Using 3D deformable template trellis to describe the movement of the lip

Foo, Say Wei; Yong, Lian; Liang, Dong

2003

Foo, S. W., Yong, L., & Liang, D. (2003). Using 3D deformable template trellis to describe the movement of the lip. Proceedings of 4th European Workshop on Image Analysis for Multimedia Interactive Services, 5(18), 445-448.

<https://hdl.handle.net/10356/91394>

https://doi.org/10.1142/9789812704337_0081

Using 3D deformable template trellis to describe the movement of the lip © copyright 2003 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.
<http://www.ieee.org/portal/site>.

USING 3D DEFORMABLE TEMPLATE TRELLIS TO DESCRIBE THE MOVEMENT OF THE LIP

SAY WEI FOO

School of EEE, Nanyang Technological University, Singapore, 639798
E-mail: eswfoo@ntu.edu.sg

YONG LIAN AND LIANG DONG

ECE Dept., National University of Singapore, Singapore, 119260
E-mail: {eleliany, engp0564}@nus.edu.sg

A novel approach to infer lip dynamics is presented in this paper. This method applies 3D deformable templates and probability trellis to track the boundaries of the lip during speaking. Experimental result shows that good accuracy of lip tracking is achieved with the proposed method.

1. Introduction

Lip reading becomes a fascinating research area in multimedia since 1990's. It is proven with many experiments that even small effort is made toward the incorporation of visual signal, the performance of a speech recognizer is improved compared with a purely acoustic recognizer. Such enhancement effect is especially prominent under noisy environment such as in bus station, airport, office and stock market. As a result, lip reading holds the promise to broaden the range of applicability of speech recognition to unfavorable environments.

Most previous studies focus on tracking the movement of the lip during speaking. In this paper, a novel approach for this purpose is also reported. By applying 3D templates, together with probability trellis, we show that the movement of the lip at various head positions can be properly tracked.

2. Tracking Strategy

The deformable template method proposed by Alan Yuille *et al* is a popular approach for lip tracking [1]. However, this is a 2D approach that the images to be processed should be the exact frontal projection of the speaker. In natural speaking, it is inevitable that the head of the speaker may rotate in different directions as illustrated in Fig. 1. Thus error is caused in 2D template matching even the lip region is properly located and segmented.

Figure 1. Change of the position of speaker's head during natural speaking

To solve this problem, 3D templates are adopted in our system. As shown in Fig. 2a, the template is affixed to a 3D ball surface so that it can rotate around three axes. Like some 2D template, the frontal projection of the template is comprised of eight Bezier curves as shown in Fig. 2b. Eleven geometric measures $g_1 \sim g_{11}$ as shown in Fig. 2c are drawn to measure the size, position and curvature of the template. Together with the rotation angles θ_1 , θ_2 and θ_3 (with respect to the three rotation axes), a fourteen-dimensional feature vector $T = [g_1, g_2, \dots, g_{11}, \theta_1, \theta_2, \theta_3]$ is thus formulated.

Figure 2. The 3D templates adopted in the proposed system

Step 1. Initialization: The input to the system is image sequence that reveals the lip region of a speaker (Fig. 1). Based on the RGB histogram of each image, the lip region is roughly located by filtering the image with a proper threshold (Fig. 3a). A template is then chosen from the template database to fit the lip region, which minimize the difference between the image and the template.

Step 2. Matching Process (Base tracking algorithm): During the process, parameters of a template are adjusted to minimize certain energy function. Considered that the hue factor is relatively independent to illumination, the energy function takes three components as defined in (1, 2, 3):

$$E_{lip} = -\frac{1}{R_1} \int_{R_1} H(x) dx \quad (1)$$

$$E_{edge} = -\frac{1}{C_1 + C_2} \int_{C_1 + C_2} |H^+(x) - H(x)| + |H^-(x) - H(x)| dx \quad (2)$$

$$E_{hole} = -\frac{1}{R_2} \int_{R_2} H(x) dx \quad (3)$$

where R_1 , R_2 , C_1 and C_2 are areas and contours as shown in Fig. 3b. $H(x)$ is certain function of the hue of the given pixel; $H^+(x)$ is the hue function of the

closest right-hand side pixel and $H^-(x)$ is that of the closest left-hand side pixel. T_{t+1} and T_t are the matched templates at time $t+1$ and t . The overall energy of the template is the linear combination of the components as defined in (4).

$$E = c_1 E_{lip} + c_2 E_{edge} + c_3 E_{hole} \quad (4)$$

where c_1 , c_2 and c_3 are positive constants. Since it is difficult to get analytical solutions to the direction of parameter adjustment: $\Delta x = -\frac{\partial E}{\partial x} \bar{r}$, minimizing E is implemented by modifying x at a decreasing step Δx as defined in (5). The optimal template is obtained by calling the above procedures iteratively.

$$\Delta x_{n+1} = \begin{cases} -\Delta x_n & \text{if } E^{n+1} > E^n \\ \Delta x_n \times e^{-1/2} & \text{otherwise} \end{cases} \quad (5)$$

Figure 3. Tracking results after each step of processing

Step 3. Constructing the Probability Trellis: In the proposed algorithm, the parameters of the template all take discrete values, where $\theta_1, \theta_2, \theta_3 = N * 0.1745 \text{ rad}$ ($N = 0, \pm 1, \pm 2, \dots$). Assume at time t , T_t^i is one of the matched templates with rotation angles $\Theta_t = (\theta_1, \theta_2, \theta_3)$. At time $t+1$, K closest angle settings $\Theta_{t+1}^1 = (\theta_1^1, \theta_2^1, \theta_3^1), \dots, \Theta_{t+1}^K = (\theta_1^K, \theta_2^K, \theta_3^K)$ are obtained from (6).

$$\Theta_{t+1}^j = \arg \min_{\Theta} [D(\Theta_{t+1}^j, \Theta_t) < D_T] \quad (6)$$

where $D(\Theta_{t+1}^j, \Theta_t)$ is the Euclidean distance between Θ_{t+1}^j and Θ_t and D_T is a predefined threshold. For the K angle settings at time t , K^2 matched templates at $t+1$ are searched with the above tracking strategy. However, only the K matches ($T_{t+1}^1, T_{t+1}^2, \dots, T_{t+1}^K$) with the highest frequency of appearances are retained. For T_t^i , the Euclidean distances with respect to $T_{t+1}^1, T_{t+1}^2, \dots, T_{t+1}^K$ are calculated via (7).

$$D^i(j) = \|T_t^i - T_{t+1}^j\|, \quad i, j = 1, 2, \dots, K \quad (7)$$

A transition probability is defined as in (8),

$$P_{ij}(t) = \exp[D^i(j)] / \sum_{k=1}^K \exp[D^i(k)] \quad (8)$$

By calculating the probabilities at each time slot, a probability trellis is constructed as shown in Fig. 4. The best path is obtained by back-tracking the nodes along the maximum probabilities.

Figure 4. The probability trellis of the matched templates

3. Experiments

The tracking results using the proposed method are illustrated in Fig. 5 and are compared with that of traditional 2D method. It is observed that the rotation of the head is better compensated with the proposed method.

Figure 5. Tracking results of the proposed method (a) and 2D template method (b)

4. Conclusion

The proposed 3D deformable template and the probability trellis are extension to the conventional 2D template method. It is applied to lip tracking with some success. The geometric features detected by the templates can be used as inputs to a recognition system. Other applications of the method may include motion detection, motion extraction and so on.

References

1. A. Yuille and P. Hallinan, "Deformable Templates," *Active Vision*, (1992).
2. T. Coianiz, L. Torresani and B. Caprile, "2D Deformable Models for Visual Speech Analysis," *Speech Reading by Humans and Machines*, (1996).
3. M.E. Hennecke, K.V. Prasad and D.G. Stork, "Using Deformable Templates to Infer Visual Speech Dynamics," *28th ACSSC*, (1994).