends on the size of the skin region to track thus to the size of the video. We have chosen for the first video 1750 particles for each hand filter and 3500 particles for the head and the double number of particles for the second video sequence. For the proposed tracker we have chosen 3500 particles for the hands and 1500 particles for the head for both video sizes. Figure 6 shows tracking results of sequences with complex occlusions; two hands occluding the face.

Results from our tracker for this sequence show better performances that the tracker in [Gia09a] since head

tracker do not exchange with any hand filter. Table 2 shows the evaluation results for head and hand for two sequences of about 3500 frames. Robust head tracking gives already better results for hands tracking.

We conducted our experiments in a laptop Intel Core i7-5500U CPU, 2.4 GHz and 16 GB of RAM. The first evaluation with only the head tracker runs at 30 frames per second, and the second with the hands and head trackers runs at 3.3 frames per second with full resolution images. Hand tracker doesn't have any particulare code optimization, that is why it takes lots more time to process frames beside the high number of particles used per hand ($3500 \times 2$). The overall evaluation framework validates the proposed method, shows the robustness of the approach and the improvements of implementing it in a tracking system.

# 6 CONCLUSION AND PERSPECTIVES

In this paper we have addressed real-time head tracking by integrating global information from an image template into particle filtering. We propose an improved particle filter based algorithm for efficient and robust head tracking. Stability is increased when other skin regions are present in the frame. The tracking is performed using the colour and the similarity measure between a template belonging to the model and each particle sample. Since the method uses colour cue and image template of the object to track, the proposed algorithm can be used in many tracking applications. However when head rotation is out-of-plane the similarity measure may by low even at the optimal position.

In future works we intend to make our method template adaptive so that the template can be updated on time. This method will be integrated in our semi-automatic

|  | Seq.1 | | Seq.2 | |
|---|---|---|---|---|
| Method | Head | Hands | Head | Hands |
| Gianni *et al.* | 90.78 | 88.3 | 74.2 | 79.4 |
| Proposed method | 99.9 | 96.65 | 99.6 | 96.7 |

**Table 2:** Tracking evaluation. Tracking the face robustly arises the quality of the results also for hands tracking.

annotation framework for sign language corpora to enhance sign temporal segmentation and glossing.

# 7 REFERENCES

[Bir98a] Birchfield,S. Elliptical head tracking using intensity gradients and color histograms. In 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Proceedings. pp.232-237, 1998.

[Bla98a] Black,M., and Jepson,A. A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions. In Computer Vision ECCV'98, pp.909-924, 1998.

[Bra01a] Braffort,A., Cuxac,C., Choisier,A., Collet,C., Dalle,P., Fusellier,I., Gherbi,R., Jausions,G., Jirou,G., Lejeune,F., et al. Projet LS-Colin. Quel outil de notation pour quelle analyse de la LS ? In Journées Recherches sur la langue des signes. UTM, Le Mirail, Toulouse, 2001.

[Bra98a] Bradski.G. Real time face and object tracking as a component of a perceptual user interface. In Fourth IEEE Workshop on Applications of Computer Vision, 1998. WACV'1998, pages pp.214-219.1998 .

[Due00a] Duetscher,J., Blake,A., and Reid,I. Articulated body motion capture by annealed particle filtering. In IEEE Computer Society Computer Vision and Pattern Recognition, vol.2, pp. 126-133. 2000.

[Gia09a] Gianni,F., Collet,C., and Dalle,P. Robust tracking for processing of videos of communications gestures. In Gesture-Based Human-Computer Interaction and Simulation, LNAI 5085, Springer-Verlag, pp.93-101, 2009.

[Gon11a] Gonzalez,M., and Collet,C. Robust body parts tracking using particle filter and dynamic template. In $18^{th}$ IEEE International Conference on Image Processing (ICIP 2011), IEEE, Brussels, Belgium, pp 529–532, 2011

[Gre05a] Greiffenhagen,M., Ramesh,V., and Comaniciu,D. Statistical modeling and performance characterization of a real-time dual camera surveillance system. Apr. 22 2005. US Patent App. 11/112, 930.

[Hag02a] Hager,G., and Belhumeur,P. Efficient region tracking with parametric models of geometry and illumination. In IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(10): pp.1025-1039, 2002

[Ima98a] Imagawa,K., Lu,S., and Igi,S. Color-based hands tracking system for sign language recognition. In Third IEEE International Conference on

Automatic Face and Gesture Recognition, pp.462-467. 1998.

[Isa98a] Isard,M., and Blake,A. Condensation-conditional density propagation for visual tracking. In International journal of computer vision 29, pp.5-28, 1998.

[Jan02a] Jang,D., Jang,S., and Choi,H. 2D human body tracking with structural kalman filter. In Pattern Recognition, 35(10): pp.2041-2049, 2002.

[Kir02a] Kiruluta,A., Eizenman,M., and Pasupathy,S. Predictive head movement tracking using a kalman filter. In IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 27, pp.326-331, 2002.

[Kov03a] Kovac,J., Peer,P., and Solina,F. Human skin color clustering for face detection. In EUROCON International Conference on Computer as a Tool, vol.2, pp.144-148. 2003.

[Mic04a] Micilotta,A., and Bowden,R. View-based location and tracking of body parts for visual interaction. In Proc. of British Machine Vision Conference, volume 2, pp.849-858. 2004.

[Mit07a] Mitra,S., and Acharya,T. Gesture recognition: A survey. In IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 37, pp.311-324, 2007.

[Num03a] Nummiaro,K., Koller-Meier,E., and Van Gool,L. An adaptive color-based particle filter. In Image and Vision Computing, 21(1): pp.99-110, 2003

[Pia10a] Piater,J., Hoyoux,T., and Du,W. Video analysis for continuous sign language recognition. In Proceedings of 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, pp.22-23. 2010.

[Rad06a] Raducanu,B., and Vitrià,Y. A robust particle filter-based face tracker using combination of color and geometric information. In Image Analysis and Recognition, pp.922-933, 2006.

[Smo14a] Smolyanskiy,N., Huitema,C., Liang,L., and Anderson,S.E. Real-time 3D face tracking based on active appearance model constrained by depth data. In Image and Vision Computing 11, vol.32, pp.860-869, 2014

[Ste01a] Stenger,B., Mendonça,P., and Cipolla,R. Model-based hand tracking using an unscented kalman filter. In Proc. British Machine Vision Conference, volume 1, pp.63-72. 2001.

[Uri15a] Uřičář,M., Franc,V., Thomas,D., Sugimoto,A., and Hlaváč,V. Real-time multi-view facial landmark detector learned by the structured output SVM. In 11th IEEE International Conference and Workshops on Automatic Face and

Gesture Recognition (FG 2015), Ljubljana, pp. 1-8. 2015.

[Vio02a] Viola,P., and Jones,M.: Robust real-time object detection. In International Journal of Computer Vision 57, pp.137-154, 2002.

[You10a] YoungJoon,C., Seung Ho,S., Kyusik,C., and TaeYong,K Real-time user interface using particle filter with integral histogram. In IEEE Transactions on Consumer Electronics 56, pp.510-515, 2010.

[Zha07a] Zhao,L., and Tao,J. Fast facial feature tracking with multi-cue particle filter. In International Conference on Image and Vision Computing. Hamilton, New Zealand. pp.7-12, 2007.

[Zha13a] Cai,Z.W., Wen,L.Y, Cao,D., Lei,Z., Yi,D., and Li,S.Z.: Person-specific face tracking with online recognition. In 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013, Shanghai, pp.1-6. 2013.

[Zho10a] Zhou,M., Liang,L., Sun,J., and Wang,Y. AAM based face tracking with temporal matching and face segmentation. In 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, pp.701-708. 2010