

# Human Lips-Contour Recognition and Tracing

Md. Hasan Tareque<sup>1</sup>  
Department of Computer Science and Engineering  
IBAIS University  
Dhaka, Bangladesh

Ahmed Shoeb Al Hasan<sup>2</sup>  
Department of Computer Science and Engineering  
Bangladesh University of Business & Technology  
Dhaka, Bangladesh

**Abstract**—Human-lip detection is an important criterion for many automated modern system in present day. Like computerized speech reading, face recognition etc. system can work more precisely if human-lip can detect accurately. There are many processes for detecting human-lip. In this paper an approach is developed so that the region of a human-lip can be detected, we called it lip contour. For this a region-based Active Contour Model (ACM) is introduced with watershed segmentation. In this model we used global energy terms instead of local energy terms because, global energy gives better convergence rate for malicious environment. At the time of ACM initialization by using  $H^\infty$  based on Lyapunov stability theory, the system gives more accurate and stable result.

**Keywords**—Watershed Model; Active contour models (ACM);  $H^\infty$  filter Contour model; Lyapunov stability theory

## I. INTRODUCTION

Automatic lips contour detection and tracking is always a crucial prerequisite process for various kinds of applications. It has been extensively utilized in the state-of-the art of audio-visual speech recognition. [1]

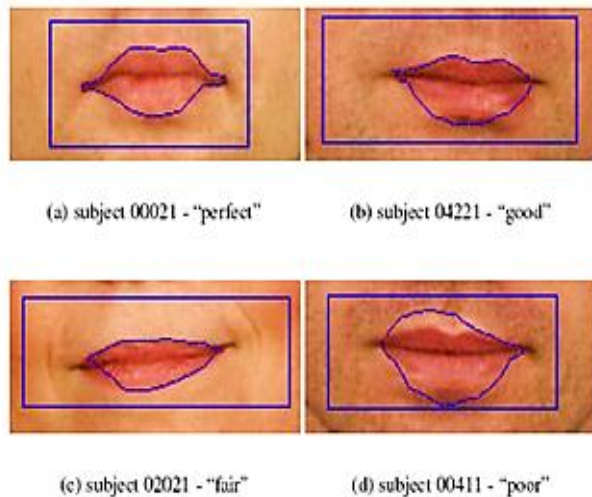


Fig. 1. Different shape of lips contour

There are several uses of lips contour detection like Audio-visual speech authentication [3], intelligent human-computer interaction Human expression recognition [2] etc. Automatic Speech Recognition (ASR) [4] systems use only Acoustic Information, for this system show poor performance in noisy environment. Bimodal Audio-Visual Systems signal often contains information that is complementary to audio

information. Again visual information is not affected by acoustic noise. For this overall performance of the combined system is better [5]-[7].

Moreover, when working with the video sequence, an automatic and precise initial contour placement for the ACM is a crucial issue. Therefore, lips tracking algorithm is adopted into the system to keep track on the lips feature points at the subsequent incoming video frame as the initialization of the ACM lips contour process. In this paper, the conventional  $H^\infty$  filtering is modified according to the LST as to assure that the system is always at the stable condition. Additionally, by properly selecting the Lyapunov function  $V(k)$ , during the system design, as the time approaches to infinity, the tracking error of the proposed  $H^\infty$  filtering would asymptotically converge to zero. This is because the Lyapunov function [11] selected for the proposed system,

$$V(k) = e^2(k)$$

This consists of a unique global minimum point at the system origin. From the simulation results demonstrated, the modified  $H^\infty$  shows an appreciable improvement in terms of reducing the lips feature points tracking error compared to the conventional  $H^\infty$  approach

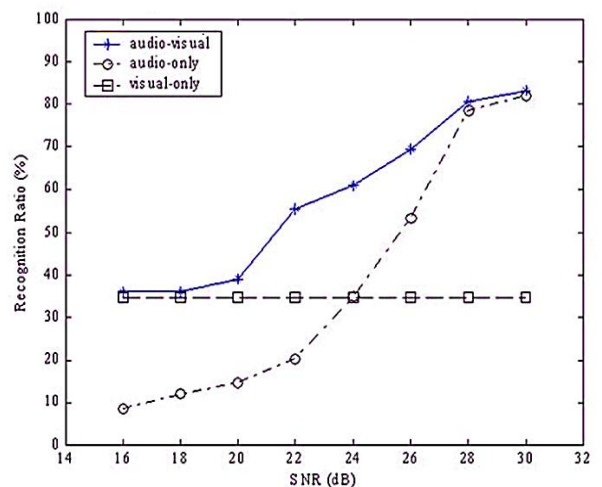


Fig. 2. SNR ratio of different system

In this paper some of the terminology will be discussed like

### A. Contour:

A contour line of a function of two variables is a curve along which the function has a constant value. Minimal Cost

Path: The optimal path from every pixel in the image to the seed point is determined by using Dijkstra's algorithm.[9]

B. Active contour model:

Active contour model, also called snakes, is a framework for delineating an object outline from a possibly noisy 2D image. This framework attempts to minimize an energy associated to the current contour as a sum of an internal and external energy.[12]

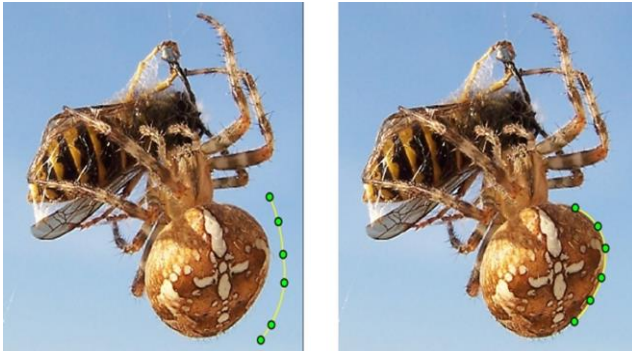


Fig. 3. Active contour model.

C.  $H^\infty$  filter:

After a decade or so of reappraising the nature and role of Kalman filters, engineers realized they needed a new filter that could handle system modelling errors and noise uncertainty. State estimators that can tolerate such uncertainty are called robust. The  $H^\infty$  filter [8] was designed for robustness.

D. Lyapunov stability theory:

The most important type is that concerning the stability of solutions near to a point of equilibrium. If all solutions of the dynamical system that start out near an equilibrium point  $x_e$  stay near  $x_e$  forever, then  $x_e$  is Lyapunov stable. [12]

II. LIPS CONTOUR DETECTION & TRACKING

In this paper, instead of utilizing the global region model, the active contour energy with the local information interpretation corresponds to the watershed segmentation is proposed. This gives better performance on the contour detection compared to the global energies calculation, particularly when the inner and outer region shares the similar image statistic. Performs is better compared to the localized ACM algorithm.

Compared to the aforementioned approach, the integration of ACM and watershed segmentation presented in this paper not only provides high deformability but is also better in handling the unclear boundary between lips and facial skin.

For video sequence, an automatic and precise initial contour placement for the ACM is a crucial issue. Keep track on the lips feature points at the subsequent incoming video frame as the initialization of the ACM [6] lips contour process.

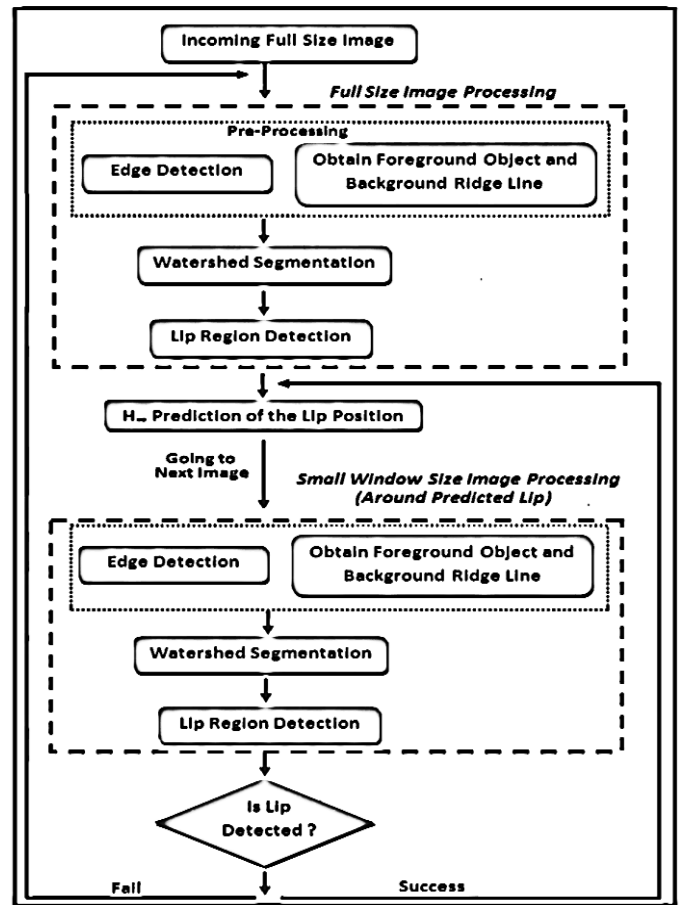


Fig. 4. Follow chart of the system.

$H^\infty$  filtering is modified according to the LST as to assure that the system is always at the stable condition. By properly selecting the Lyapunov function,  $V(k)$ , as the time approaches to infinity, the tracking error of the proposed  $H^\infty$  filtering would asymptotically converge to zero. This is because the Lyapunov function selected for the proposed system,

$$V(k) = e^{2k}$$

The overall process of the proposed lips contour detection and tracking is demonstrated in Figure 5. The input image is first sent to the lips localization to roughly obtain the lips location. Subsequently, the watershed region segmentation and at the same time, also the ACM initialization are working on the lips image. The output from the watershed and the ACM initial contour position are then passed to the localized region-based ACM for further lips contour detection. Once the lips contour is successfully detected, the feature points from the detected contour are extracted and used by the modified  $H^\infty$  lips tracking process to keep track on the initial contour position of the succeeding image. The lips localization is triggered only once per 20 rounds of the overall lips detection process.

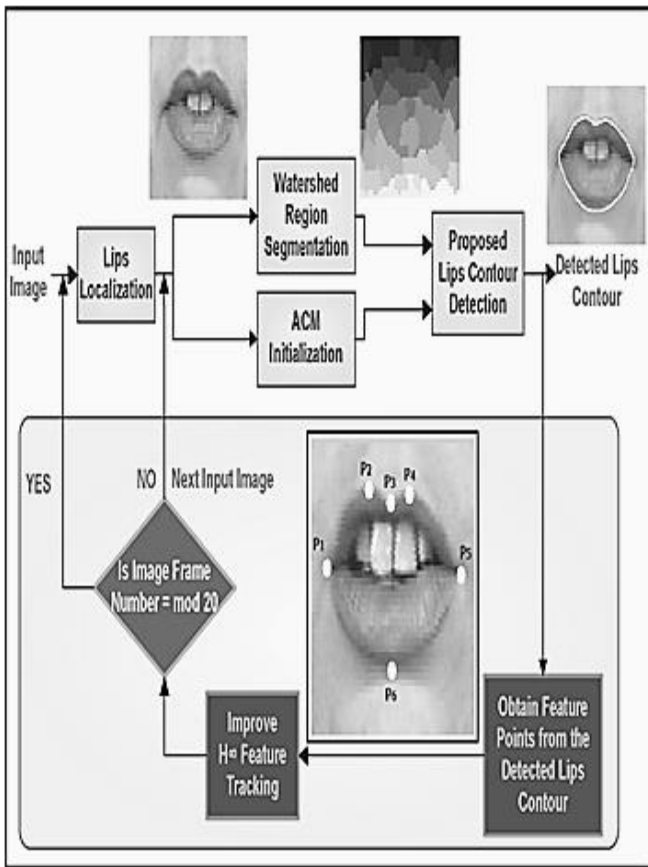


Fig. 5. Overview of the overall system.

### III. LIPS CONTOUR DETECTION WITH LOCALIZED WATERSHED-BASED ACTIVE CONTOUR MODEL

Each pixel is situated at a certain altitude levels. Where the white pixel (intensity value = 255) is referred to as the maximum altitude while black pixel (intensity value = 0) is known as the minimum altitude.

The watershed algorithm which applied in this paper is based on the rain-flow simulation. By employing the falling rain concept, the raindrop falls from the highest altitude to the lowest (known as catchment basin) according to the steepest descent order.

The coloured watershed regions are known as the local mask region  $M$ , for a particular point within the narrow band. The local exterior region is illustrated in orange colour while the red region is known as the local interior region.

The local mask region can be mathematically construed:

$$M(x, y) = \begin{cases} 1, & \text{if } \|x - y\| < w \\ 0, & \text{if otherwise} \end{cases}$$

Where  $w$  is known as the segmented watershed region situated at the local point to be analyzed, and including the segmented watershed regions located at the points of north, east, south and west from the local point. The energy function,

$F$  of an image,  $i(x, y)$  could be mathematically written as follows:

$$F(C, u, v) = \alpha \cdot \text{Length}(C) + \lambda \int_{in} |i(x, y) - u|^2 dx dy + \lambda \int_{out} |i(x, y) - v|^2 dx dy$$

Where  $\alpha \geq 0$ ,  $\lambda > 0$ ,  $u$  and  $v$  are the average intensities levels inside and outside of the curve  $C$ , respectively.

By applying the level set approach onto the variation of the ACM, the unknown  $C$  is substituted with the unknown value  $\phi$ ,  $\phi(x, y)$  is taken as the signed distance function, where  $\phi(x, y) > 0$  inside the curve,  $\phi(x, y) < 0$  outside the curve, and  $\phi(x, y) = 0$  on the curve.

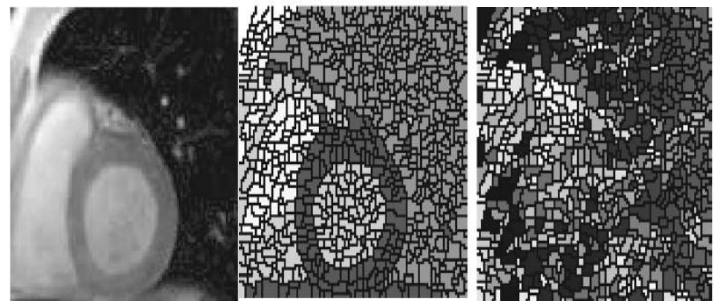


Fig. 6. a) Original image (b) expected watershed segmentation (c) over-segmentation.

So by applying Sobel filtering & watershed segmentation the system can gain expected segmentation.

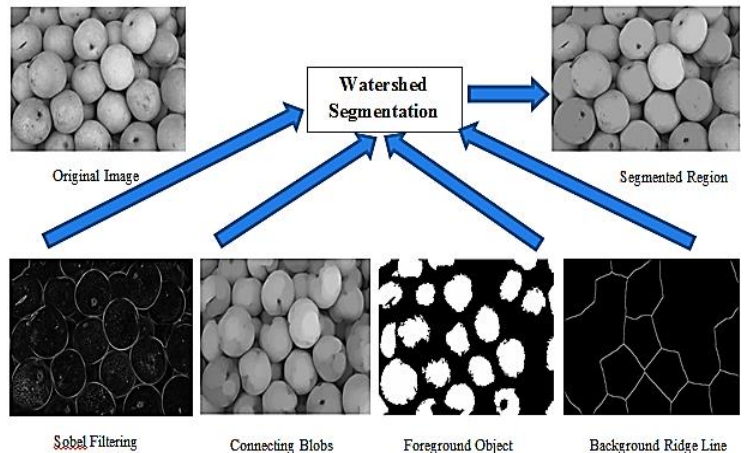


Fig. 7. The outcomes of the pre-processing procedures.

### IV. MODIFIED $H^\infty$ BASED ON LYAPUNOV STABILITY THEORY FOR LIPS TRACKING SYSTEM

An active shape model sample grey level perpendicular to the lip contour and centered at the model points.

After obtaining the lips contour from the proposed localized ACM system as discussed in the previous section, six feature



points from the detected lips region is tracked using the modified  $H^\infty$  filtering. [9]

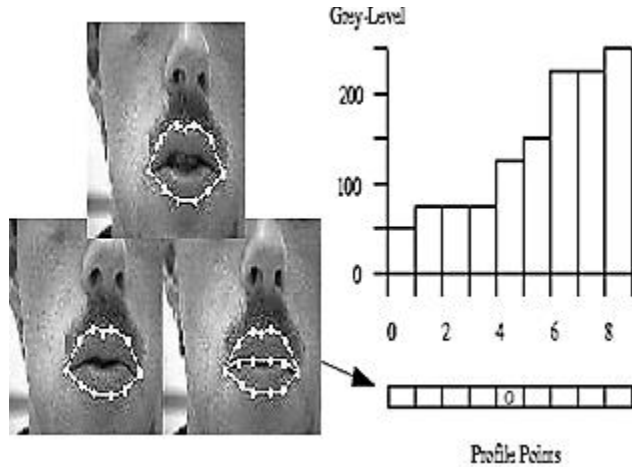


Fig. 8. Profile points & grey-level.

After obtaining the lips contour from the proposed localized ACM system as discussed in the previous section, six feature points from the detected lips region are tracked using the modified  $H^\infty$  filtering which is elaborated in this section. The feature points include the right and left lip corners, the lower central point, and three points of the Cupidon's bow as depicted in Fig. 1. The six feature points ( $P_1 - P_6$ ) have to be first extracted from the detected contour.  $P_1$  and  $P_5$  are, respectively, obtained from the right and left most of the contour. Whereas  $P_6$  situated at the bottom of the lips contour while  $P_3$  at the upper contour that has the similar vertical coordination as  $P_6$ .  $P_2$ , which is the left Cupidon's bow is obtained by calculating the maximum altitude between  $P_1$  and  $P_3$ , while  $P_4$  (the right Cupidon's bow) is the maximum altitude between  $P_5$  and  $P_3$ .

$H^\infty$  filtering is mathematically represented according to the state-space concept. State equation:

$$X_k = AX_{k-1} + w_k$$

Where  $X_k = [x_{k-p+1}^T x_{k-p+2}^T \dots x_k^T]^T$ ,  $A$  is known as the  $P \times P$  dimensional state transition matrix, which bring forward the state value,  $X_k$  from time  $k$  to  $k + 1$ ,  $w_k$  is the  $P$ -dimensional model error that analogous covariance matrix is known as  $Q_k = E [w_k w_k^T]$ .

Observation equation:  $y_k = CX_{k-1} + v_k$

Where  $y_k$  is known as the  $N$ -dimensional sequential observation vector,  $C$  is the  $N \times P$  dimensional observation function,  $v_k$  is the observation error that analogous covariance matrix is known as  $R_k = E [v_k v_k^T]$ .

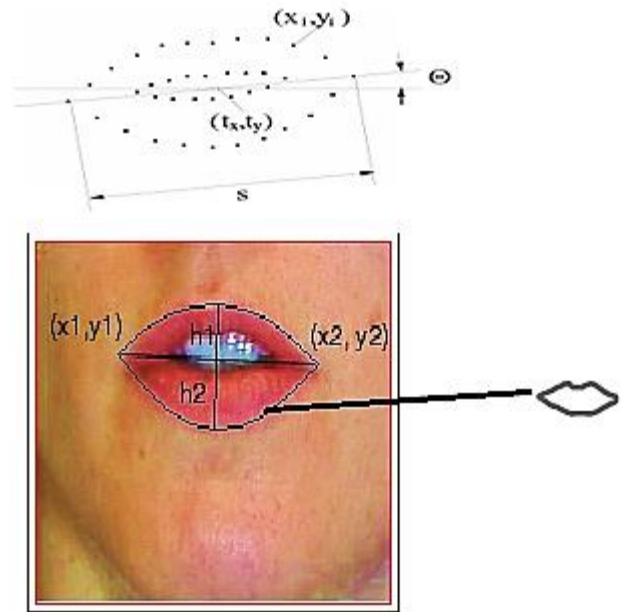


Fig. 9. Lip feature extraction and tracking.

In this paper a new gain for the  $H^\infty$  filtering parameter updating rules is proposed.

The modified  $H^\infty$  is interpreted as follows.

*Theorem:* With the given linear parameter vector,  $H_k$  and desired output,  $d_k$ , the state vector,  $x_k$  is updated as follows:

$$x_k = x_{k-1} + g_k \alpha_k$$

The modified  $H^\infty$  adaptation gain that fulfills the condition of  $\Delta v < 0$  is proposed as follows:

$$g_k = \left[ I - \frac{F_k P_{k-1} L_k F_k^T e_{k-1} - P_k}{\alpha_k} e_{k-1} \right] \frac{H_k}{\|H_k\|^2}$$

where

$$L_k = I - \gamma Q P_k + H_k^T V^{-1} H_k P_k$$

$$P_k = F P_k L_k F^T + W.$$

Where  $\gamma$  is known as the user-defined performance bound,  $Q$ ,  $W$  and  $V$  are the respective weighting matrices for the estimation error, process noise, and measurement noise. The priori estimation error,  $\alpha_k$ , is denoted as follows:

$$\alpha_k = d_k - x_{k-1}^T H_k.$$

When the time,  $k$ , approaches infinity, the tracking error,  $e_k$ , would be asymptotically converged to the value of zero.

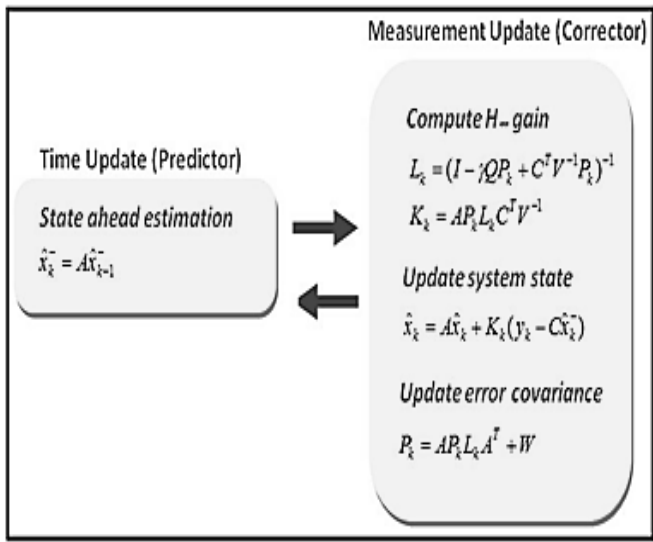


Fig. 10. The process flow of the  $H^\infty$  filter algorithm

## V. FUTURE WORK

This method can be applied in 3-D system by some modification. Face recognition, emotion detection, speech recognition from a video etc. area can achieve a high percentage of accuracy by applying lips detection method.

## VI. CONCLUSION

It is well known that visual information from lip shapes and movements helps improve the accuracy and robustness of a speech recognition system. Hence, the development of an accurate and robust algorithm for extracting lip features relevant to speech information becomes vital. Watershed-based active contour model and modified  $H^\infty$  filtering is illustrated in this paper. Compared to the global region-based ACM, the performance of the proposed localized ACM is successfully improved. The modified  $H^\infty$  tracking approach validates improved lips feature points tracking compared to the conventional  $H^\infty$  algorithm [11].

By applying Shape Constraint factor this method could be more robust. Given an image, the region of the lip may be located by using Shape Models. The overall system would be implemented into the audio-visual speech recognition for further research in the future. This system can be used to identify face or face recognition system.

But the system has High time complexity. Tracking through large number of frames point P1-P6 sometimes difficult to identify in robust situation. Sometime the skin & lip color are too similar that hard to distinguish. As a result contour detection might be tuff. For different expression like smiley, sad etc. finding accurate contour of the lip is a challenge.

## VII. ACKNOWLEDGMENT

A special thanks to Prof. Dr. Mohammad Mahfuzul Islam for giving us a chance to work with this topic. Also a superior thanks goes to Siew Wen Chin, Kah Phooi Seng, Li-Minn Ang, King Hann Lim, there works really was very helpful.

## REFERENCES

- [1] B. J. Borgstrom and A. Alwan, "A low-complexity parabolic lip contour model with speaker normalization for high-level feature extraction in noise-robust audiovisual speech recognition," *IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum.*, vol. 38, no. 6, pp. 1273–1280, Nov. 2008.
- [2] Y. Zhang, Q.-J. Liu, Y.-H. Li, and Z. Li, "Intelligent wheelchair multi-modal human-machine interfaces in lip contour extraction based on PMM," in *Proc. IEEE Int. Conf. ROBIO*, Dec. 2009, pp. 2108–2113.
- [3] M.-I. Faraj and J. Bigun, "Audio-visual person authentication using lip-motion from orientation maps," *Pattern Recognit. Lett.*, vol. 28, no. 11, pp. 1368–1382, Aug. 2007.
- [4] J. A. Dargham, A. Chekima, and S. Omatu, "Lip detection by the use of neural networks," *Artif. Life Robot.*, vol. 12, nos. 1–2, pp. 301–306, 2008.
- [5] K. S. Jang, "Lip contour extraction based on active shape model and snakes," *Int. J. Comp. Sci. Network Security*, vol. 7, no. 10, pp. 148–153, Oct. 2007.
- [6] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [7] S. Lankton and A. Tannenbaum, "Localizing region-based active contours," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2029–2039, Nov. 2008.
- [8] N. Eveno, A. Caplier, and P.-Y. Coulon, "Accurate and quasi-automatic lip tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 5, pp. 706–715, May 2004.
- [9] C. S. Wen et al., "Lips detection for audio-visual speech recognition system," in *Proc. Int. Symp. ISPADS*, 2009, pp. 1–4.
- [10] W.-Y. Chang, C.-S. Chen, and Y.-D. Jian, "Visual tracking in highdimensional state space by appearance-guided particle filtering," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1154–1167, Jul. 2008.
- [11] Siew Wen Chin, Kah Phooi Seng, and Li-Minn Ang, "Lips Contour Detection and Tracking Using Watershed Region-Based Active Contour Model and Modified  $H^\infty$ ," *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 22, No. 6, June 2012.
- [12] N. Widynski, S. Dubuisson, and I. Bloch, "Integration of fuzzy spatial information in tracking based on particle filtering," *IEEE Trans. Syst. Man Cybern. Part B: Cybern.*, vol. 41, no. 3, pp. 635–649, Jun. 2011.