LIP TRACKING USING ADAPTIVE FUZZY PARTICLE FILTER IN THE CONTEXT OF CAR DRIVING SIMULATOR UNDER LOW CONTRAST NEAR-INFRARED ILLUMINATION

Parisa Darvish Zadeh Varcheie, Langis Gagnon

R & D Department, CRIM, 405 Ogilvy Avenue, Suite 101, Montreal (QC), Canada, H3N 1M3
Email: { parisa.darvish, langis.gagnon} @crim.ca

ABSTRACT

A real-time lip tracking on very low contrast images acquired under near-infrared illumination is presented. We developed a modified particle filter tracker based on fuzzy logic that is appropriate for non-linear modeling and robust to the non-Gaussian noise. Fuzzy model is used to normalize the particle filter samples weights. Fuzzy membership functions are applied to geometric and appearance features. Lip modeling and tracking are done by sampling around lip regions using a particle filter and scoring sample features are done based on a fuzzy rule. The performance of the tracking algorithm is evaluated for different people with various mouth changes, such as smile and speech. More than 78% of the lip corners are correctly detected within distances less than 5% of the lip length from the ground truth.

Index Terms— lip tracking, particle filter, fuzzy modeling

1. INTRODUCTION

Lip detection and tracking have many applications in pattern recognition, such as speech and facial expression recognition. This paper presents a real-time lip tracking method designed for car driving simulator within the SPEED-Q project [1]. The goal of SPEED-Q is to develop a simulation environment with a multi-sensor data acquisition and analysis system for driving performance assessment, cognitive load measure and training. The subjects are asked to drive in a simulator and then react to the monitored scenarios (Fig. 1). Their cognitive load will vary according to the complexity level of the driving task. Facial expressions (FE) are important cognitive load markers in the context of car driving. FE can be defined globally or locally using the Facial Action Coding System (FACS) [2]. FACS is based on muscular activity underlying momentary changes in facial aspects where each change is coded as a facial Action Unit (AU). A certified FACS coder has manually analyzed 74 video sequences of 12 people, acquired in our driving simulator, in order to identify the set of most important AUs depicted by car drivers. A total of 140 instances of AUs were identified, for which the most common were eye blinks, brow lowerer, jaw drops, lips apart, lip cor-

ner puller and lips suck. We have implemented a real-time eye blink detector that has been integrated in the car simulator [3]. We present here our progresses regarding real-time lip points tracking on the particular facial images of the simulator, that is, very low contrast frontal face images (i.e. since normally the driver during driving looks at the front of the car) acquired under near-infrared illumination (Fig.2). Lip points tracking is the basic step for the detection of all mouth-related AUs. Among the facial features, mouth and lips are difficult to track since they are highly deformable. Geometric and appearance based features are two general methods used in face recognition [4]. Using both geometric and appearance features might be the best choice for certain FE [5]. Several works have been done in lip tracking, but most of them are pixel color based ([6, 7]) which are not applicable to the near-infrared illumination. Some others extract and track the whole lip contour and mouth changes [8, 9], but time processing of contour-based methods is high. Also, in our application, we need to know the exact lip corners position and their distances during the tracking. Here, lip modeling and tracking are done based on sampling around lip regions using a particle filter and scoring sample features based on a fuzzy rule. We combine fuzzy logic approach which is appropriate for non-linear model, with particle filter which is robust to non-Gaussian noise. The performance of the tracking algorithm for different people with various mouth changes, such as smile and talk, is evaluated. Section 2 presents our proposed lip tracking algorithm, and Section 3 gives performance results. Section 4 concludes the paper.

2. METHODOLOGY

An important refinement in the context of the SPEED-Q project is that the tracking algorithm must not lose the lip throughout the video and must not be distracted by head movement. The particle filter is a Bayesian method that recursively estimates the state of the tracking target as a posterior distribution with a finite set of weighted samples. It operates in two prediction and update phases. Particle samples are generated during the prediction phase and are related to the state of the tracked target.

In our problem, the lip is modeled by four lip anchor

points which are the left, right, up (centre point on outer edge of upper lip according to the Face and Gesture Recognition Working group (FGNET) [10] annotation), and down (centre point on outer edge of lower lip) lip fiducial points (called corners in the rest of this paper) around mouth region (see Fig. 1). The state of the particle filter at each time t and for each lip corner is defined as a vector P_t of the coordinates of each four lip corners:

$$P_t = (x(t), y(t)) \tag{1}$$

The state vector P_t is determined manually in the initial step of the tracking algorithm. The following geometric and appearance-based features with their measures are used as the observation model for our particle filter tracking method:

1. $\phi_p(s)$, Euclidean distance between the sample coordinates and the correspondent previous position of the lip corner, is the first measure on a geometric feature. This distance is the input of a normalized Gaussian membership function $\bar{\phi}_p(s)$ given by:

$$\bar{\phi}_p(s) = e^{\left(-\frac{\phi_p(s)}{2\sigma^2}\right)}. \tag{2}$$

where σ has been determined experimentally equal to 10 (for image size 640×480).

2. $\phi_t(s)$, Euclidean distance between the sample coordinates and the detected correspondent corner coordinates obtained of longest trace of the current edge lip image, is the second measure on a geometric feature. $\phi_t(s)$ is imported to a normalized Gaussian membership function, $\bar{\phi}_t(s)$, which is the same as $\bar{\phi}_p(s)$.

3. $\phi_g(s)$, 2-D correlation coefficient between cropped $\alpha_x \times \alpha_y$ (30×30) gray level image around each sample coordinates, and the same size gray level template around correspondent previous position of the lip corner, is used as the measure for this appearance feature. $\bar{\phi}_g(s)$ is the considered fuzzy membership function and applied to $\phi_g(s)$ as :

$$\bar{\phi}_g(s) = \frac{\phi_g(s) + 1}{2} \tag{3}$$

4. $\phi_h(s)$, Euclidean distance between the normalized gray level histogram of the cropped $\alpha_x \times \alpha_y$ gray level image around each sample coordinates, H_c and the normalized gray-level histogram of the same size region around previous position of the lip corner H_L is the measure for this appearance feature as:

$$\phi_h(s) = \sqrt{\sum_n (H_c(s)[n] - H_L[n])^2} \tag{4}$$

where n is the histogram bin number. The membership function $\bar{\phi}_h(s)$ is considered as:

$$\bar{\phi}_h(s) = 1 - \frac{\phi_h(s)}{\sqrt{2}} \tag{5}$$

5. $\phi_e(s)$, 2-D correlation coefficient between cropped $\alpha_x \times \alpha_y$ edge images around each sample coordinates, and the same size cropped edge images around correspondent previous position of the lip corner is used as the measure for this appearance feature. $\bar{\phi}_e(s)$ is the considered fuzzy membership function the same as $\bar{\phi}_g(s)$ and applied to $\phi_e(s)$.

6. $\phi_{px}(s)$ and $\phi_{py}(s)$, Euclidean distances between the normalized x and y projection histogram pattern, and the cropped $\alpha_x \times \alpha_y$ edge image around each sample coordinates, with the normalized x and y projection histograms of the same size region around previous position of the lip corner are the measures used for x and y histogram projections respectively the same as $\phi_h(s)$. Their correspondent membership functions,$\bar{\phi}_{px}(s)$ and $\bar{\phi}_{py}(s)$ are similar to $\bar{\phi}_h(s)$.

7. Classification results of a GentleAdaBoost classifier [11] is the last appearance and robust feature. Four GentleBoost classifiers with ((48 ($images\ responded\ to\ Gabor\ filters$) $+ 1$ ($grayscale\ image$)) $\times 13 \times 13$ ($pixels\ image$) =) 8,281 Gabor features are trained for each lip corner. Confidence level $\phi_c(s)$ [11] of the classifier is used as a measure and the sigmoid function over this measure is applied as fuzzy membership function given by:

$$\bar{\phi}_c(s) = \frac{1}{e^{-\alpha(\phi_c(s)-\beta)} + 1} \tag{6}$$

At each time t, samples are determined by replacement from the previous sample set S_{t-1} at time $t-1$,

$$S_{t-1} = \{x_i^{t-1}, w_i^{t-1}\}_{i=1}^N \tag{7}$$

where N is the number of samples, x_i^{t-1} is the ith sample coordinates at time $t-1$ and w_i^{t-1} is the related weight to i^{th} sample . Sample set S_{t-1} is an approximation of posterior distribution of the target state at time $t-1$. The weight of the sample w_i^{t-1} is obtained by adding all membership values derived from discussed membership functions:

$$w_i^{t-1} = \bar{\phi}_p(s_i^{t-1}) + \bar{\phi}_t(s_i^{t-1}) + \bar{\phi}_g(s_i^{t-1}) + \bar{\phi}_h(s_i^{t-1}) +$$
$$\bar{\phi}_e(s_i^{t-1}) + \bar{\phi}_{px}(s_i^{t-1}) + \bar{\phi}_{py}(s_i^{t-1}) + \bar{\phi}_c(s_i^{t-1}) \tag{8}$$

Among all N samples, N_{s_i} samples with high probabilities (weights) are selected and the samples with small probabilities may be never selected. The particle filter state in two consecutive frames does not change significantly. It is typically a translation of sample coordinates around its previous position. In addition, for each frame we insert strong corners coordinates determined by Harris corner detector to candidate samples coordinates. Indeed we use an additional source of samples which is obtained by applying Harris corner detector
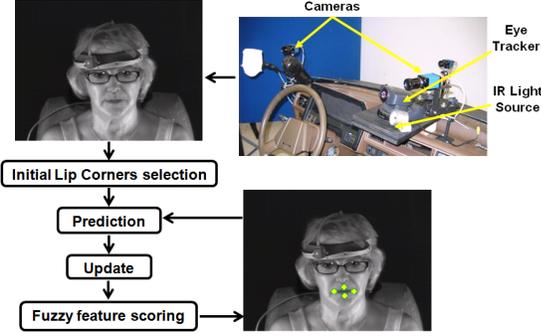
Fig. 1. Car driving simulator (top right) and lip tracking algorithm (left)

Fig. 2. Circles with radius of 5%, 10%, 20% and 50% of normal lip length around the left and right lip corners

around a $\beta_x \times \beta_y$ (e.g. 50×50) image pixel size of previous position of the correspondent lip corner. At each time t, samples are reproduced in state space by a dynamical first order auto-regressive model given by:

$$P_t = P_{t-1} + w_t \qquad (9)$$

P_t and P_{t-1} are the particle filter states at time t and $t-1$ respectively. w_t is a multivariate Gaussian random variable and it correlates to random translation of the sample corner coordinates. Each lip corner s_p is the best sample in each time t which has the maximum weights and selected by:

$$s_p = argmax_{si \in S}\{\bar{\phi}_p(s_i) + \bar{\phi}_t(s_i) + \bar{\phi}_g(s_i) + \bar{\phi}_h(s_i) +$$
$$\bar{\phi}_e(s_i) + \bar{\phi}_{px}(s_i) + \bar{\phi}_{py}(s_i) + \bar{\phi}_c(s_i)\} \quad (10)$$

The lip tracking algorithm using this adaptive fuzzy particle filter (Fig. 1) can be summerized as the following:

Step 1. Initialization
1- Manually select four up, down, left and right lip corners.
Step 2. Prediction
1- For each lip corner, apply Harris corner detector to a region around previous position of the correspondent lip corner.
2- Insert the corners coordinates resulting from Harris corner detector to the available sample coordinates from previous frame.
3- Choose samples from samples at time $t-1$, S_{t-1}, based on the probabilities determined by equation (8).
4- Reproduce samples utilizing equation (9).
Step 3. Update
1- Update new samples weights from obtained observation measurement using equation (8).
Step 4. Fuzzy feature scoring
1- Select the best sample using equation (10).
2- Increase time t and iterate from step 2 to step 4.

3. EXPERIMENTAL RESULTS AND ANALYSIS

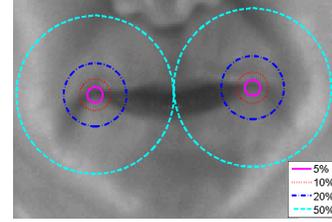The test dataset is composed of uncompressed video sequences of 18 subjects sitting in the driving simulator (one sequence per subject, with an average of 20 minutes videos at 30 fps). We evaluate the tracking algorithm on those video parts where the lip and mouth movements from the frontal view are present (e.g. smile, talking, lip stretch, lip press, etc.). In Fig. 2, some circles with the radius of 5%, 10%, 20% and 50% of normal lip length around the left and right lip corners are shown. TPR and TF in Table 1 and Table 2 are the two metrics used to evaluate our method . TPR calculates the target localization accuracy according to the number of frames with lip corner located correctly (TP) and wrongly (FP). TF indicates the lack of continuity of the tracking system for a single lip corner track [12] versus the number of frames where the true lip corner is not detected (T_m) and total number of frames (NF). Mouth changes related to each person is described below the Table 1. The TPR and $\overline{TPR}$ values for all four corners are enhanced when the circle radius increased. More than 78% of the detected lip corners have distances less than 5% of the lip length from the ground truth. The TPR and $\overline{TPR}$ results of the left and right lip corners are higher than TPR and $\overline{TPR}$ results of up and down lip corners. In addition, these values for the up lip corner are more elevated than down lip corner. Furthermore, there is less tracking fragmentation for left and right lip corners than the up and down lip corners. Indeed left, right and up lip corners are stronger than down lip corners. The texture of left and right lip corners are more significant than up and down lip corners. Typically, left and right lip corners should have almost the same TPR, TF, $\overline{TPR}$, and $\overline{TF}$ values because of symmetric constraint. Here, small variation in the video sequences is not uniform in both left and right sides of the driver and the right side is brighter than the left side. This asymmetrical lighting has serious effect on the appearance based features of the observation model. We have used histogram equalization to reduce this effect, but there is still a small difference between left corner and right corner tracking results.

4. CONCLUSION

A real-time fuzzy particle filter lip tracking on very low contrast images was presented. We combine fuzzy logic approach with particle filter to have robustness to non-Gaussian noise

Table 1. TPR of left, right, up and down lip corners tracking.

		TPR%=TP/(TP+FP)%							
rad	LC	F_1	F_2	F_3	F_4	F_5	F_7	F_7	$\overline{TPR}$%
5%	l	89	95	71	87	59	94	92	**83**
	r	91	95	68	94	92	93	93	**89**
	u	82	90	66	84	77	87	86	**81**
	d	73	87	69	80	72	85	84	**78**
10%	l	94	98	74	93	76	95	94	**89**
	r	94	98	73	96	94	96	94	**92**
	u	85	92	70	85	82	90	87	**84**
	d	76	89	72	83	75	89	86	**81**
20%	l	97	99	80	93	92	97	96	**93**
	r	99	99	81	100	97	98	97	**95**
	u	89	94	79	87	84	92	90	**87**
	d	91	94	81	84	78	91	89	**86**
50%	l	99	100	93	100	100	100	99	**98**
	r	100	100	91	100	100	100	99	**98**
	u	97	98	89	92	88	95	93	**93**
	d	96	97	88	91	83	93	92	**91**

rad:radius circle size, LC: lip corner, l: left lip corner, r: right lip corner, u: up lip corner, d: down lip corner, F_1: talk, F_2: smile, F_3: lip press, F_4: lip suck, F_5: lip pucker, F_6: lip funneler, F_7: lip tightener, $\overline{TPR}$: average of TPR values.

Table 2. TF of left, right, up and down lip corners tracking. $\overline{TF}$: average of TF values.

	TF%= T_m/NF%							
LC	F_1	F_2	F_3	F_4	F_5	F_6	F_7	$\overline{TF}$%
l	0.1	0.3	0.2	0.5	0.4	0.7	0.6	**0.4**
r	0.05	0.2	0.1	0.4	0.5	0.6	0.8	**0.37**
u	1.4	2.3	2.6	2.9	3	3.1	3.4	**2.67**
d	1.7	2.5	2.9	3.1	3.6	3.2	3.3	**2.9**

and non-linear tracking model. Fuzzy model is applied to obtain the sample weight. Lip modeling and tracking are done based on sampling around lip regions using particle filter and scoring sample features based on a fuzzy rule. The performance of the tracking algorithm for different people with various mouth changes was evaluated. Future work will be to add a SVM classifier to recognize the correspondant AU due to the lip changes for detection of unsafe driver behaviour.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Auto21, "Safe platform for evaluating/enhancing driver qualifications," http://www.auto21.ca/en/subcontent.php?page=ae2105, [Online; accessed 15-December-2009].

[2] P. Ekman and W.V. Friesen, "Facial action coding system: : A technique for the measurement of facial movement," *Consulting Psychologists Press, Palo Alto*, 1978.

[3] M. Lalonde, D. Byrns, L. Gagnon, N. Teasdale, and D. Laurendeau, "Real-time eye blink detection with GPU-based SIFT tracking," *Canadian Conference on Computer and Robot Vision (CRV)*, pp. 481–487, 2007.

[4] K. Delac and M. Grgic, "Face recognition," *I-Tech Education and Publishing, Austria*, 2007.

[5] S. Beauchemin, P. D. Z. Varcheie, L. Gagnon, D. Laurendeau, M. Lavalliere, T. Moszkowicz, F. Prel, and N. Teasdale, "COBVIS-D: A computer vision system for describing the cephalo-ocular behavior of drivers in a driving simulator," *International Conference on Image Analysis and Recognition (ICIAR)*, pp. 604–615, 2009.

[6] R. Stiefelhagen, U. Meier, and J. Yang, "Real-time lip-tracking for lipreading," *In Eurospeech 9*, 1997.

[7] B. Abboud and G. Chollet, "Appearance based lip tracking and cloning on speaking faces," *IEEE International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 301–305, 2005.

[8] S. Lucey, S. Sridharan, and V. Chandran, "Initialized eigenlip estimator for fast lip tracking using linear regression," *IEEE International Conference on Pattern Recognition (ICPR)*, pp. 3182–3185, 2000.

[9] M. Lievin, P. Delmas, P.Y. Coulon, F. Luthon, and V. Fristot, "Automatic lip tracking: Bayesian segmentation and active contours in a cooperative scheme," *IEEE International Conference on Multimedia Computing and Systems*, vol. 1, pp. 9691–9697, 1999.

[10] FGNET, "Face and gesture recognition working group," http://www-prima.inrialpes.fr/FGnet/html/home.html, [Online; accessed 13-September-2009].

[11] P. Darvish Zadeh Varcheie and L. Gagnon, "Progress report of CRIM's activities for the SPEED-Q project for the period of April 2008 to March 2009," *Technical Report, CRIM-09/04-2, Montreal, CRIM*, 2009.

[12] F. Yin, D. Makris, and S.A. Velastin, "Performance evaluation of object tracking algorithms," 2007, IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (PETS).