Detection of Facial Parts based on ABLATA

Siddhartha Choubey
Shri Shankaracharya Technical Campus, Bhilai

Vikas Singh
Shri Shankaracharya Technical Campus, Bhilai

Abha Choubey
Shri Shankaracharya Technical Campus, Bhilai

ABSTRACT
Facial feature detection from standard 2D RGB images is a well-researched field but out of prolific techniques there isn’t much efficacy is achieved in the previous studies that can extract feature data even for a low quality images in real time. Hence, we propose an algorithm based on Attribute Based Level Adaptive Algorithm (ABLATA) which use recursive data estimates for this task. While the recursive data estimates learns the relation between patches of the localized segmented blocks and the location of nodes covering the region of the required regional properties of the face.

Keywords
Face detection, ABLATA, feature selection.

1. INTRODUCTION
Facial feature extraction has been an active area of research in computer vision for increasing applications like human computer interactions [1-5]. In the previous studies had reported results which similar tasks has been carried out effectively over medium quality images [2-7]. Out of which regression forest method has been a useful tool to for daunting tasks involving task efficiency for operations associated with computer vision applications [6]. This method employs probability estimation over the parametric mapping of the given image and subsequent depth patches, while ensuring that the position & orientation of the imagery objects are encoded through various learning schemes. Whereas, other related methods like Houhg forest estimation & pose estimation detects objects from 2D slandered RGB images from rich training data sets [7-13]. Even though, most of the available techniques haven’t yet been successful in providing real time performances which largely restricts its applications and in addition with that this methods have been ineffective against low quality images.

Now, there are other models popularly known as active appearance model which uses textures above the area of facial region to regressively fit a linearly generative model to a masked test image [8,9,13]. However, the performance of such algorithms are easily influenced by the environmental changes like lighting variation, complexity of the model, and an unrequired affinitive for the average face; this lead such methods to give poor performance against untrained objects and remain untouched to complete the tasks with low resolution images [14-22].

This work is focused on detection and segmentation of micro-facial expression from input images using temporal cues. The feature extraction process is implemented by the custom modeled ABLATA algorithm [23-27]. The proposed methods aids feature extraction without using template matching techniques and thus eliminating unnecessary time consumption during the requisite computational operations. This work exploits the advantages of ABLATA for automatically segmenting and recognizing human facial expression from 2D standard RGB facial images. The novelty of this architecture is that it achieves detection time close to the human accuracy while processing images in real-time, while the data is matched from the previous study of human face encoding [24].

2. Methodology
Facial Feature Detection Through ABLATA

Thus, based on weighted spatial localization of neighboring pixels the threshold value of the cluster is determined through t using ABLATA [32]. Mathematically, an image is a two dimensional (2 D) function, \( f(x, y) \), where \( x \) and \( y \) are the coordinate values in spatial domain or plane; and the magnitude of \( f(x, y) \) is the intensity value of pixel at \( \text{(x, y)} \). If \( x \), \( y \) and the magnitude of \( f(x, y) \) in an image are discrete quantities then the image is said to be digital image. Image may be represented as two dimensional matrices whose elements are intensities of pixels present in image. Almost all image processing related operations operate on these pixels either in spatial domain or in frequency domain or transform domain.

The function \( f(x, y) \) can be expressed as:

\[
\begin{pmatrix}
    f(0,0) & \cdots & f(0,N_y-1) \\
    \vdots & \ddots & \vdots \\
    f(N_x-1,0) & \cdots & f(N_x-1,N_y-1)
\end{pmatrix}
\]  (1)

Now, each digital image has certain finite number of elements characterized by some coordinate values and intensity value.

The coordinate indicates the position of pixel in an image. In Equation (1) the image elements \( f(N_x-1,N_y-1) \) represent the maximum number of resolution starting from \( f(0,0) \).

Suppose that ‘\( f \)’ is the set of categorized pixels band and ‘\( P \)’ is a uniformity predicate defined over groups of connected pixels. Segmentiation is simply a partitioning of the set \( F \) into a set of connected subjects or regions (\( P_1, P_2, \ldots, P_n \)) such that \( \bigcup \ P_i = F \) with \( P_i \cap P_j = \emptyset \) when \( i \neq j \). The uniformity predicate \( i = 1 \) pixels represented as \( P \) (\( P_i \)) is true for all regions \( P_i \cup P_j \) and is false when \( P_i \) is adjacent to \( P_j \).

The thresholding algorithm for binary images is applied as:

\[
f_i := \sum_{r=0}^{m} r \cdot \left( g_i \leq g_{\text{block}} < g_{i+1} \right) \]  (2)

where, \( r(.) \) is the mean value; \( g_{\text{block}} \) and \( g_{i+1} \) are the lower bound and upper bound respectively of the given thresholding pixel boundary condition.

The unnatural bias for partitioning is avoided by selecting small sets of points and different measure of dissociation. The problem with such criterion for thresholding is that it does not
consider association with clusters. In order to circumvent this problem, the cost of thresholding at runtime as a function of the total pixel threshold to all those levels formed in the above step is determined and taken in account through the pixel association rule.

Thus, we have the generic equation normalization is defined as:

\[
f_t(N_x, N_y) = \sum_{i=0}^{r} \left( \frac{\text{cut} (N_x - 1, V)}{\text{assoc} (N_x - 1, N_y)} + \frac{\text{cut} (N_x, N_y)}{\text{assoc} (N_y, V)} \right)
\]

where, \( \text{assoc} (N_x, N_y) = \sum_{u \in m \in n} W(u, v) \) is the total connection from pixels of set A to all set B. By using this definition of the disassociation between the groups, small isolated points are partitioned out and will no longer have distinct \( N \) values, since the cut value will almost be a large percentage of the total connections from the small set to all other pixels. If no other level changes are found then terminate the operation. The mechanism of segmented image is finally generated after extraction operation.

Thus, let us suppose that levels based dependencies between different facial parts can be expressed as \( p(A|B) \) where \( A \) is the sets of nodes estimated by ABLATA in previous steps during normalization and \( B \) is the voting element for \( A \) which express the facial description for the local sub patches in form of a sets of nodes given by the training sets of a number of images on iteration basis. Thus, the location of different parts is dependent i.e.,

\[
p(A|B) = \prod_{i=1}^{N} p(f_i|B) \quad (4)
\]

But since, equation 4 creates a consolidated regions \( R \) from Levels \( L \) and Pixels \( P \). Therefore, the entropy \( H \) of two positions can be mathematically defined as:

\[
H(B) = \frac{N \sum_{i=1}^{n} p(A|B)(A \in B)}{||B||} 
\]

\[
t_{h}(R_i, L_i) = \begin{cases} 
H(B), & \text{if } f_t(R_i, L_i) < S \\
(H(A \cap B), & \text{if } f_t(R_i, L_i) > S
\end{cases} 
\]

\[
S = t_{h} \frac{1}{\text{area}} 
\]

where, \( S \) is the splitting function and \( f(R_i, L_i) \) which is the function represented in equation 3 recursively with \( R_i \& L_i \) as its hierarchical input parameters & \( t_h \) is the segmented region with integrated facial features.

**Algorithm: Detection of facial Feature from ABLATA**

**Input:** \( f(x, y) \) Input Image & B i.e., the voting elements of nodes for a facial expression.

**Output:** \( f'(x, y) \) with \( t_h \) the Segmented Facial Portion in allmage.

**Step1:** Read Image:

\[
f(x, y) = \begin{bmatrix}
  f(0, 0) & \cdots & f(0, N_y - 1) \\
  \vdots & \ddots & \vdots \\
  f(N_x - 1, 0) & \cdots & f(N_x - 1, N_y - 1)
\end{bmatrix}
\]

**Figure.1: Sample Image Read.**

**Step2:** Initialize Thresholding through ABLATA:

\[
f_t = \sum_{i=0}^{m} r \left( (g_i < f_{\text{block}} < g_{i+1}) \right)
\]

**Step3:** Estimate Normalization of Each pixel in conjugation with each other:

\[
f_t(N_x, N_y) = \sum_{i=0}^{r} \left( \frac{\text{cut} (N_x - 1, V)}{\text{assoc} (N_x - 1, N_y)} + \frac{\text{cut} (N_x, N_y)}{\text{assoc} (N_y, V)} \right)
\]

**Figure. 2: Normalization & formation of levels based on attribution from ABLATA.**

**Step4:** Find levels based dependencies between different facial parts:

Begin for \( I \): \( A \)

\[
p(A|B) = \prod_{i=1}^{N} p(f_i|B)
\]
Step 5: Find entropy $H(B)$ for each mutual value of $B$ and $A$

\[ H(B) = \sum_{n=1}^{N} \frac{\sum_{i \in B} (A \setminus B)}{||B||} \]

Step 6: Calculate final segmented regions $t_i$ of facial images:

\[ t_i(R_i, L_i) = \begin{cases} H(B), & \text{if } f_i(R_i, L_i) < S \\ H(A \cap B), & \text{if } f_i(R_i, L_i) > S \end{cases} \]

\[ S = \frac{t_i}{\sum_{i=1}^{N} t_i} \]

Step 6: Repeat Steps 4-6 until whole Image is traversed.

3. CONCLUSION

Figure 5: Detected Facial Parts

Figure 6(A): Error Evaluation while detecting facial parts through the proposed method.

Figure 6(A): Comparison of detection through several voting points i.e., $B$ or fiducial point used to detecting facial parts through the proposed method & through humans
As shown in fig.6(A) above the total number of facial detection parts is improved over the course while ensuring that the error involved in detection remains nominal. Also, in fig. 6(B) we have compared the rate of detection of the facial parts by the proposed algorithm and by the human [24]; which shows that error ratio involved in detection remains comparable to that of human subjects. Thus, the algorithm that we have presented performs the requisite computational operations in a real-time for facial feature detection based on the improved concept of ABLTA (fig 7). This study effectively estimate the positional & several other facial landmarks which once were conditional to the probability ensembles with the global face properties. Our work will benefit other studies involving modeling the facial appearance and its location through other facial feature points which is conditional pose authored by the subjects. In our future work, we intend to model the properties like that of emotion detection on top of the proposed study to carry out other relevant computer vision tasks.

![Sample instances of detection using the proposed study](image)

4. REFERENCES


