LIMITATION INVESTIGATION TOWARD LIPS RECOGNITION

Yun-Fu Liu, *Student Member*, *IEEE*, Chao-Yu Lin, and Jing-Ming Guo, *Senior Member*, *IEEE*

Department of Electrical Engineering,
National Taiwan University of Science and Technology,
Taipei, Taiwan

E-mail: yunfuliu@gmail.com, chaoyulin78@gmail.com, jmguo@seed.net.tw

***ABSTRACT*-**In this paper, the impact of the lips for facial recognition is investigated. In the first stage of the proposed system, a Fast Box Filtering (FBF) is proposed to generate a noise-free source with high processing efficiency. Afterward, five various mouth corners are detected though the proposed system, in which it is also able to resist beard and rotation problems. For the feature extraction, two geometric ratios and 10 parabolic related parameters are adopted for further recognition through the Support Vector Machine (SVM). Experimental results demonstrate that when the number of subjects is fewer or equal to 36, the Correct Accept Rate (CAR) is greater than 98%, and the False Accept Rate (FAR) is smaller than 0.064% (CAR>95.6%, FAR<0.083%| #Subjects$\leq$54). Moreover, the processing speed of the overall system achieves 34.43 fps (frame/sec) which meets the real-time requirement.

Keywords: Lips recognition, lips detection, feature extraction, pattern recognition, lips analysis.

1. INTRODUCTION

Nowadays, biometric systems are widely used in identification and recognition applications, due to the bio-invariant characteristics of some specific structures such as fingerprint, face, and iris. Among these features, face recognition is able to work at greater distance between the prospective users and camera than other types of features. Yet, one critical issue of the face recognition system is that the system cannot work well if the target face is partially covered. Thus, considering a smaller part of a face for further identification/recognition can be an effective way to solve this problem.

Lips belong to a part of face, and the lips contour of it has been employed on different applications such as verbal or expression recognitions, yet this feature is accompanied with additional features for recognition. For example, Çetingül et al. [1] proposed a method utilizing the lips' contour, while the additional voice information is also required. In this study, an identification system which solely adopts the features of lips' contour is proposed. In contour detection, some of the former methods employ the color information [2]-[3].

To overcome the above lighting problems, some iterative-based methods which only rely on the characteristic of contour are proposed [4]-[5]. To cope with this, Li and Cheung [6] utilized the grayscale statistic of the lips' connecting part to locate the mouth corners, and used the grayscale variation to detect the contours of the upper and lower lips. This method is able to detect the lips, yet the problem of rotated lips (image) cannot be solved. Moreover, normally the boundary of the lower lip is rather complex, this above method cannot fully cope with these issues.

In this study, both of the lips detection and recognition are discussed. To detect the lips contour, more grayscale characteristics from lips and various contrast enhancements are employed to resist the influences from various skin colors and environments. In addition, a few iterative strategies are used to obtain an accurate contour of lips, while the real-time requirement is still maintained. Moreover, a complete algorithm is designed for detecting the corner of the lower lip. In feature extraction and recognition parts, 12 various normalized features are used to distinguish the identities of different subjects through the Support Vector Machine (SVM) [7].

2. MOUTH CORNERS DETECTION

To extract the features for recognition, five different points as shown in Fig. 1 are demanded. Figure 2 illustrates the proposed algorithm, and the details are discussed below.

2.1. Preprocessing

In this study, the face region is directly detected by the powerful method namely AdaBoost [8] from an image, as an example shown in Fig. 3 with the red box. For further refining the possible region of the lips, a sub-region is roughly extracted, as an example shown in Fig. 3 with the yellow box.

For easing the influences from the camera noise and various lighting changes, and achieving a lower computation complexity simultaneously, the proposed Fast Box Filtering (FBF) and the well-known histogram stretching method are used to obtain a contrast enhanced and smoothed result. The first step of the proposed FBF method is to obtain an integral image as derived by

$$r_{i,j} = g_{i,j} + r_{i,j-1} + r_{i-1,j} - r_{i-1,j-1}, \quad (1)$$

where $r_{i,j}$ and $g_{i,j} \in G$ denote the integral value and the original grayscale value, respectively. Each of the obtained $r_{i,j}$ denotes the summation result of the whole grayscale values within its bottom-left region. Afterward, the response of the box filtering can be obtained by the following calculation,

$$b_{i,j} = [r_{i+\text{round}\left(\frac{\omega}{2}\right),j+\text{round}\left(\frac{\omega}{2}\right)} - r_{i-\text{round}\left(\frac{\omega}{2}\right)-1,j+\text{round}\left(\frac{\omega}{2}\right)} - r_{i+\text{round}\left(\frac{\omega}{2}\right),j-\text{round}\left(\frac{\omega}{2}\right)-1} + r_{i-\text{round}\left(\frac{\omega}{2}\right)-1,j-\text{round}\left(\frac{\omega}{2}\right)-1}]/\omega^2, \quad (2)$$

where $b_{i,j}$ denotes the smoothed result; the odd parameter ω denotes the size of the employed filter size; the function round$(\cdot)$ denotes the round down operation.

The rough lips region (bounding with the obtained (i_2, j_2) and (i_3, j_3) positions) is processed with the proposed FBF and the histogram stretching [3] method. The enhanced smoothed image is named G^{FBF} in this study, and which is widely-used in this paper.

2.2. Left and right mouth corners

The boundary between of the two lips is always darker than their neighboring regions because of shadow. Figure 4 shows some examples to demonstrate this observation. This characteristic is adopted in this work to robustly detect the left and right mouth corners.

2.3. Upper lip corner

Figure 4 demonstrates that the characteristic of the upper lip corner (the P3 labeled in Fig. 1) has a strong variation around the

boundary between the philtrum and the upper lip. The following two equations describe the two characteristics,

$$g_{i+1,j}^{FBF} - g_{i,j}^{FBF} \leq \eta^{top}, \quad (3)$$

$$g_{i+1,j}^{FBF} \geq T^{FBF} - w. \quad (4)$$

where $g_{i,j}^{FBF}$ and $g_{i+1,j}^{FBF}$ are searching within a line region of one pixel width. This line is perpendicular to $\overline{P1P2}$ and intersects the middle point $(\frac{i^{P1}+i^{P2}}{2}, \frac{j^{P1}+j^{P2}}{2})$. The different between the upper point $g_{i+1,j}^{FBF}$ and lower $g_{i,j}^{FBF}$ should have a positive difference and smaller than a threshold η^{top}, moreover the grayscale of the upper point $g_{i+1,j}^{FBF}$ is also higher than the threshold T^{FBF} to describe the skin brightness.

2.4. Middle point

The strategy for detecting P5 (the middle point of the lips) is similar to that of the P3. The boundary between the two lips is always darker than the rest of the lips region because of shadow. This characteristic is described as below,

$$g_{i,j}^{FBF} \leq g_{min}^{middle} + w, \quad (5)$$

$$g_{i-1,j}^{FBF} - g_{i,j}^{FBF} \geq \eta^{middle}, \quad (6)$$

where g_{min}^{middle} denotes the minimal grayscale in the search line region.

2.5. Lower lip corner

Comparing with the detection of the other corners, the detection of the bottom corner is more complex because of shadow and the grayscale difference is unstable between the colors of skin and lower lip. Figure 4 shows some instances, such as the color of the lower lip in Fig. 4(a) (focus on the region around the lower lip corner) is brighter than the color of the neighboring skin, and which is totally conversed in the case of Fig. 4(b). The case which is shown in Fig. 4(c) is rather ambiguous between color the lower lip and the color of the neighboring skin. Consequently, an algorithm is proposed to fully cope with these conditions, and Fig. 5 shows the corresponding flowchart.

Two local regions namely LSR1 and LSR2 (Lower lip Search Region) contained in G^{FBF} are defined. These regions exclude lots of unrelated parts for the searching of the lower lip corner (P4). The proposed method only searches P4 from P1, since which is sufficient and also directly reduces the computation requirement. The corresponding examples about the related LSRs are illustrated in Fig. 6. Based upon a reasonable suggestion that the P4 normally appears only on the line (called central axis) which is perpendicular to $\overline{P1P2}$, and intersects at the P5, thus the search region should around the central axis. The captured lips may suffer from rotation, the search region has to be enlarged to cover all of the possible locations of P4. Regarding top and bottom boundaries of LSR1 and LSR2 can be determined through the possible positions of P4, and the search region size is affected by the $D^{\overline{P3P5}}/D^{\overline{P1P2}}$ (the normalized heights of upper).

Before explaining the strategy about the lower lip's corner determination algorithm as shown in Fig. 5, the properties of the three points $P4^1$, $P4^2$, and $P4^3$, and their possible positions are described and illustrated in Fig. 7 for better understanding. The upper part of the $P4^1$ is brighter than its lower part, and vice versa for $P4^2$. These two scenarios are handled by the Method 1, and the remaining $P4^3$ is handled by the Method 2.

2.5.1. Method 1

This subsection explains the details of the Method 1, containing the image enhancement and feature points detection methods. To exclude the unrelated too bright or dark grayscales and further increase the contrast, an equation formulated as below is used,

$$g_{i,j}^{ELSR1} = \begin{cases} 0, \text{ if } g_{i,j}^{LSR1} < T^{LSR1,2}/\tau \\ 255, \text{ if } g_{i,j}^{LSR1} > T^{LSR1,1} \\ \frac{g_{i,j}^{LSR\,1}}{T^{LSR\,1,1} - T^{LSR\,1,2}/\tau} \times 255, \text{O. W.} \end{cases}, \quad (7)$$

where the two of thresholds $T^{LSR1,1}$ and $T^{LSR1,2}$ are yielded by the Isodata method [9], where the first threshold is calculated from all of the $g_{i,j}^{LSR1}$ contained in G^{LSR1}, and the second threshold is calculated from the candidates which meets $g_{i,j}^{LSR1} \leq T^{LSR1,1}$; parameter $\tau \in [1,2]$ denotes a factor which controls the excluded area. The region LSR1 ∩ central axis is adopted for locating the $P4^1$ and the $P4^2$.

2.5.2. Method 2

This method is designed for detecting P4 through the whole contour of the lower lip. Since the lower lips from the processed images mostly contain no shadow, it provides a clear contour for the following detection. First, since the left half of the lower lip's contour is tracked, G^{LSR2} is enhanced with the following equation,

$$g_{i,j}^{ELSR2} = \begin{cases} 0, \text{ if } g_{i,j}^{LSR2} < T^{LSR2,2} \\ 255, \text{ if } g_{i,j}^{LSR2} > T^{LSR2,1} \\ \frac{g_{i,j}^{LSR\,2}}{T^{LSR\,2,1} - T^{LSR\,2,2}} \times 255, \text{O. W.} \end{cases}. \quad (8)$$

This mechanism and the corresponding parameters are identical to that used in Eq. (7).

To track the left half of the lower lip contour, the search starts from the position P1 to the central axis. First, the contour contains the right column of P1 ($j^{P1}+1$) is detected, and the search region is only with three-pixel-height (the next search area is the detected contour points $(i, j+1)$, $(i-1, j+1)$, and $(i-2, j+1)$.

Then case when no contour point is detected in the current search region, the previous detected contour point's horizontal position is estimated as the contour point.

3. Lips Recognition

To effectively distinguish each individual, the two independent features, aspect ratio (λ_A) and curvature (λ_C), are adopted for recognition. The dimension of the feature vector $\mathbf{F} = \{\lambda_A, \lambda_C\}$ is 12, where the features contained in λ_A are the normalized heights of the upper and lower lips calculated as below,

$$\lambda_A = \left\{\overline{P3P5}/\overline{P1P2}, \overline{P4P5}/\overline{P1P2}\right\}, \quad (9)$$

where the $\overline{P1P2}$ is the normalized term. The remaining 10 features are included in λ_C, which is constructed with the parameters of five parabolic lines. To describe these lines, Fig. 8 shows a conceptual diagram for the related points for better understanding, where the lines are in blue; feature points detected by the algorithm given in the previous section are in red; the rest are in green. Notably, the names shown in this figure are widely-used for the following explanation.

To obtain the five parabolic lines, the four curves including the upper lip's C13 and C23, and lower lip's C14 and C24 are initialized. Focusing on the lower lip, due to the complex shadow, the precise contour is difficultly determined. Thus, two parabolic lines are utilized for describing the contour of the lower lip in this study.

Comparing with bottom lip, the upper lip does not suffer from shadow, thus a robust contour can be provided for further recognition. First, the K3 point is estimated for yielding the upper lip's two initial parabolic lines, named C13 and C23. The K3 point is expected as the vertex which constructs the initial line with point P1 or P2, and the line is very close to the exact contour of the lip. The steps for detecting the real contour can be summarized as below:

Step 1. Rough Search Region (RSR) construction: This step is to ease the influence caused from other unrelated parts. The RSR is used to ensure all the required contour points are within this region. The spaces of size $1 \times (|i^{K3} - i^{P3}|)$ toward top, and down of the same column construct the RSR.

Step 2. Contour detection: Searching from the mouth corner (P1 or P2) to the center of the mouth column by column, the method and parameters employed to detect the upper lip corner as discussed in subsection 2.3 are employed. Focusing on the Step 1 of the candidate selection strategy discussed in subsection 2.3, it is modified for providing a better accuracy. Given the case the first one search operation (P1's right column), the next contour point is involved in the region RSR ∩ ER, where ER denotes the Eight-neighbor Region of the previous detected contour point (the P1 in this example). Some conceptual diagrams are illustrated in Fig. 9, and the cases when various numbers of candidates are detected are also discussed.

The most top of the detected contour from the RSR constructed with C13 is considered as K13 (K23 for C23). To further describe the contour of the upper lip, the detected contour points near the C3 are considered as another parabolic line for feature extraction.

Five parabolic lines containing two lower lip's initial curves and upper lip's three detected contours are given as

$$i = \varphi_a j^2 + \varphi_b j + \varphi_c, \quad (10)$$

where the coefficients φ_a, φ_b, and φ_c are used to describe each parabolic line. Since each curve contains many points, the average coefficients are considered as the features. However, due to fact that φ_c only describes the intercept characteristic which is sensitive to the size of region detected by the AdaBoost, this coefficient is excluded. In addition, since the size of the lips can be affected by the distance between the camera and the user, to avoid the influence, $\lambda_C = \{\varphi_a^k / \overline{\text{P1P2}}, \varphi_b^k / \overline{\text{P1P2}}\}$, where k denotes various parabolic lines for recognition. In this study, the LibSVM [7] is adopted as the classifier for recognition.

4. Experimental Results

In the experiments, since no former public dataset for lips recognition is available, the database [10] refined from the Cohn-Kanade database [11] which contains 97 different subjects and in total 8795 samples of size 640x490 is utilized. The Cohn-Kanade database is formerly used for facial expression recognition, thus it contains various expressions which are not suited for lips recognition. Consequently, the images with neutral expression are adopted to construct the refined database for used, in which in total 2158 samples are involved. Notably, one of the 97 subjects is not used in the refined database since this individual does not have neutral expression. In addition, 1444 samples are used for the training with the LibSVM, and the remaining 714 samples are adopted for the following experiments.

Figure 10 shows the processing speed comparison between the traditional Box Filtering (BF) and the proposed Fast Box Filtering (FBF). The simulation platform gears with Intel Pentium 4 2.2 GHz Processor and 2GB RAM. The results demonstrate that the efficiency of the proposed FBF is simply inferior to that of the BF when filter is of size 3x3 since the integral image is required for constructing with the proposed method. In addition, the accuracy of the detected mouth corners with various filter size (ω) using FBF is simulated for optimization. The corresponding results are organized in Table I, in which the distances between the manually labeled mouth corners and the detected mouth corners are used for estimation.

The comparison of the mouth corner detection between Li-Cheung's method [6] and the proposed method is provided in Table I. The results demonstrate that their method cannot accurately detect those mouth corners. Yet, Li-Cheung's method can achieve the processing speed of 68.92 fps which is faster than the proposed corner detection of 45.24 fps, since lots of the computations are consumed on the proposed lower lip corner's detection. Nonetheless, the processing speed on this stage of the proposed method still not jeopardizes the real-time processing requirement of the overall proposed system.

Two criterions, Correct Accept Rate (CAR) and False Accept Rate (FAR), are used to estimate the performance of the proposed recognition system. Performance is shown in Fig. 11.

5. Conclusions

In this study, the ultimate capability of the lips for facial biometrics is investigated. Thus, a lips recognition system with well-designed parameters is proposed to achieve surprisingly recognition accuracy when only a partial face is available. Several stages are involved in the proposed system, and the corresponding contributions are organized below: 1) The proposed Fast Box Filtering (FBF) provides a faster processing speed than typical Box Filtering scheme when a greater filter size is required. 2) The mouth corners detection scheme demands fewer iteration compared to former schemes in the literature, and thus the overall recognition system can operate in real-time fashion. 3) This proposed lips recognition is able to handle the critical issues when beard or rotation is involved.

References

[1] H. E. Çetingül, E. Erzin, Y. Yemez, and A. M. Tekalp, "Multimodal speaker/speech recognition using lip motion, lip texture and audio," Signal Process, vol. 86, no. 12, pp. 3549-3558, 2006.

[2] M. M. Hosseini and S. Ghofrani, "Automatic lip extraction based On wavelet transform," in Proc. WRI Global Congress on Intelligent Systems, vol. 4, pp. 393-396, 2009.

[3] R. C. Gonzalez and R. E. Woods, Digital Image Processing 2/E, Upper Saddle River, NJ: Prentice-Hall, 2002.

[4] H. Seyedarabi, W. Lee, and A. Aghagolzadeh, "Automatic Lip Tracking and Action Units Classification Using Two-Step Active Contour and Probabilistic Neural Networks," Canadian Conference on Electrical and Computer Engineering, no. 4054825, pp. 2021-2024, 2007.

[5] M. Rizon, M. Karthigayan, S. Yaacob, and R. Nagarajan, "Japanese Face Emotions Classification Using Lip Features," in Proc. Geometric Modeling And Imaging (GMAI), no. 4271734, pp. 140-144, 2007.

[6] M. Li and Y. M. Cheung, "*Automatic lip localization under face illumination with shadow consideration*," Signal Processing, vol. 89, no.12, pp. 2425-2434, 2009.

[7] R. E. Fan, P. H. Chen, and C. J. Lin. "Working Set Selection Using Second Order Information for Training Support Vector Machines," the Journal of Machine Learning Research, vol. 6, pp. 1889-1918, 2005.

[8] P. Viola and M. Jones, "Robust real-time face detection," International Journal of Computer Vision, vol. 57, no 2, pp. 137-154, May 2004.

[9] A. El-Zaart, "Images thresholding using Isodata technique with gamma distribution," Pattern Recognition and Image Analysis, vol. 20, no. 1, pp. 29-41, 2010.

[10] Lip Recognition Database: http://msp.ee.ntust.edu.tw/public file/FaceRecSet.rar

[11] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," IEEE Intl. Conf. Automatic Face and Gesture Recognition, Grenoble, France, pp.46-53, 2000.

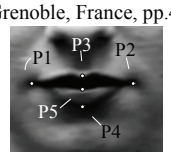


Fig. 1. Five feature points of lips.

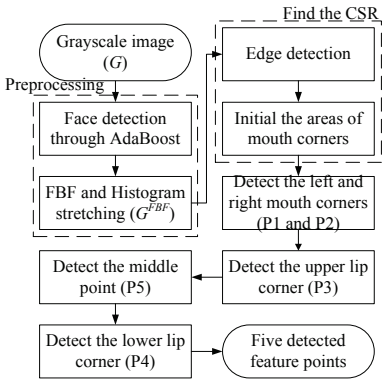


Fig. 2. Flowchart of the proposed lips detection scheme.

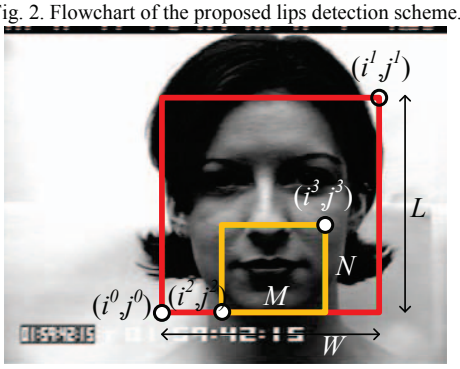
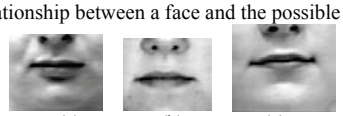


Fig. 3. Relationship between a face and the possible lips region.



(a) (b) (c)

Fig. 4. Three different lips images with and without shadow at the bottom of lower lip.

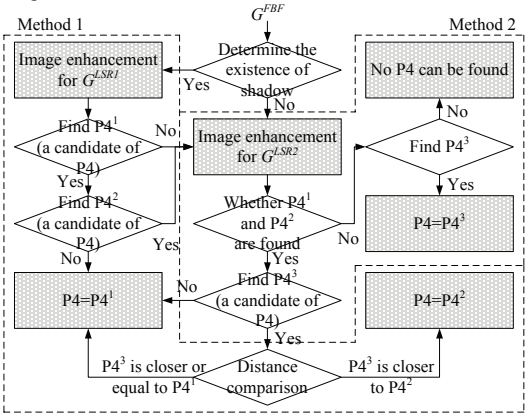
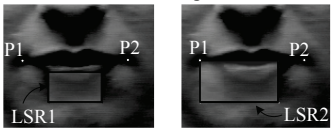
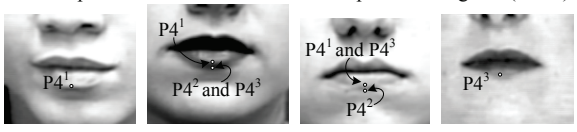


Fig. 5. Flowchart of the lower lip's corner determination.



(a) (b)

Fig. 6. Examples of the two different Lower lip Search Regions (LSRs).



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

Fig. 7. Practical detected P4 candidates. (a) Only $P4^1$ is found. (b)-(c) Two cases when $P4^1$ and $P4^2$ are found. (d) No shadow is detected, and $P4^3$ is found.

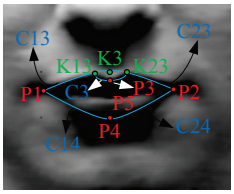
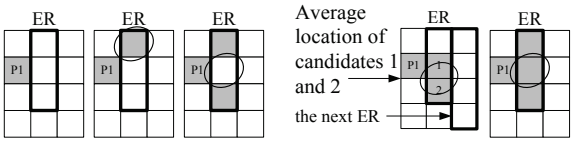


Fig. 8. Points and curves definitions for λ_C feature extraction.



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

Fig. 9. Various upper lip contour search scenarios. (a) No candidate is found in ER. Numbers of (b) one, (c)-(d) two, and (e) three candidates are found in ER.

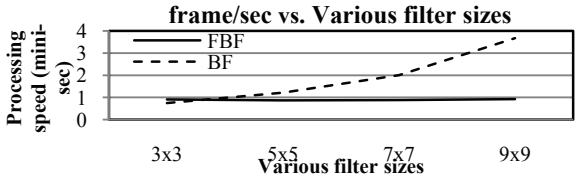
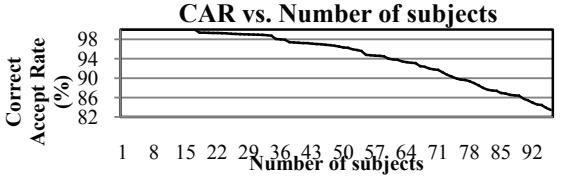
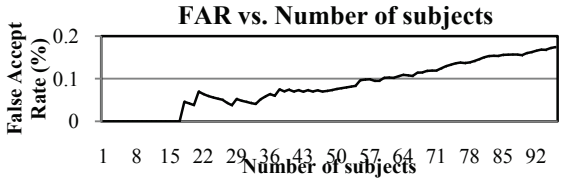


Fig. 10. Comparison between the proposed Fast Box Filtering (FBF) and traditional Box Filtering (BF).



(a)



(b)

Fig. 11. Performance of the proposed recognition system. (a) CAR. (b) FAR.

TABLE I. AVERAGE ACCURACY OF THE FIVE DETECTED MOUTH CORNERS. (DATA: THE MEAN <STANDARD DEVIATION> OF THE DISTANCE DIFFERENCE (UNITS: PIXELS))

Filter Size (ω) / Mouth corners	3x3	5x5	7x7	9x9	Li-Cheung's method [6]
P1	0.09 <0.16>	1.23 <0.72>	2.74 <1.43>	4.74 <2.31>	40.45 <11>
P2	0.08 <0.15>	1.27 <0.78>	3.11 <1.91>	5.65 <3.22>	41.4 <11.8>
P3	0.12 <0.22>	0.97 <0.52>	1.6 <1.91>		14 <8.45>
P4	1.36 <1.16>	1.01 <1.12>	2.14 <1.54>		25 <7.12>
P5	0.08 <0.15>	0.7 <0.61>	1.47 <0.78>		33.6 <12.5>