# Visual tracking of hands, faces and facial features as a basis for human-robot communication

Maria Pateraki and Haris Baltzakis and Panos Trahanias

*Abstract*— **This paper presents an integrated approach for tracking hands, faces and specific facial features (eyes, nose, and mouth) of multiple persons in image sequences. For hand and face tracking, we employ a state-of-the-art blob tracker which is specifically trained to track skin-colored regions. The skin-color tracker is extended by incorporating an incremental probabilistic classifier used to maintain and continuously update the belief about the class of each tracked blob, which can be left-hand, right hand or face as well as to associate hand blobs with their corresponding faces. Then, in order to detect and track specific facial features within each detected facial blob, a hybrid method consisting of an appearance-based detector and a feature-based tracker is employed. The proposed approach is intended to provide input for the analysis of hand gestures and facial expressions that humans utilize while engaged in various conversational states with robots that operate autonomously in public places. It has been integrated into a system which runs in real time on a conventional personal computer which is located on the mobile robot itself. Experimental results confirm its effectiveness for the specific task at hand.**

## I. INTRODUCTION

In this paper we propose an integrated approach to identify and track human hands, human faces and specific facial features in image sequences. The proposed approach is mainly intended to support natural interaction of multiple persons with autonomously navigating robots that guide visitors in museums and exhibition centers and, more specifically, to provide input for the analysis of hand gestures and facial expressions that humans utilize while engaged in various conversational states with a robot. The proposed approach extends, combines and integrates a set of state-of-the-art techniques to solve three different but closely related problems: (a) identification and tracking of human hands and human faces which are detected as skin-colored blobs, (b) robust classification of the identified tracks to faces and hands, and, finally, (c) identification and tracking of specific facial features (eyes, nose and mouth) within each recognized facial blob.

For the first of the above defined problems (identification and tracking of human hands and faces) a variety of approaches have been reported in the literature [1]. Several of them rely on the detection of skin-colored areas,e.g. [2]. The idea behind this family of approaches is to build appropriate color models of human skin and then classify image pixels based on how well they fit to these color models. On top of that, various segmentation techniques are used to cluster skin-colored pixels together into solid blobs that correspond to human hands and/or human faces.

In contrast to blob tracking approaches, model based ones [3] do not track objects on the image plane but, rather, in a hidden model-space. This is commonly facilitated by means of sequential Bayesian filters such as Kalman or particle filters. The state of each object is assumed to be an unobserved Markov process which evolves according to specific dynamics and which generates measurement predictions that can be evaluated by comparing them with the actual image measurements. Model based approaches are computationally more expensive and often require the adoption of additional constraints for the dynamics of the system and for the plausibility of each pose but they inherently provide richer information regarding the actual pose of the tracked human as well as the correspondence of specific body parts with the observed image.

In the field of facial feature detection and tracking a number of approaches have already been presented in the existing literature [1]. Still, complexities arising from inter-personal variation (i.e. gender, race), intra-personal changes (i.e. pose, expression) and inconsistency of acquisition conditions render the task difficult and challenging. Related methods can be categorized on the basis of their inherent techniques. Color-based approaches were exploited in earlier systems by analyzing prior knowledge of color properties of facial features. Although this category of approaches is sensitive to illuminations and head pose changes, it still gains attention in the literature [4], as it succeeds fast detection. Shape- or model-based approaches represent salient facial features via a model and its parameters are optimized to fit to the observations. Earlier examples included deformable templates, graph matching, active contours, Hough transformation, e.g. [5], as well as Active appearance models (AAM) [6]. Later, many derivatives based on AAM have been proposed,e.g. [7] and although they may lead to accurate feature detection results, they may also converge to incorrect local minima due to improper initializations and feature variances and with a cost in time. Approaches based on machine learning techniques, like Principal Components Analysis, Neural Networks and Adaboost Classifiers,e.g [8] are relatively robust to illumination differences, but require a large number of images for training and are computationally

This work was partially supported by the European Commission under contract numbers FP6-045388 (INDIGO project), FP7-248258 (First-MM project) and FP7-270435 (JAMES project).

M. Pateraki and H. Baltzakis are with the Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH), 71110 Heraklion, Crete, Greece.

P. Trahanias is with with the Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH), 71110 Heraklion, Crete, Greece and the Department of Computer Science, University of Crete, 71409 Heraklion, Crete, Greece.

{pateraki,xmpalt,trahania}ics.forth.gr

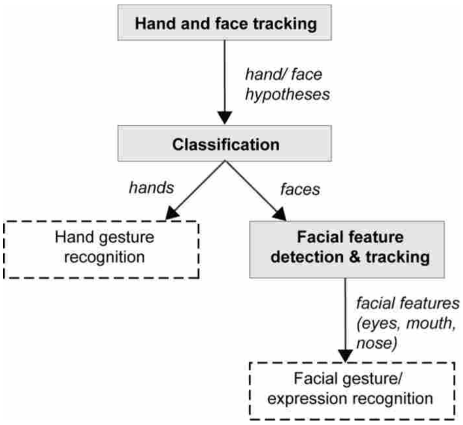
Hand and face tracking
hand/ face hypotheses
Classification
hands
faces
Hand gesture recognition
Facial feature detection & tracking
facial features (eyes, mouth, nose)
Facial gesture/ expression recognition

Fig. 1. Block diagram of the proposed system for hands and face tracking.

less efficient in the case of high resolution video sequences.

## II. METHODOLOGY

A block diagram of the components that comprise the proposed approach is depicted in Fig. 1. The first block in Fig. 1 is the hand and face color-based tracker. The second step of the proposed system involves the classification of the resulting tracks into tracks that belong to facial blobs and tracks that belong to hands; left and right hands are also classified separately in this step.

Hand trajectories are forwarded to the hand-gesture recognition system (not described in this paper), while facial regions are further analyzed in order to detect and track specific facial features (eyes, nose and mouth) and to facilitate facial gestures and expression recognition at a later processing stage of the system (also not part of this paper).

In the following sections we describe each of the above mentioned components in detail.

## III. HAND AND FACE TRACKING

In this work, hand and face regions are detected as solid blobs of skin-colored, foreground pixels and they are tracked over time using the propagated pixel hypotheses algorithm analyzed in detail in [9]. This specific tracking algorithm allows the tracked regions to move in complex trajectories, change their shape, occlude each other in the field of view of the camera and vary in number over time.

Initially, the foreground area of the image is extracted by the use of a background subtraction algorithm [10]. Then, foreground pixels are characterized according to their probability to depict human skin and then grouped together into solid skin color blobs using hysteresis thresholding and connected components labeling. The location and the speed of each blob is modelled as a discrete time, linear dynamical system which is tracked using the Kalman filter equations. Information about the spatial distribution of the pixels of each tracked object (i.e. its shape) is passed on from frame to frame by propagating a set of pixel hypotheses, uniformly sampled from the original object's projection, to the target frame using the object's current dynamics, as estimated by the Kalman filter. The density of the propagated pixel hypotheses provides the metric which is used in order to associate observed skin-colored pixels with existing object tracks in a way that is aware of each object's shape and the uncertainty associated with its track.

Figure 2 demonstrate the operation of the employed hand and face tracker on a test sequence which involves a man performing hand gestures in an office environment. As can be easily observed, the tracker succeeds in keeping track of all the three hypotheses, despite the occlusions and the blob merging events introduced at various fragments of the sequence.

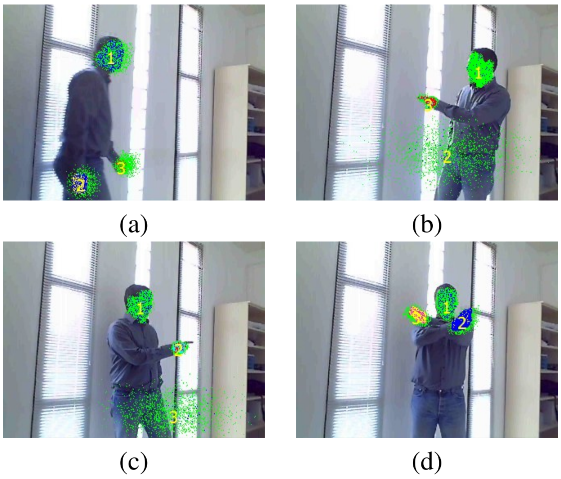

Fig. 2. Indicative tracking results in four frames of the office image sequence used in the previous example. In all cases the algorithm succeeds in correctly tracking the three skin-colored regions.

## IV. CLASSIFYING BETWEEN HANDS AND FACES

To proceed with higher level tasks, like hand gestures and facial expressions recognition, one has to distinguish between tracks that belong to hands and tracks that belong to faces. Moreover, for hand tracks, one has to know which tracks belong to left hands and which tracks belong to right hands. Towards this goal, we have developed a technique that incrementally classifies a track into one of three classes: faces, left hands and right hands.

The input of the technique is a feature vector $O_t$ which is extracted at each time instant $t$ and is used to update the belief of the robot $B_t$ regarding the class $F$ of each track. The feature vector $O_t$ consists of the following components:

- The periphery-to-area ratio $r_t$ of the current track's blob. The ratio $r_t$ is normalized to the corresponding ratio of a circle and provides a measure of the complexity of the blob's contour. It is expected that hands will generally have more complex contours than faces, i.e. larger values for $r_t$.
- The vertical and the horizontal components $u_t$ and $v_t$ of the speed of a tracked skin-colored blob. The intuition behind this choice is that hands are generally expected to move faster than faces and faces are not expected to have large vertical components in their motion.
- The orientation $\theta_t$ of the blob. It is expected that faces will tend to have orientations close to $\pi/2$.

- The location $l_t$ of the blob within the image. This location is relative to the location of each possible head hypothesis and it is normalized according to the radius of this head, as it will be explained later in this section.

We define the belief $B_t$ of the robot at time instant $t$ to be the probability that the track belongs to class $f$, given all observations $O_i$ up to time instant $t$. That is:

$$
\begin{aligned}
B_t &= P(F=f|O_1,\ldots,O_{t-1},O_t) && (1)\\
&= \frac{P(O_t|F=f,O_1,\ldots,O_{t-1})P(F=f|O_1,\ldots,O_{t-1})}{P(O_t|O_1,\ldots,O_{t-1})} && (2)
\end{aligned}
$$

Since the denominator $P(O_n|O_1,\ldots,O_{t-1})$ is independent of $F$, we can substitute it with $1/\alpha$ and we obtain

$$
\begin{aligned}
B_t &= \alpha P(O_n|F=f,O_1,\ldots,O_{t-1})P(F=f|O_1,\ldots,O_{t-1}) && (3)\\
&= \alpha P(O_n|F=f,O_1,\ldots,O_{t-1})B_{t-1} && (4)
\end{aligned}
$$

The above equation defines an incremental way to compute $B_t$, i.e. to classify the track by incrementally improving the belief $B_t$ based on the previous belief $B_{t-1}$ and the current observations.

By assuming the Markov property and the independence assumptions indicated by Figure 3, the computation of $B_t$ can be simplified as:

$$B_t = \alpha P(O_t|F=f)B_{t-1} \qquad (5)$$

In order to compute the term $P(O_t|F=f)$ in the right hand of (5), we assume the naive Bayes classifier depicted in the graph of Fig. 3 (b).

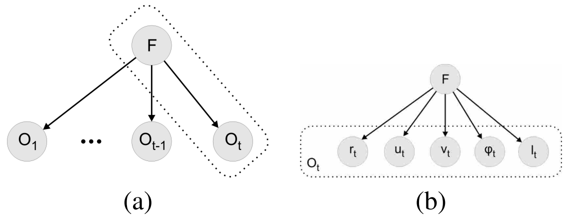

Fig. 3. (a) Bayes graph encoding the independence assumptions of our approach, (b) The naive Bayes classifier used to compute the $P(O_t|F=f)$.

According to this graph, we have at time instant $t$:

$$
\begin{aligned}
P(O_t|F) &= \frac{P(F,O_t)}{P(F)} && (6)\\
&= P(r_t|F)P(u_t|F)P(v_t|F)P(\theta_t|F)P(l_t|F) && (7)
\end{aligned}
$$

All the probabilities in the right side of (7) can be estimated according to training data. Hence they are encoded and stored in appropriate look-up tables, thus permitting real-time computations.

The lookup tables for $P(r_t|F)$, $P(u_t|F)$, $P(v_t|F)$ and $P(\theta_t|F)$ are depicted in Fig. 4. They are 1D lookup tables encoding the relevant quantity ($r$, $u$, $v$, or $\theta$) with the probability of appearance of this quantity in the training set. These lookup tables are identical for left hands and right hands but they are different in the case of faces. This is because, the relevant quantities are not expected to vary significantly between left and right hands but, as can be easily observed in Fig. 4, they differ significantly in the case of faces.

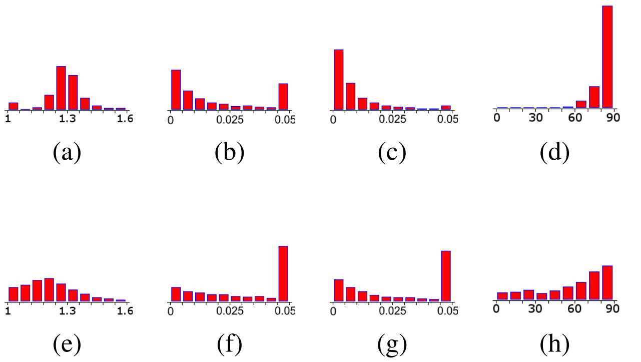

Fig. 4. 1D Look-up tables used for the computation of Equation (7). (a): $P(r_t|F=face)$, (b): $P(u_t|F=face)$, (c): $P(v_t|F=face)$, (d): $P(\theta_t|F=face)$, (e): $P(r_t|F=hand)$, (f): $P(u_t|F=hand)$, (g): $P(v_t|F=hand)$, (h): $P(\theta_t|F=hand)$.

$P(l_t|F)$, which is the probability of a blob being observed at location $l_t$ given its class $F$, is computed and stored differently for faces and differently for hands.

For faces, $P(l_t|face)$ is retrieved as the probability for a facial blob to be centered at this specific image location $l_t$. Obviously, the 2D lookup table for $P(l_t|face)$ depends on the actual application at hand and involves assumptions about the pose of the camera and the relative location of the human(s) with respect to the camera. In our case, which involves a human-robot interaction application, we assumed a camera placement such that the field of view of the camera includes the upper body part of one or more humans standing at a convenient distance between 0.5m and 2m in front of the robot. The actual lookup table that we compiled and used in our experiments is depicted in Fig. 5(a).

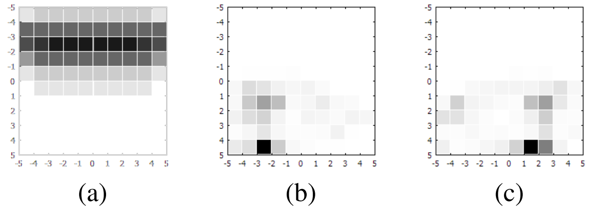

Fig. 5. 2D Look-up tables used for the computation of $P(l_t|F)$ in Equation (7). (a) for faces, (b) for left hands, (c) for right hands.

For hands, $P(l_t|left\,hand)$ and $P(l_t|right\,hand)$ are computed relatively to the location of the corresponding person's face. Since we don't know which is the corresponding person's face, we marginalize over all possible face hypotheses. That is, for $P(l_t|right\,hand)$ we have:

$$P(l_t|right\,hand) = \sum_h P(l_t|right\,hand, h=face)P(h=face) \qquad (8)$$

and similarly for the left hand:

$$P(l_t|left\,hand) = \sum_h P(l_t|left\,hand, h = face)P(h = face) \tag{9}$$

Figures 5(b) and 5(c) depict the resulting lookup tables for $P(l_t|right\,hand, h = face)$ and $P(l_t|left\,hand, h = face)$.

Evidently, the discriminative power of each of the above described features is significantly related to the application scenario at hand. That is, different training data should be used for different applications, which is especially true for the speed components $u_t$ and $v_t$ and for the expected image location $l_t$ of hand and facial blobs. In all experiments reported in this paper we have trained our classifier assuming a human-robot interaction setup which involves human(s) standing at a distance of approximately $1m$ from a camera which is placed at approximately 1.2 meters above the ground (i.e. the robots chest).

## V. DETECTION AND TRACKING OF FACIAL FEATURES

For tracking individual facial features within each detected facial blob, we utilize a hybrid approach by integrating an appearance-based detector and a feature-based tracker for the eyes, the nose and the mouth. The combined approach inherits advantages from both approaches permitting robust identification of the facial features, correct maintenance of feature IDs among frames, as well as real-time computations.

The overview of the implemented approach is illustrated in Fig. 6 and is based on three steps.

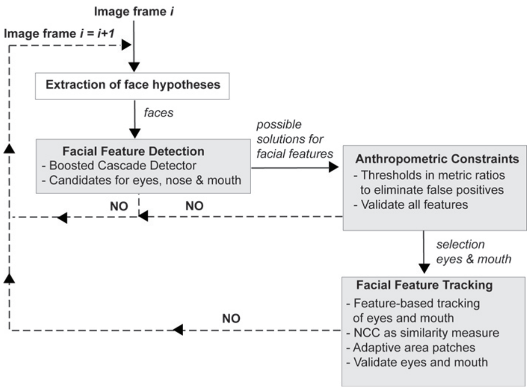
Image frame *i*
Image frame *i = i+1*
Extraction of face hypotheses
faces
Facial Feature Detection
- Boosted Cascade Detector
- Candidates for eyes, nose & mouth
possible solutions for facial features
Anthropometric Constraints
- Thresholds in metric ratios to eliminate false positives
- Validate all features
NO
NO
selection eyes & mouth
Facial Feature Tracking
- Feature-based tracking of eyes and mouth
- NCC as similarity measure
- Adaptive area patches
- Validate eyes and mouth
NO

Fig. 6. Diagram of the proposed approach for detection and tracking of facial features.

For the initial detection of facial features we use the Boosted Cascade Detector of Viola and Jones [11] and the available implementation in the OpenCV open source library. In our case, for the detection of the features within each face blob, individual sets of Haar-like features for eyes, nose and mouth are utilized and the method is initialized with frontal-view faces.

An important factor which affects both the reliability of detection and the tracking accuracy of facial features is the size of the detected face blob. According to Tian [12], facial features become hard to detect when the face region is smaller than a threshold of approximately 70 × 90 pixels. Therefore, the procedure of facial feature detection and tracking is only activated when the face blob satisfies the above size requirements.

After all features have been detected, specific anthropometric constraints are applied in order to cast out false positives. Motivated by the work of Sohail and Bhattacharya [13], we have collected a large set of measurements from images depicting faces in frontal view. The collected measurements were used to built an anthropometric model of the human face and to define the necessary thresholds and validation gates used to filter out false positive detections. The selected validation criteria involve the location and the size of the eyes, the nose and the mouth. Landmarks on other regions such as the eyebrows, used, for example, in [13], were not selected because they often proved to be occluded by hair, eyeglasses or, in some cases, they were entirely non-existent.

More specifically, we define the following criteria:

- All four selected features (eyes, nose, mouth) should be detected.
- The normalized sizes of the two eyes and mouth should be within certain bounds.
- The normalized distance between the midpoint of the eye centers and nose tip should be approximately 0.6. That is $D_2/D_1 \simeq 0.6$, where $D2$ is the distance between points $P_3$ and $P_4$ (see Fig. 7).
- The normalized distance between the midpoint of eye centers and mouth center should be approximately $\simeq$ 1.2. That is $D_3/D_1 \simeq 1.2$, where $D_3$ is the distance between points ($P_3$ and $P_5$).

The above defined criteria are applied for each facial blob, following the detection of facial features.

For blobs that pass the above criteria, the tracking procedure is invoked. It is to be noted that tracking is only performed for the two eyes and for the mouth region. The nose region is not tracked because it's actual location is not considered important for our target application, which is expression recognition and visual speech detection.

Tracking is based on template matching, using as eye and mouth templates the detected areas from each face using the normalized cross-correlation (NCC) measure as matching score/quality measure. The selection of NCC as quality measure is justified as only small deviations in the relative positions of the feature areas with respect to the position of the face blob in the image are expected. The position with the maximum similarity score within each search area for eyes and mouth is selected as the new feature

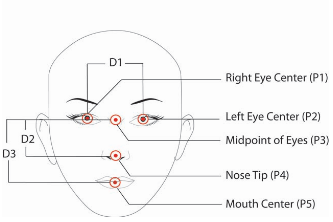

Fig. 7. Landmarks in the Anthropometric Face Model.

position and the size of templates is updated with a factor in every consecutive frame. The width and height factor are computed by the ratios of the template width and height to the respective face width and height. The matching score is used to block results of low reliability and if it is below a certain threshold, detection is reinvoked. With this approach there is a significant gain in processing time, allowing for real-time computations. For example for the case of images with size 640x480 pixels and a face of approximate size of 200 x 200 pixels, detection of each feature area is approximately at 200 - 400 ms whereas tracking is below 10 ms using a standard computer.

## VI. EXPERIMENTAL RESULTS WITH REAL WORLD DATA

Figure 8 depicts results from a sequence with multiple persons. The number of persons varies with time and there are many occlusions which cause frequent disappearances/reappearances of both hand and face blobs. Despite these difficulties, the tracking algorithm succeeds to correctly identify all visible faces and hands. In Fig. 9 results are shown from the same sequence with emphasis on facial feature localization within the provided face blobs, in cases of multiple persons performing different facial expressions and head gestures.

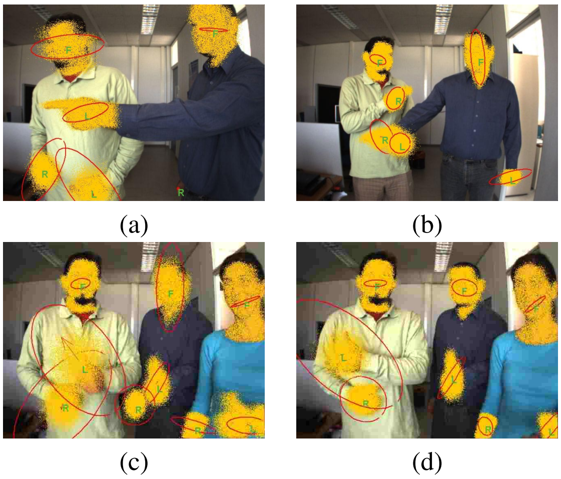
(a) (b) (c) (d)

Fig. 8. Four frames of a sequence depicting multiple persons performing hand and facial gestures in an office environment. Throughout the sequence performers enter and leave the scene and there are frequent occlusions, which result in a varying number of hand and face hypotheses.

Figure 10 presents hand and face classification results for various frames of the office sequence of Fig. 2. Blobs classified as faces are marked with an “F”, left hands are marked with an “L”, and right hands are marked with an “R”. The proposed approach has been successful in classifying the three observed tracks and it also managed to maintain its belief over the whole sequence. To quantitatively assess the tracker’s performance with respect to class discrimination and robust tracking over time, Fig. 11 depicts the belief of each of the three tracks of the office sequence, as it evolves over the first 500 frames of this sequence. As can be easily observed, the belief of each track is initially uncertain but very soon it stabilizes to the correct class. The belief stays stable to the correct classes thoughout the whole sequence consisting of a total of 2600 frames (for clarity of presentation, only the first 500 frames are shown in these graphs).

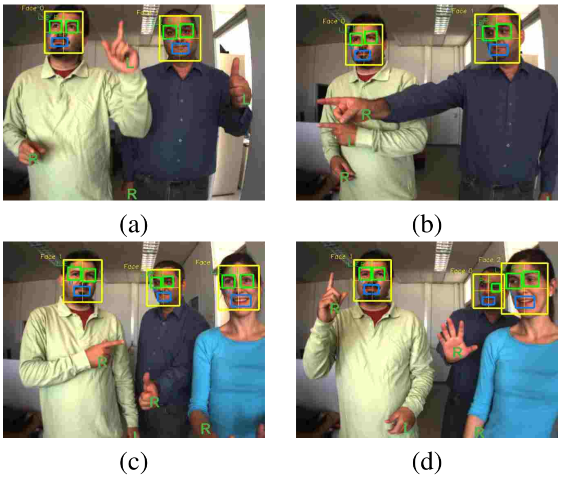
(a) (b) (c) (d)

Fig. 9. Four frames of a sequence depicting multiple persons performing hand and facial gestures in an office environment, focusing on the tracking of facial features. Throughout the sequence performers enter and leave the scene, and perform gestures in various distances from the camera.

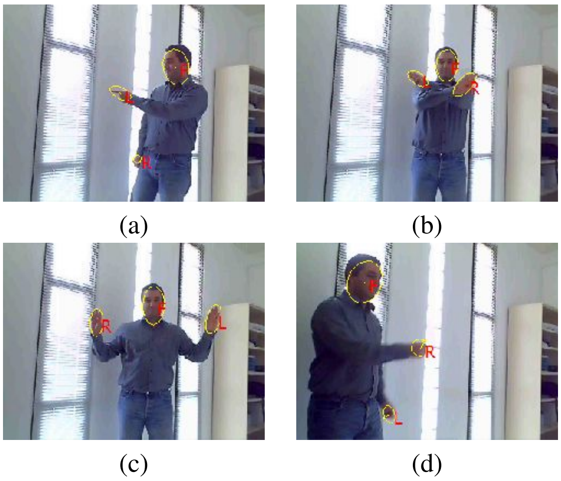
(a) (b) (c) (d)

Fig. 10. Four frames of a sequence depicting a person performing various hand gestures in an office environment.

Figure 12 depicts some facial feature tracking results from two additional, close-up, sequences captured in an exhibition center. The first sequence comprised of a total number of 1100 image frames, whereas the second sequence of 650 image frames. In all our experiments the algorithm successfully tracked the skin-colored blobs corresponding to faces, following convergence curves which were very similar to the ones depicted in Fig. 11. Eyes, nose and mouth regions were also correctly localized and tracked, even in cases of usual off-plane head rotations and different facial expressions. Table I shows quantitative results, verified by a human supervisor, of the two sequences of Fig. 12. The high true positive (TP) percentages for each facial region indicate successful localization when the respective region is visible.

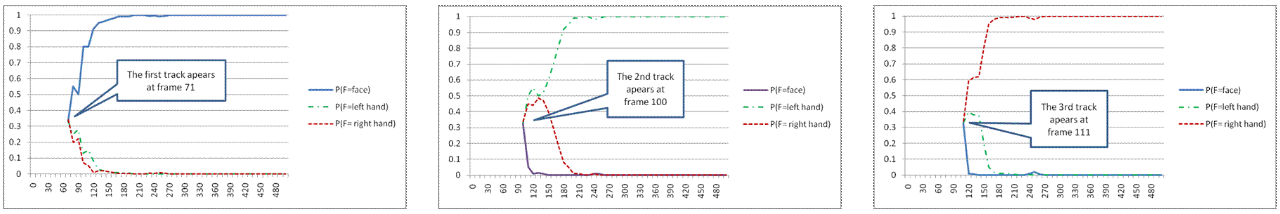

Fig. 11. The belief of each of the three tracks of the office sequence, as it evolves over the first 500 frames. The solid blue lines correspond to the probability of each blob being a face blob, the dot-dashed green lines correspond to left hands and the dashed red lines corresponds to right hands.

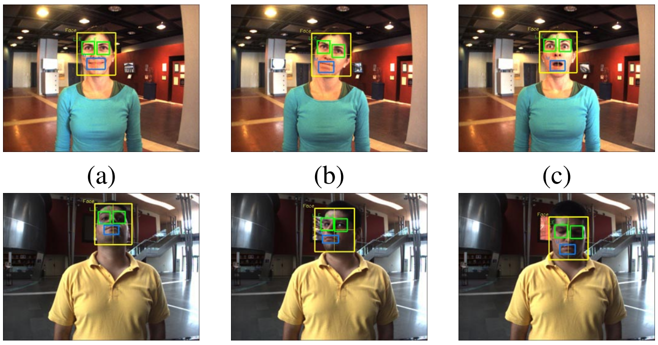
(a) (b) (c)
(d) (e) (f)

Fig. 12. Frames of different sequences captured in an exhibition center that show results from facial feature tracking of a person.

The false negative (FN) percentages are generally low with the exception for the right eye of sequence 2 due to lower signal content in this area.

TABLE I

PERCENTAGES OF TRUE POSITIVE (TP), TRUE NEGATIVE (TN), FALSE POSSITIVE (FP) AND FALSE NEGATIVE (FN) RESULTS FOR THE TWO SEQUENCES OF FIG. 12.

| | **TP** (%) | **TN** (%) | **FP** (%) | **FN** (%) |
|---|---|---|---|---|
| Sequence 1 | | | | |
| Mouth | 95.09 | 0 | 0.22 | 4.69 |
| Left Eye | 95.98 | 0 | 0 | 4.02 |
| Right Eye | 93.30 | 2 | 0.22 | 4.64 |
| Sequence 2 | | | | |
| Mouth | 93.31 | 0 | 1.18 | 5.51 |
| Left Eye | 94.49 | 4 | 0 | 3.94 |
| Right Eye | 85.83 | 6 | 0 | 11.81 |

More experimental results from different application environments (office, exhibition center) can be found in `http://www.ics.forth.gr/~pateraki/handfacetracking.html`.

## VII. CONCLUSIONS

In this paper we have presented an integrated approach for tracking of hands, faces and facial features of multiple persons in image sequences, intended to support natural interaction with autonomously navigating robots in public spaces and, more specifically, to provide input for the analysis of hand gestures and facial expressions that humans utilize while engaged in various conversational states with the robot.

For hand and face tracking, the skin-color tracker was extended by incorporating an incremental probabilistic classifier which was used to maintain and continuously update the belief about the class of each tracked blob which can be a left-hand, a right hand or a face. Facial feature detection and tracking was performed via the employment of state-of-the-art appearance-based detection coupled with feature-based tracking, using a set of anthropometric constraints.

Experimental results have confirmed the effectiveness of the proposed approach proving that the individual advantages of all involved components are maintained, leading to implementations that combine accuracy, efficiency and robustness.

Besides using the proposed methodology to give input for the analysis of hand gestures and facial expressions, it possesses characteristics that constitute it suitable for more general activity recognition tasks and tasks related to robot learning by demonstration. Future work will involve relevant investigations in the mentioned areas.

## REFERENCES

[1] M.H. Yang, D. Kriegman, and D. Ahuja. Detecting faces in images: A survey. *IEEE Trans. PAMI*, 24(1):34–58, 2002.

[2] H. Baltzakis, A. Argyros, M. Lourakis, and P. Trahanias. Tracking of human hands and faces through probabilistic fusion of multiple visual cues. In *Proc. International Conference on Computer Vision Systems (ICVS)*, pages 33–42, Santorini, Greece, May 2008.

[3] B. Stenger, A. Thayananthan, P. H. S. Torr, and R. Cipolla. Model-based hand tracking using a hierarchical bayesian filter. *IEEE Trans. PAMI*, 28(9):1372–1384,, September 2006.

[4] T. Nguyen, V. Nguyen, and H. Kim. Robust feature extraction for facial image quality assessment. In Y. Chung and M. Yung, editors, *Information Security Applications*, volume 6513 of *Lecture Notes in Computer Science*, pages 292–306. Springer Berlin / Heidelberg, 2011.

[5] A.L. Yuille, P.W. Hallinan, and D.S. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99–111, 1992.

[6] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[7] Y. Zhou, Y. Li, Z. Wu, and M. Ge. Robust facial feature point extraction in color images. *Engineering Applications of Artificial Intelligence*, 24:195–200, 2011.

[8] S. Phimoltares, C. Lursinsap, and K. Chamnongthai. Locating essential facial features using neural visual model. In *Proc. Intl. Conf. on Machine Learning and Cybernetics, 2002*, volume 4, pages 1914–1919, Nov. 2002.

[9] H. Baltzakis and A. Argyros. Propagation of pixel hypotheses for multiple objects tracking. In *Proc. International Symposium on Visual Computating (ISVC)*, Las Vegas, Nevada, USA, November 2009.

[10] Chris Stauffer and W. Eric L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 2246–2252, June 1999.

[11] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[12] Y. Tian. Evaluation of face resolution for expression analysis. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, pages 82–89. IEEE Computer Society, 2004.

[13] A.S. Md Sohail and P. Bhattacharya. Detection of facial feature points using anthropometric face model. In Ernesto Damiani, Kokou Ytongnon, Peter Schelkens, Albert Dipanda, Louis Legrand, and Richard Chbeir, editors, *Signal Processing for Image Enhancement and Multimedia Processing*, volume 31 of *Multimedia Systems and Applications*, pages 189–200. Springer US, 2008.