



Image coding based on maximum entropy partitioning for identifying improbable intensities related to facial expressions

SEBA SUSAN*, NANDINI AGGARWAL, SHEFALI CHAND and AYUSH GUPTA

Department of Computer Science and Engineering, Delhi Technological University, Delhi 110042, India
e-mail: seba_406@yahoo.in

MS received 26 October 2014; revised 2 April 2015; accepted 31 May 2016

Abstract. In this paper we investigate information-theoretic image coding techniques that assign longer codes to improbable, imprecise and non-distinct intensities in the image. The variable length coding techniques when applied to cropped facial images of subjects with different facial expressions, highlight the set of low probability intensities that characterize the facial expression such as the creases in the forehead, the widening of the eyes and the opening and closing of the mouth. A new coding scheme based on maximum entropy partitioning is proposed in our work, particularly to identify the improbable intensities related to different emotions. The improbable intensities when used as a mask decode the facial expression correctly, providing an effective platform for future emotion categorization experiments.

Keywords. Variable length coding; Huffman coding; Shannon–Fano coding; entropy coding; maximum entropy partitioning; facial expression recognition; non-extensive entropy with Gaussian gain.

1. Introduction

Facial expression analysis has assumed great significance in the recent advances made in human–machine interaction. The facial expressions depict various human emotions such as pain, surprise, sorrow and happiness [1, 2] which if interpreted correctly by machine learning algorithms can be of immense practical use, such as for the assistance of elderly [3] and infants [4], providing medical aid to patients in pain who are unable to communicate [5–7], or devising modern day gadgets and applications that react to human emotions [8], a few examples being humanoid robots that recognize faces and facial expressions [9, 10], expression glasses for facial expression recognition [11], adaptive music controller [12], household gadgets, etc. Some online facial expression recognition applications include the FaceReader [13] and LAFTER [14]. However these are limited to certain range of emotions only such as a positive or negative emotion, whether the person is laughing or not, etc. which reduces emotion recognition to a binary problem. In the actual scenario, decoding the human emotion is a most complex task due to the complex state of the human mind, and the presence of mixed emotions both positive and negative cannot be entirely ruled out. The studies linking facial expressions to human psychology and behavior span back to several decades [4, 7, 15, 16]. Since then, several researches have been directed towards developing more efficient machine learning algorithms for

deciphering human emotions from facial expressions [17]. Automated facial expression analysis usually proceeds by segmenting the skin pixels, locating the eye, nose or mouth coordinates, and cropping the face followed by normalization to a standard size [18–24]. Automated face cropping tools such as the Viola Jones Face detector [25] are readily downloadable from the web. The cropped and normalized face is subjected to feature-extraction and subsequent categorization into classes, be it the class of the subject (face recognition) or the class of the emotion (emotion recognition/facial expression analysis). The facial features extracted play a great role in the accuracy of the classification. Texture features are more than often used for representing faces, examples being the local binary pattern (LBP) for images/video frames [26] and the hybrid local binary pattern along three orthogonal planes (LBP-TOP) for video blocks [27]. Texture cues are particularly suited for the face and facial expression recognition problem since they represent patterns directly associated with the shape and structure of a person’s face or dynamic patterns in subsections of the face associated with some emotion, such as the shape or appearance of the mouth and lips, or the creases in the forehead or near the eyes and the crinkling of the nose [28–31]. Grid-based facial expression analysis using localized texture cues and a normalized similarity index for matching decodes human emotions successfully and detects the presence of mixed, compatible emotions [32]. Similar lines of reasoning have been used by several face models such as the extended Active Shape Model [33] for locating landmark points in an upright face.

*For correspondence

In this paper, we seek to highlight the improbable, imprecise and indistinct image intensities that are indicative of the emotion being portrayed in the cropped facial image. We achieve this by devising a novel image coding scheme based on maximum entropy partitioning which assigns longer codes to those low probability symbols/intensities that are difficult to distinguish from others in terms of their probabilities of occurrences and lack a proper definition. The paper is organized as follows: section 2 reviews the conventional Huffman and Shannon–Fano image coding schemes, section 3 outlines our overall methodology for locating improbable intensities related to emotions by information-theoretic coding, section 4 discusses the new image coding scheme for identifying improbable intensities based on the maximum entropy based partitioning. Section 5 analyzes the experimental results and section 6 draws the overall conclusions.

2. Huffman and Shannon–Fano coding of images: a basic review

We start by discussing the two most important and popular information-theoretic coding techniques. The Huffman [34] and Shannon–Fano coding schemes [35] construct optimal codes for a set of symbols $\{x\}$ and their probabilities $\{p(x)\}$. Compared to Shannon–Fano coding, the Huffman coding is more efficient in terms of the average code-length generated. However the Shannon–Fano codes are within one bit of their ideal theoretical length of $-\log p(x)$ [36]. The two techniques are now described in detail.

The Huffman coding algorithm is a source encoding technique used for unequal length coding. It is described in the context of images [37] below.

Step 1 Construct a tree based on the gray-level frequencies obtained from the image histogram. This is achieved using the following steps:

- i. Sort the probabilities from the image histogram in the descending order.
- ii. Add the two lowest probabilities to form a node at the higher level.
- iii. Repeat until the highest node is a single node.

Step 2 Assign binary codes to each gray level/node as follows:

- i. Assign null code to the highest node
- ii. Whenever a node splits into two, assign a ‘0’ to the left child and a ‘1’ to the right child.
- iii. Repeat until the terminal nodes are reached.

Huffman coding is an example of unequal length coding [38] where the histograms of the symbols are peaked. It is also an efficient coding technique which assigns the shortest codes to the most probable intensities and the longest codes to the least probable intensities. It is stated to

have the highest efficiency and the shortest average code-length [39]. Huffman coding is popularly used for image cryptography [40] and image steganography [41] in which the constructed binary codes are assigned to image intensity levels. In our work we further investigate its use for decoding facial expressions by highlighting the low probability facial intensities related to emotions.

The Shannon–Fano coding scheme constructs sub-optimal codes which are not as nearly efficient as that of Huffman codes. The steps for constructing Shannon–Fano codes are described below.

Step 1: Given a list of symbols with their probabilities, arrange them in the decreasing order of the probabilities.

Step 2: Divide this list into two partitions such that the sum of probabilities in one partition roughly equals the sum of probabilities in the other.

Step 3: Assign a ‘0’ to the first partition and a ‘1’ to the second partition.

Step 4: Repeat steps 2 and 3 iteratively on each partition created, until only one symbol is left in each partition and all symbols have been assigned codes.

The above coding scheme works best when the symbols are more or less equally likely and a single bit is sufficient to divide them into two partitions. An example of an application is text compression where Shannon–Fano coding is more useful than Huffman and Arithmetic coding [42]. A reason why Shannon–Fano coding is otherwise almost never used is its low efficiency when the mismatch between the sums of the probabilities of the two partitions is high, as often happens with real-world images having peaked histograms or intensities with unequal probabilities.

3. Locating improbable intensities in an image by information-theoretic coding

In this section we first describe pixel masking based on the thresholding of code-length. The K symbols for image coding are the gray-level intensities in the image quantized to K bins. The overall methodology for selective pixel masking based on the longer length of the codes assigned is given below.

1. Apply Information-Theoretic coding (such as Huffman and Shannon–Fano coding) to generate separate codewords for each intensity value.
2. Check the length of the codewords assigned to each intensity value.
3. Map the code-length back to each pixel position by closest bin matching with pixel intensity. Plot the histogram of the code-lengths associated with the pixels in the image.
4. Determine the optimal threshold value of the length of codewords t_len from the histogram of code-lengths.

5. For each pixel
 - If $(\text{length}(\text{codeword})) \leq t_len$
Assign 0 (color code for black) to the corresponding pixel
 - else if $(\text{length}(\text{codeword})) > t_len$
Assign 1 (color code for white) to the corresponding pixel.
6. The image so constructed is a binary mask which highlights the improbable intensities in the image that are assigned longer code-lengths by information-theoretic coding and are indicative of the specific emotion being portrayed.
7. Alternatively, substituting the 0's in the above result with the original image intensity values and substituting the 1's with 0's would yield the masked original image devoid of any emotions.

The selective pixel masking procedure outlined above is demonstrated for Huffman coding of two real-world images in figure 1, with threshold $t_len = 8$ arbitrarily selected for best results for a 256-bin histogram ($K = 256$).

In figure 1(b), the image intensities excluding the improbable ones are shown. The missing portion includes parts of the face of the person in the picture which is significant information. In figure 1(d) the faces of the two girls in the picture are low probability intensities as compared to the background shades and hence are assigned longer codes.

We observe from the above two examples that applying Huffman coding to random photographs would identify faces or at least parts of it as being improbable events and these are assigned longer codes. This highlights the potential of Information-theoretic coding for segregating pixels into categories of the improbable intensities (which include parts of faces in the photographs) and the more probable intensities (which include the majority of the background).

Theorem 1 Symbols that are assigned longer codes contribute significant information that helps in identification.

Proof In the variable length coding techniques like Huffman and Shannon–Fano coding, the shorter codes are assigned to the symbols that are highly probable and predictable and thus carry less information since they are more or less expected to occur. The idea behind the encryption is that guessing this highly probable symbol will not give much information to the attacker hence lesser number of bits is assigned to cut the computational cost. As the length of the codeword increases, the guesswork becomes more difficult for the attacker and the symbols now pertain to useful information that must be hidden to secure the primary message content [43]. The expected value of the guesswork $E(G(W_l))$ (defined as the number of guesses for the symbol W_l till the symbol is guessed) increases proportionally with the length of the codeword l [43].

3.1 Motivation for our problem

We start with the assumption that it is the improbable and indistinct intensities in an image that contain the traces of emotion in a human face. Recent literature highlights the study of facial micro-expressions which are subtle changes in facial expressions when a person tries to conceal his real emotions [44]. The identification of facial intensities particular to an emotion makes the recognition task easier. The dynamic changes in facial expressions play a greater role in emotion recognition as compared to the static information in video sequences [43]. A recent work on face recognition from the videos of subjects dynamically portraying emotion through facial expressions emphasizes the need for capturing the salient spatial information in the video, which is the face itself that appears in every frame irrespective of its dynamically changing facial expression [45]. This was achieved using the 3D-Difference Theoretic Texture Features (DTTF) that relies on averaging the pixel-defined gray level difference statistics throughout the entire volume of the video block to extract the high probability facial intensities contained in every frame. The averaging procedure would subdue the effect of the intensity changes pertaining to the dynamically changing emotion and helps in recognizing the face irrespective of its emotion. In our work on facial expression analysis, on the other extreme, we seek to highlight the improbable intensities in the facial image that would indicate characteristics of the emotion being portrayed. The face with the improbable intensities highlighted would appear differently for different emotions. In collaboration with texture analysis techniques, a future emotion categorization tool is visualized. We use information-theoretic coding in a novel attempt to identify the parts of the face that are vulnerable and specific to different emotions. The vulnerable intensities are defined as those that are assigned longer codes on account of their being improbable and difficult to distinguish in terms of their occurrence frequency which is neither too low nor too high. The more obvious and predictable facial intensities do not give any clues of the emotion being portrayed.

In a cropped frontal face image the improbable intensities mostly pertain to:

- i. The dilation of the eyes and the arch of the eyebrows
- ii. The creases in the forehead and around the eyes
- iii. The crinkling of the nose and the dilated ends of the nostrils
- iv. The shape of the mouth which is the most expressive part of the face whether it is open, closed, drooping, shut tightly or stretched in a smile
- v. The shadows along the sides of the face where lighting is less change with different emotions

Our aim is to locate and identify the ‘difficult’ intensities listed above which not only have low probabilities but also separate out last due to their indistinctiveness and lack of

proper definition with respect to the rest of the intensities in the facial image. This refines our present definition of ‘vulnerable’ intensities related to facial expressions as those having low and indistinct probability values and having the longest codes assigned.

4. Proposed image coding using maximum entropy based partitioning for analyzing facial expressions

In this section we describe in detail the new image coding technique for facial expression analysis using a novel level-partitioning based on the maximum entropy principle. The maximum entropy principle has been used before in applications such as image thresholding [46] to create optimal partitions of probability histograms. The central idea is that for the optimal partition, the sum of entropies of the two partitions is maximum since each of the two subsets contains equally likely members at that point. We apply this principle to create the partitions in Shannon–Fano coding which at present relies on roughly matching the sums of probabilities in the two subsets. The steps follow the order of Shannon–Fano coding in their execution and code assignment procedure; however, the partitions between symbols at each level are determined from the positive entropy spike or simply the maximum entropy among all possible partitions, as explained below with the help of relevant notations and formulae.

Shannon entropy for a complete probability distribution is defined as

$$H(\mathbf{p}) = - \sum_i p_i \log p_i. \quad (1)$$

The entropy in (1) is additive i.e. for the statistically independent case,

$$H(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}) + H(\mathbf{y}). \quad (2)$$

The steps of the proposed coding algorithm for a set of symbols (image intensities quantized to K bins in our case) and their associated probabilities (derived from the image histogram) are outlined below.

Step 1: Given a list of symbols with their associated probabilities, arrange them in the decreasing order of the probabilities.

Step 2: Divide the set/subset of sorted symbols tentatively into two probability partitions $\mathbf{p1}$ and $\mathbf{p2}$ (all possible partitions are iteratively tested starting with the lone top symbol versus the rest in iteration $k = 1$ and ending with all the symbols versus the last symbol in the last iteration).

Step 3: Find the entropy values $H_k(\mathbf{p1}_N)$ and $H_k(\mathbf{p2}_N)$ corresponding to the two probability partitions $\mathbf{p1}$ and $\mathbf{p2}$ for the k th iteration, after normalizing the probabilities first so that they sum up to 1 within each partition as shown below.

$$\mathbf{p1}_N = \frac{p1}{\sum_i p1_i} \quad (3)$$

$$\mathbf{p2}_N = \frac{p2}{\sum_i p2_i} \quad (4)$$

$$H_k(\mathbf{p1}_N) = - \sum_{x \in \mathbf{p1}_N} x \log x \quad (5)$$

$$H_k(\mathbf{p2}_N) = - \sum_{y \in \mathbf{p2}_N} y \log y \quad (6)$$

Step 4: Find the weighted sum of the two (equiprobable) entropies $H_k(\mathbf{p1}_N)$ and $H_k(\mathbf{p2}_N)$ to generate the entropy of the source $H_k(S)$ for the k th iteration as

$$H_k(S) = \frac{1}{2} H_k(\mathbf{p1}_N) + \frac{1}{2} H_k(\mathbf{p2}_N) \quad (7)$$

Step 5: Repeat steps 3–4 for all possible partitions k starting from the top with at least one symbol being in one partition

Step 6: Find the partition that corresponds to maximum entropy as

$$opt = \arg \max_{\forall k} H_k(S) \quad (8)$$

Step 7: Assign a code ‘0’ to the first partition and a code ‘1’ to the second partition.

Step 8: Repeat steps 2–7 recursively on all subsets of the partitions created until no more partitions are possible and all the symbols have been individually assigned codes.

We describe next two theorems which further helps in strengthening the validity of the proposed maximum entropy based partitioning and a propose a further improvisation of results by the substitution of the extensive Shannon entropy by the non-extensive entropy with Gaussian gain.

Theorem 2 Maximum entropy partitioning assigns longer codes to groups of symbols in an image that are indistinct in terms of their number of occurrences and assume a uniform distribution among themselves.

Proof The occurrence of an event such as a smile on a face is related to the co-occurrence of several low probability intensities other than the normal predictable facial intensities. To detect the closely related intensities associated with a smile such as the creases in the skin near the mouth and the contour of the lips we search for uniform probability densities in the probability graph of the facial image. Our iterative procedure described above, zeroes in on these inseparable intensities towards the end, assigning them the longest codewords. Therefore, to isolate image intensities that contribute to an event we have to find groups of image intensities which have a similar occurrence frequency. They would thus share a uniform probability

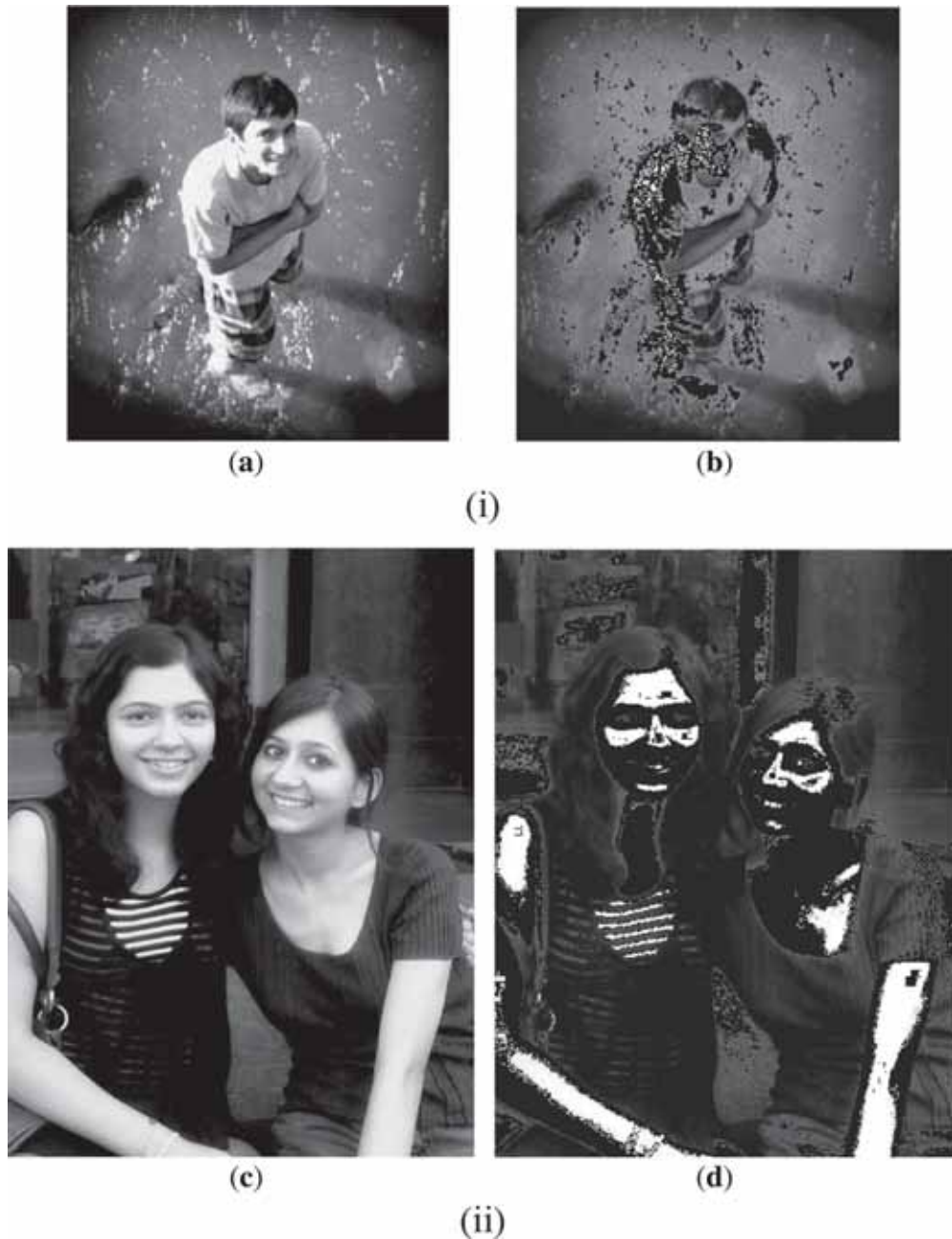


Figure 1. (i) The masked original image using Huffman coding: (a) Original image 1. (b) Reconstructed image after masking the improbable intensities by 0 (black). (ii) Highlighting the improbable intensities in the image: (c) Original image 2. (d) Reconstructed image after highlighting the improbable intensities by 1 (white).

distribution among themselves. The maximum entropy principle in general, states that the optimal probability partition is obtained at the point where the sum of entropies of the two partitions is maximum or in other words when the two partitions individually approximate uniform distributions. For instance from (7), the entropy between two sub-partitions is stretched to the maximum limit when the distributions in both partitions are uniform

$$H_k(S)_{\max} = - \underset{p^{1_i \rightarrow \frac{1}{P}}}{Lt} \left[0.5 \sum_{i=1}^P p_{1_i} \log p_{1_i} \right] - \underset{p^{2_i \rightarrow \frac{1}{Q}}}{Lt} \left[0.5 \sum_{i=1}^Q p_{2_i} \log p_{2_i} \right] \quad (9)$$

or when

Sym 1	0.48	0.48	0.48	0.48	0.48	0.48	0.52	1.00
	1	1	1	1	1	1	0	
Sym 2	0.18	0.18	0.18	0.18	0.21	0.31	0.48	
	000	000	000	000	01	00	1	
Sym 3	0.13	0.13	0.13	0.13	0.18	0.21		
	001	001	001	001	000	01		
Sym 4	0.07	0.07	0.08	0.13	0.13			
	0100	0100	011	010	001			
Sym 5	0.06	0.06	0.07	0.08				
	0101	0101	0100	011				
Sym 6	0.05	0.05	0.06					
	0110	0110	0101					
Sym 7	0.022	0.03						
	01110	0111						
Sym 8	0.008							
	01111							

Figure 2. Example showing Huffman coding for eight symbols: Sym 1 – Sym 8 with their probabilities shown in black and the codes constructed shown in red.

Sym 1	0.48	0.48	0.48	0.48	0.48
	0	0	0	0	0
Sym 2	0.18	0.18	0.18	0.18	0.18
	1	10	100	100	100
Sym 3	0.13	0.13	0.13	0.13	0.13
	1	10	101	101	101
Sym 4	0.07	0.07	0.07	0.07	0.07
	1	11	110	1100	1100
Sym 5	0.06	0.06	0.06	0.06	0.06
	1	11	110	1101	1101
Sym 6	0.05	0.05	0.05	0.05	0.05
	1	11	111	1110	1110
Sym 7	0.022	0.022	0.022	0.022	0.022
	1	11	111	1111	11110
Sym 8	0.008	0.008	0.008	0.008	0.008
	1	11	111	1111	11111

Figure 3. Example showing Shannon–Fano coding for eight symbols: Sym 1 – Sym 8 with their probabilities shown in black and the codes constructed shown in red.

Table 1. Proposed image coding scheme using Shannon entropy based partitioning with symbol partitions shown by different color and fonts.

Symbol	Shannon entropy	Code assigned	Symbol	Shannon entropy	Code assigned	Symbol	Shannon entropy	Code assigned
0.48		0	0.48		00	0.48		000
0.18	1.656	0	0.18	0	00	0.18		001
0.13	2.132	0	0.13	0.5816(max)	01	0.13		01
0.07	2.363(max)	1	0.07		10	0.07		1000
0.06	2.327	1	0.06	1.182	10	0.06	0.6891	1001
0.05	2.188	1	0.05	1.569	10	0.05	0.6902(max)	101
0.022	2.024	1	0.022	1.669(max)	11	0.022		110
0.008	1.519	1	0.008	1.314	11	0.008		111

$$H_k(S) \rightarrow \log P + \log Q. \tag{10}$$

_{max}

In the special case when the probability density function (pdf) of the image is more uniform than its partitions, we

proceed by partitioning the lowest probability from others and repeating the procedure on the remaining probabilities.

Theorem 3 The non-extensive entropy with Gaussian gain is more suitable than the extensive Shannon entropy for the proposed maximum entropy based partitioning of symbols

Table 2. Proposed image coding scheme using non-extensive entropy based partitioning with symbol partitions shown by different color and fonts.

Symbol	Non-extensive entropy	Code assigned	Symbol	Non-extensive entropy	Code assigned	Symbol	Non-extensive entropy	Code assigned
0.48		<i>0</i>	0.48		<i>00</i>	0.48		<i>00</i>
0.18	1.308	<i>0</i>	0.18	0.7358(max)	<i>01</i>	0.18		<i>010</i>
0.13	1.612	<i>0</i>	0.13	0.6817	<i>01</i>	0.13		<i>011</i>
0.07	1.724(max)	<i>1</i>	0.07		<i>10</i>	0.07		<i>1010</i>
0.06	1.718	<i>1</i>	0.06	1.2492	<i>10</i>	0.06	1.1427	<i>1011</i>
0.05	1.641	<i>1</i>	0.05	1.5528	<i>10</i>	0.05	1.1438(max)	<i>100</i>
0.022	1.559	<i>1</i>	0.022	1.5664(max)	<i>11</i>	0.022		<i>110</i>
0.008	1.258	<i>1</i>	0.008	1.2876	<i>11</i>	0.008		<i>111</i>

Table 3. Coding of the intensities of the *Happy* image of *S052* using various methods.

Sorted probabilities from histogram	Symbol	Huffman code	Shannon–Fano code	Proposed method using non-extensive entropy partitioning	Proposed method using Shannon entropy partitioning
0.2318	11	00	00	000	0000
0.1573	10	100	010	001	0001
0.1365	8	101	011	010	0010
0.1362	9	110	100	011	0011
0.1210	7	111	101	1000	0100
0.0966	6	011	110	1001	0101
0.0626	5	0100	1110	101	011
0.0268	4	01011	11110	1100	100
0.0163	3	010100	111110	1101	101
0.0129	2	0101000	1111110	1110	110
0.0020	1	0101001	1111111	1111	111

Table 4. Coding of the intensities of the *Surprise* image of *S052* using various methods.

Sorted probabilities from histogram	Symbol	Huffman code	Shannon–Fano code	Proposed method using non-extensive entropy partitioning	Proposed method using Shannon entropy partitioning
0.1984	11	10	00	000	0000
0.1874	10	11	010	001	0001
0.1700	9	001	011	010	001
0.1461	8	010	100	011	010
0.0833	7	0001	101	1000	011
0.0758	6	0110	1100	1001	100
0.0518	5	0111	1101	101	1010
0.0495	4	00000	1110	1100	1011
0.0243	3	000010	11110	1101	1100
0.0109	2	00000110	111110	1110	1101
0.0026	1	00000111	111111	1111	111

Proof The non-extensive entropy with Gaussian gain [47] is defined as

$$H(P) = \sum_{i=1}^n p_i e^{-p_i^2}. \quad (11)$$

It was originally introduced as an identifier of regular textures. It is adept at representing the sparse probabilities

in images. It was proved superior in performance to other forms of entropy such as: Renyi, Shannon, Pal & Pal, Tsallis entropies for texture categorization experiments that use entropy as a single feature. It is especially good at detecting regularity in sparsity due to its single-sided Gaussian information gain function $e^{-p_i^2}$ which highlights the contribution of the low probabilities lying in the bell of the Gaussian curve. This is well suited for our experiments where we are searching through improbable intensities for

Table 5. Coding of the intensities of the *Sad* image of *S052* using various methods.

Sorted probabilities from histogram	Symbol	Huffman code	Shannon–Fano code	Proposed method using non-extensive entropy partitioning	Proposed method using Shannon entropy partitioning
0.3091	11	00	00	000	000
0.1796	10	010	01	001	0010
0.1050	9	100	100	010	0011
0.1005	7	101	1010	011	0100
0.0965	8	110	1011	1000	0101
0.0933	6	111	110	1001	0110
0.0588	5	0110	1110	1010	0111
0.0290	4	01110	11110	1011	100
0.0153	3	011110	111110	1100	101
0.0108	2	0111110	1111110	1101	110
0.0022	1	0111111	1111111	111	111

Table 6. Coding of the intensities of the *Anger* image of *S052* using various methods.

Sorted probabilities from histogram	Symbol	Huffman code	Shannon–Fano code	Proposed method using non-extensive entropy partitioning	Proposed method using Shannon entropy partitioning
0.3918	11	1	00	000	000
0.1742	10	001	01	001	001
0.1089	9	011	100	0100	0100
0.1019	8	0000	101	0101	0101
0.0881	7	0001	110	011	011
0.0509	6	0101	1110	1000	1000
0.0428	5	01000	11110	1001	1001
0.0200	4	010011	111110	101	101
0.0121	3	0100100	1111110	1100	1100
0.0085	2	01001010	11111110	1101	1101
0.0007	1	01001011	11111111	111	111

the occurrence of uniform distributions. The sum of non-extensive entropies, in a subsequent work on crowd motion monitoring [48], indicated anomalies by positive spikes, a feature attributed to the non-additive nature of the non-extensive entropy with Gaussian gain. Additional experiments on the dynamic growth of the hidden layer in the Multi-Layer Perceptron neural network, identify the occurrence of anomalous training inputs in the process of incremental learning by positive entropy spikes [49]. In applications to image processing, uniform distributions within an image were localized by mean-shift clustering using the maximum non-extensive entropy of the mode-probabilities [50]. In this work, we expect to utilize the non-additivity of the sum of non-extensive entropies of the two partitions, to indicate to us which is the optimal partition by a positive entropy spike.

The maximum non-extensive entropy is achieved when

$$H_k(S)_{\max} = Lt_{p \rightarrow \frac{1}{p}} \left[0.5 \sum_{i=1}^P p1_i e^{-p1_i^2} \right] + Lt_{p \rightarrow \frac{1}{q}} \left[0.5 \sum_{i=1}^Q p2_i e^{-p2_i^2} \right] \quad (12)$$

or when

$$H_k(S)_{\max} \rightarrow e^{-\left(\frac{1}{p}\right)^2} + e^{-\left(\frac{1}{q}\right)^2}. \quad (13)$$

Comparing the two entropies, the logarithmic sum in (10) roughly takes into account the noise in both partitions and gives a partition that is globally optimum for the whole image. The exponential sum in (13) on the other hand gathers the uniform densities only in one partition and accumulates all other intensities including noise in the second partition, producing a positive spike for this combination. Any small addition to a uniform density that disturbs its uniformity is amplified exponentially and that combination is rejected as having lower entropy. Therefore the non-extensive entropy is more useful to us in our quest for localizing uniform densities accurately in the probability graph.

4.1 Automated computation of the threshold for code-length using maximum entropy partitioning

We extend the maximum entropy principle for automatically computing the threshold for code-length t_len , in order to create the binary mask explained in section 3.

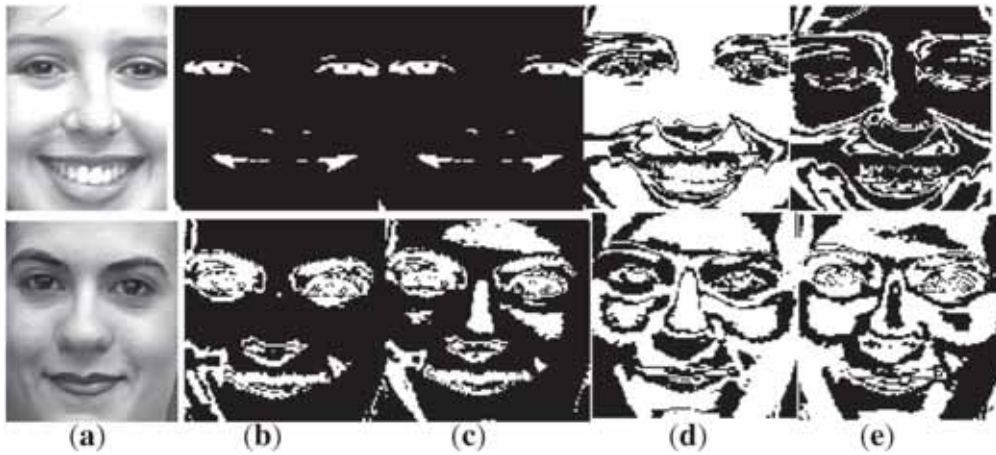


Figure 4. Intensities related to the *Happy* emotion (row-wise from top to bottom) shown in white in a black background, identified using different coding techniques, (shown here for two subjects *Top row*: Subject *S052*, *Bottom row*: Subject *S130*). (a) Cropped original images, (b) Huffman coding, (c) Shannon–Fano coding, (d) proposed method using maximum (Shannon) entropy partitions and (e) proposed method using maximum (non-extensive) entropy partitions.

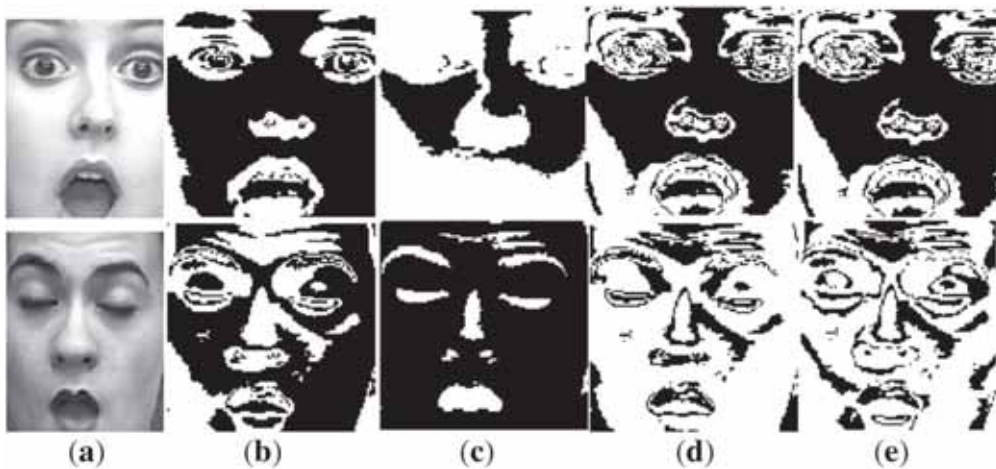


Figure 5. Intensities related to the *Surprise* emotion (row-wise from top to bottom) shown in white in a black background, identified using different coding techniques, (shown here for two subjects *Top row*: Subject *S052*, *Bottom row*: Subject *S130*). (a) Cropped original images, (b) Huffman coding, (c) Shannon–Fano coding, (d) proposed method using maximum (Shannon) entropy partitions and (e) proposed method using maximum (non-extensive) entropy partitions.

The steps for automated computation of the code-length threshold are detailed below:

1. For the K codes constructed by any image coding scheme for K symbols, enlist the code-lengths associated with each symbol/intensity
2. Map the code-length to each pixel position in the original image by closest bin matching with pixel intensity
3. Construct the probability histogram of the resulting code-length matrix
4. To find the optimal partition of symbols/intensities of this probability histogram do:
 5. Set the partition point $k = 1$
 6. Create partitions of symbols from Symbol 1 to k (Partition 1) and Symbol $k + 1$ to K (Partition 2)
 7. Find the non-extensive entropy with Gaussian gain within each partition using (11), after first normalizing the probabilities in each partition so that they sum to one
 8. Find the weighted sum of entropies of the two partitions using (7)
 9. Increment k by 1 and repeat steps 6–9 until $k = K - 1$
 10. Find the value of k at which maximum entropy is achieved using (8). Let this value be called opt

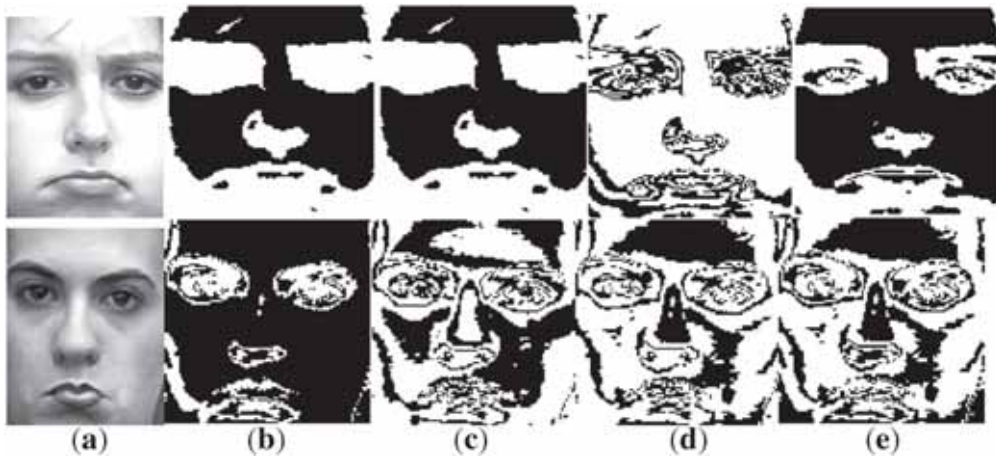


Figure 6. Intensities related to the *Sad* emotion (row-wise from top to bottom) shown in white in a black background, identified using different coding techniques, (shown here for two subjects *Top row*: Subject *S052*, *Bottom row*: Subject *S130*). (a) Cropped original images, (b) Huffman coding, (c) Shannon–Fano coding, (d) proposed method using maximum (Shannon) entropy partitions and (e) proposed method using maximum (non-extensive) entropy partitions.

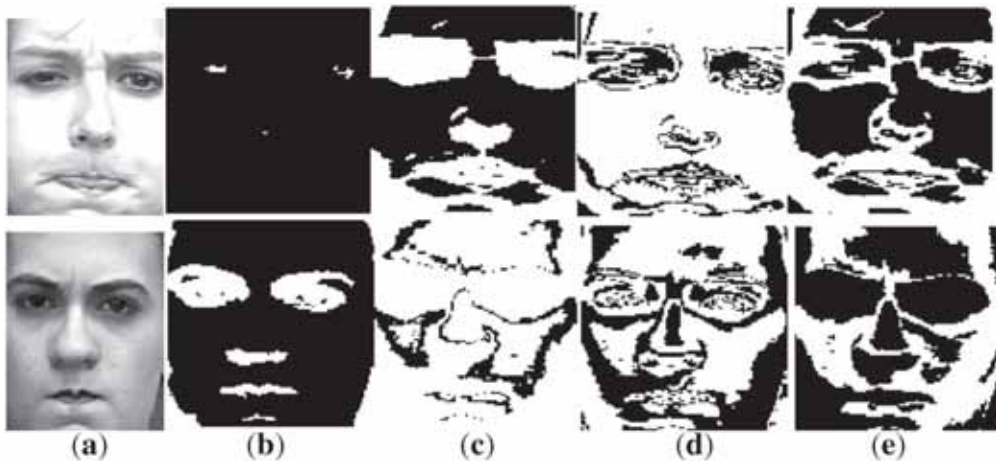


Figure 7. Intensities related to the *Anger* emotion (row-wise from top to bottom) shown in white in a black background, identified using different coding techniques, (shown here for two subjects *Top row*: Subject *S052*, *Bottom row*: Subject *S130*). (a) Cropped original images, (b) Huffman coding, (c) Shannon–Fano coding, (d) proposed method using maximum (Shannon) entropy partitions and (e) proposed method using maximum (non-extensive) entropy partitions.

Table 7. Evaluation of the image coding results in terms of the *average code-length* for various encoding techniques for different emotions of the two subjects *S052* and *S130* (proposed method shown in gray).

Encoding scheme	S052- Happy	S130- Happy	S052- Surprise	S130- Surprise	S052- Sad	S130- Sad	S052- Anger	S130- Anger
Huffman	2.9894	3.0750	3.0525	2.9813	2.9023	3.0627	2.6945	2.9939
Shannon–Fano	2.9894	3.0709	3.0665	2.9936	2.9179	3.0627	2.7228	3.0214
Non-extensive entropy based partitioning	3.2748	3.2489	3.2474	3.5295	3.3035	3.3401	2.8142	3.1360
Shannon entropy based partitioning	3.8810	3.4968	3.5202	3.5502	3.6359	3.7359	3.3258	3.4685

11. The threshold is decided as $t_{len} = opt$, i.e. all the (longer) code-lengths in the second partition indicate the symbols of interest (in our case, the image intensities that pertain to emotion)
12. Referring to the code-length matrix constructed in step 3, all the pixels satisfying the criterion of longer code-lengths in step 11 are indicated by 1 (white) in the binary mask and the remaining pixels by 0 (black)

Table 8. Evaluation of the binary masks created by measuring the *average information (Shannon entropy)* for various encoding techniques for different emotions of the two subjects *S052* and *S130* (proposed method shown in gray).

Encoding scheme	S052- Happy	S130- Happy	S052- Surprise	S130- Surprise	S052- Sad	S130- Sad	S052- Anger	S130- Anger
Huffman	0.2839	0.3351	0.6931	0.2186	0.2785	0.4392	0.2178	0.3733
Shannon–Fano	0.2839	0.4241	0.1065	0.4573	0.6859	0.5585	0.3364	0.5343
Non-extensive entropy based partitioning	0.5263	0.6031	0.4998	0.6909	0.5689	0.6743	0.5659	0.5343
Shannon entropy based partitioning	0.2839	0.6823	0.6723	0.6921	0.6921	0.4368	0.5973	0.6740

	Happy	Surprise	Sad	Anger
Happy	1	0	1	0
Surprise	0	1	0	1
Sad	0	1	1	0
Anger	1	0	0	1

(a)

	Happy	Surprise	Sad	Anger
Happy	1	1	0	0
Surprise	1	1	0	0
Sad	0	1	1	0
Anger	0	1	0	1

(b)

	Happy	Surprise	Sad	Anger
Happy	1	0	1	0
Surprise	0	1	1	0
Sad	0	1	1	0
Anger	0	0	1	1

(c)

	Happy	Surprise	Sad	Anger
Happy	2	0	0	0
Surprise	1	1	0	0
Sad	1	0	1	0
Anger	0	0	0	2

(d)

	Happy	Surprise	Sad	Anger
Happy	1	0	1	0
Surprise	0	1	1	0
Sad	1	0	1	0
Anger	0	0	1	1

(e)

	Happy	Surprise	Sad	Anger
Happy	1	0	0	1
Surprise	0	1	0	1
Sad	0	0	1	1
Anger	0	0	1	1

(f)

	Happy	Surprise	Sad	Anger
Happy	1	1	0	0
Surprise	1	1	0	0
Sad	0	0	1	1
Anger	0	1	0	1

(g)

	Happy	Surprise	Sad	Anger
Happy	1	0	1	0
Surprise	0	1	1	0
Sad	1	0	1	0
Anger	0	0	1	1

(h)

Figure 8. Fuzzy clustering of the DTF texture cues (results shown in terms of confusion matrices) extracted from the result images of different methods. (a) Huffman coding, (b) Shannon–Fano coding. (c) proposed method using maximum (Shannon) entropy partitions, (d) proposed method using maximum (non-extensive) entropy partitions methods, (e) cropped mouth image [53], (f) 3×3 grid partitioning of original cropped image [27], (g) cropped eye and mouth canny output [55], and (h) 6×7 weighted grid partitioning of original image [56].

5. Experimental results and discussions

The experiments are performed on MATLAB 7.9 platform on an Intel Pentium IV processor. The codes constructed using Huffman, Shannon–Fano coding and the proposed maximum entropy based partitioning (for both Shannon entropy and the non-extensive entropy with Gaussian gain) are compared for their coding performance and their ability to identify the gray-level facial intensities associated with different emotions. Figure 2 and figure 3 demonstrate the

Huffman and Shannon–Fano code construction for a synthetic sample set of eight symbols $K = 8$ arranged in the decreasing order of their probabilities. In both these coding schemes, we observe that all the lower rung probability intensity levels without exception are assigned longer codes, some of them being very long. These intensity levels when highlighted are of no consequence in a facial image since they are very low in number and may not indicate any useful information regarding the facial expression. The codes constructed by the proposed maximum entropy based

partitioning are shown for Shannon entropy in table 1 and for the non-extensive entropy with Gaussian gain in table 2, with all stages of coding shown including the maximum entropy point at each stage used to generate the partitions. It is noted from both tables 1 and 2 that the maximum entropy based partitioning assigns relatively longer codes to the indistinct intensity levels having probabilities 0.07 and 0.06 which are in the middle of the range of intensities, as opposed to the conventional encoding shown in figures 2 and 3. The non-extensive entropy based partitioning method is further able to distinguish the probability of 0.48 (highest probability) from the subset of {0.18, 0.13} as rightly belonging to two groups in table 2, while Shannon entropy is unable to discern that, and clubs {0.48, 0.18} together as seen from table 1. Compared to the Shannon entropy induced partitions, we observe that the non-extensive entropy partitions are closer to uniform distributions interpreting the maximum entropy point correctly. The proposed method overall focuses on the ‘improbable’ intensities that are difficult to distinguish from their peers in terms of their frequency of occurrence, rather than the lowest probabilities, since these are more likely to reflect on emotions due to their subtle and imprecise nature.

We next implement our image coding scheme on four different sample images of facial expressions (*Happy, Surprise, Sad, Anger*) of the subjects *S052* and *S130* of the Cohn-Kanade facial expression dataset [2]. The gray level intensities are contained in 11 bins of the histogram $K = 11$ for the ease of depiction of values in tables 3, 4, 5 and 6. We determine the gray-level intensities that are indicative of emotions by using the information-theoretic coding techniques: Huffman, Shannon–Fano and the proposed maximum entropy partitioning (using both the entropies) as per procedures outlined in sections 3 and 4. The threshold for code-length t_len is selected in an automatic manner by maximum (non-extensive) entropy partitioning, for all the four image coding schemes, as discussed in section 4. The binary facial mask with the improbable intensities (code-length $> t_len$) highlighted are visually depicted in figures 4, 5, 6 and 7 for all the four image coding methods, with the codes constructed shown in tables 3, 4, 5 and 6 respectively. While the Huffman and Shannon–Fano codes predictably form a tree or pyramid type structure, as observed from tables 3, 4, 5 and 6, the new image coding scheme though more compact and less code-length efficient assumes an unpredictable coding pattern with the middle intensities assigned longer codes. The proposed coding is sufficiently efficient in terms of average code length: $L = \sum_{j=1}^{K=11} p_j l_j$, $j = 1, 2, \dots, 11$, where l_j is the code-length of the j th symbol and p_j is the probability of occurrence of the j th symbol, though not nearly as code-length efficient as the Huffman and Shannon–Fano codes as shown in table 7. The non-extensive entropy based partitioning is more code-length efficient than Shannon entropy based partitioning as observed from the code-lengths in table 7 and comparable to that of

Huffman and Shannon–Fano code-lengths. The average information or Shannon entropy of the binary mask B containing the selected intensities pertaining to emotion is computed from the normalized probability distribution of the highlighted features as $-\sum_{h \in B} p_{hN} \log p_{hN}$, where $\sum_{h \in B} p_{hN} = 1$, and the results are summarized in table 8. These results indicate that the proposed entropy based partitioning contributes to higher average information than either Huffman or Shannon–Fano codes due to the code-assignment procedure which assigns longer codes to mid-level probabilities than to the lowest probabilities. The significance of the segmented information is thus higher for maximum entropy based partitioning as deduced from table 8. The binary masks containing the segmented improbable facial intensities in figures 4, 5, 6 and 7 show visually different results for the two subjects due to the contrast between the subtlety of the emotion being displayed by subject *S130* (bottom row of all figures) and the highly intense emotions displayed by the subject *S052* (top row of all figures). A texture analysis of the facial binary mask is required to measure the binary pattern to check for similarities between the masks for the same facial expression of the two subjects. The difference theoretic texture features or DTF texture cues [51] are computed from all the binary masks in figures 4, 5, 6 and 7 and clustered for the four emotions using the Fuzzy c-means (FCM) algorithm [52]. The DTF is a low dimensional feature set superior to LBP, LBPV, filter banks for affine-invariant texture classification. Also it gives high accuracies with distance classifiers and hence is suitable for distance-based clustering algorithms. The clustering results shown as confusion matrices in figure 8 indicate the highest clustering accuracy for the proposed non-extensive entropy based maximum entropy partitioning. It is the only method which clusters the *Happy* and *Anger* facial expressions of the two subjects correctly, a fact also corroborated from the segmentations in figures 4(e) and 7(e) which show the same pattern despite a change of intensity of emotions for the two subjects. We also compare our results with four benchmark methods of facial expression recognition namely, (i) cropped mouth image [53] (ii) 3×3 grid partitioning of original cropped image [54] (iii) cropped eye and mouth canny output [55], (iv) 6×7 weighted grid partitioning of original image [56]. The clustering results of any comparison method do not show the two subjects clubbed into one cluster for any one emotion due to the highly different nature of emotion intensity portrayed by the two subjects for all emotions.

While the Huffman and Shannon–Fano coding do not show any emotionally significant intensities in figures 4–7 (b, c) the proposed entropy based partitioning especially using non-extensive entropy correctly identifies all traces of emotions. The segmented face highlighting the facial features shows distinct results in all four cases of facial expressions and could be used for categorizing emotions which forms our future work.

6. conclusions

In this paper, we propose a new variable length coding especially for identifying the improbable image intensities that help in decoding facial expressions. The new image coding scheme assigns longer codes to ‘improbable and indistinct’ intensities which are more likely to represent emotion in a facial image than the intensities with distinct probability values (very low and very high). The resultant coding though not as code-length efficient as Huffman or Shannon–Fano, identifies the vulnerable facial intensities much better. These intensities are used as a binary pattern for emotion categorization with successful results as proved by the experimentation on sample facial expressions from a benchmark dataset and comparison with state of the art methods. The non-extensive entropy with Gaussian gain is proved superior to the extensive Shannon entropy for the proposed maximum entropy based partitioning.

References

- [1] De la Torre Fernando and Jeffrey F Cohn 2011 Facial expression analysis. In: *Visual analysis of humans*, pp. 377–409. Springer London
- [2] Kanade T, Cohn J F and Tian Y 2000 Comprehensive database for facial expression analysis. In: *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pp. 46–53
- [3] Zwakhalen Sandra M G, Jan P H Hamers, Huda H Abu-Saad and Martijn P F Berger 2006 Pain in elderly people with severe dementia: a systematic review of behavioural pain assessment tools. *BMC Geriatr.* 6(1): 3a
- [4.] IZARD Carroll E et al 1983 Changes in facial expressions of 2-to 19-month-old infants following acute pain. *Dev. Psychol.* 19(3): 418
- [5] Edwards Jane, Henry J Jackson and Philippa E Pattison 2002 Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review. *Clin. Psychol. Rev.* 22(6): 789–832
- [6] Prkachin Kenneth M 1992 The consistency of facial expressions of pain: a comparison across modalities. *Pain* 51(3): 297–306
- [7] Schwartz Gary E, Paul L Fair, Patricia Salt, Michel R Mandel and Gerald L Klerman 1976 Facial expression and imagery in depression: an electromyographic study. *Psychosomatic Med.* 38(5): 337–347
- [8] Picard Rosalind W 2000 Toward computers that recognize and respond to user emotion. *IBM Syst. J.* 39(3.4): 705–719
- [9] Bartlett Marian Stewart, Gwen Littlewort, Ian Fasel and Javier R Movellan 2003 Real time face detection and facial expression recognition: development and applications to human computer interaction. In: *CVPRW'03, Conference on Computer Vision and Pattern Recognition Workshop 2003*, vol. 5, pp. 53–53
- [10] Fong Terrence, Illah Nourbakhsh and Kerstin Dautenhahn 2003 A survey of socially interactive robots. *Robotics Autom. Syst.* 42(3): 143–166
- [11] Scheirer Jocelyn, Raul Fernandez and Rosalind W Picard 1999 Expression glasses: a wearable device for facial expression recognition. In: *CHI'99 Extended Abstracts on Human Factors in Computing Systems*, pp. 262–263. ACM
- [12] Lyons Michael J and Nobuji Tetsutani 2001 Facing the music: a facial action controlled musical interface. In: *CHI'01 extended abstracts on Human factors in computing systems*, pp. 309–310. ACM
- [13] Den Uyl M J and Van Kuilenburg H 2005 The FaceReader: online facial expression recognition. In: *Proceedings of Measuring Behavior*, vol. 30
- [14] Oliver Nuria, Alex Pentland and François Bérard 2000 LAFTER: a real-time face and lips tracker with facial expression recognition. *Pattern Recognit.* 33(8): 1369–1382
- [15] Boucher Jerry D and Gary E Carlson 1980 Recognition of facial expression in three cultures. *J. Cross-Cultural Psychol.* 11(3): 263–280
- [16] Tranel Daniel, Antonio R Damasio and Hanna Damasio 1988 Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology* 38(5): 690–690
- [17] Cohen Ira, Nicu Sebe, Ashutosh Garg, Lawrence S Chen and Thomas S Huang 2003 Facial expression recognition from video sequences: temporal and static modeling. *Comput. Vis. Image Understand.* 91(1): 160–187
- [18] Hills Peter J and Michael Pake J 2013 Eye-tracking the own-race bias in face recognition: revealing the perceptual and socio-cognitive mechanisms. *Cognition* 129(3): 586–597
- [19] Lienhart Rainer and Jochen Maydt 2002 An extended set of haar-like features for rapid object detection. In: *Proceedings of International Conference on Image Processing. 2002.* vol. 1. IEEE
- [20] Magalhães Filipe et al 2013 Compressive sensing based face detection without explicit image reconstruction using support vector machines. *Image analysis and recognition.* Springer Berlin Heidelberg, pp. 758–765
- [21] Reisfeld Daniel and Yehezkel Yeshurun 1992 Robust detection of facial features by generalized symmetry. In: *International Conference on Pattern Recognition*, pp. 117–117. IEEE Computer Society Press
- [22] Roth Dan, Ming-Hsuan Yang and Narendra Ahuja 2000 A SNOw-based face detector. *Urbana* 51: 61801
- [23] Susan Seba and Pooja Kadyan 2013 A supervised fuzzy eye pair detection algorithm. In: *5th International Conference on Computational Intelligence and Communication Networks (CICN), 2013*, pp. 306–310. IEEE
- [24] Zhang Lun et al 2007 Face detection based on multi-block lbp representation. *Advances in biometrics.* Springer Berlin Heidelberg, pp. 11–18
- [25] Viola Paul and Michael J Jones 2004 Robust real-time face detection. *Int. J. Comput. Vis.* 57(2): 137–154
- [26] Ojala T, Pietikaenen M, Maenepae T 2002 Multi-resolution gray scale and rotation invariant texture classification with LBP. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(7): 971–987
- [27] Zhao G and Matti Pietikäinen 2007 Dynamic texture recognition using local binary patterns with application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* 29(6): 915–928
- [28] Hernández Benjamín, Gustavo Olague, Riad Hammoud, Leonardo Trujillo and Eva Romero 2007 Visual learning of

- texture descriptors for facial expression recognition in thermal imagery. *Comput. Vis. Image Understand.* 106(2): 258–269
- [29] Liao Shu, Wei Fan, Albert CS Chung and Dit-Yan Yeung 2006 Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features. In: *IEEE International Conference on Image Processing, 2006* pp. 665–668
- [30] Lyons Michael, Shigeru Akamatsu, Miyuki Kamachi and Jiro Gyoba 1998 Coding facial expressions with gabor wavelets. In: *Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998*, pp. 200–205
- [31] Manjunath Bangalore S and Wei-Ying Ma 1996 Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.* 18(8): 837–842
- [32] Seba Susan and Gitin Kakkar 2015 Decoding facial expressions using a new normalized similarity index. In *2015 Annual IEEE India Conference (INDICON)*, pp. 1–6. IEEE
- [33] Milborrow Stephen and Fred Nicolls 2008 Locating facial features with an extended active shape model. In: *Computer Vision–ECCV 2008*, pp. 504–513. Springer Berlin Heidelberg
- [34] Huffman David A 1952 A method for the construction of minimum redundancy codes. *Proc. IRE* 40(9): 1098–1101
- [35] Fano Robert M and Wintringham W T 1961 Transmission of information. *Phys. Today* 14: 56
- [36] Shannon C E 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* 27: 379–423
- [37] Chanda Bhabatosh and Dwijesh Dutta Majumder 2004 *Digital image processing and analysis*. PHI Learning Pvt. Ltd
- [38] Jeon Byeung-woo, Jechang Jeong and Ju-ha Park 1996 *Apparatus for variable-length coding and variable-length-decoding using a plurality of Huffman coding tables*. U.S. Patent 5,528,628, issued June 18
- [39] Takishima Yasuhiro, Masahiro Wada, and Hitomi Murakami 1995 Reversible variable length codes. *IEEE Trans. Commun.* 43(234): 158–162
- [40] Zhou Jiantao, Zhiqin Liang, Yan Chen and Oscar C Au 2007 Security analysis of multimedia encryption schemes based on multiple Huffman table. *Signal Process. Lett. IEEE* 14(3): 201–204
- [41] Nag Amitava, Sushanta Biswas, Debasree Sarkar and Partha Pratim Sarkar 2011 A novel technique for image steganography based on DWT and Huffman encoding. *Int. J. Comput. Sci. Security* 4(6): 497–610
- [42] Al-Laham Mohammed and Ibrahiem MM El Emary 2007 Comparative study between various algorithms of data compression techniques. *Int. J. Comput. Sci. Netw. Security* 7(4): 281
- [43] Christiansen Mark M et al 2013 Brute force searching, the typical set and Guesswork. *IEEE International Symposium on Information Theory Proceedings (ISIT), 2013*
- [44] Xiaohua Huang, Guoying Zhao, Xiaopeng Hong, Wenming Zheng, and Matti Pietikäinen 2016 Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing* 175: 564–578
- [45] Seba Susan and Roni Chakre 2016 3D-difference theoretic texture features for dynamic face recognition. In *2016 International conference on computational techniques in information and communication technologies (ICCTICT)*, pp. 227–232. IEEE
- [46] Pal Nikhil R and Sankar K Pal 1989 Entropic thresholding. *Signal Process.* 16(2): 97–108
- [47] Susan Seba and Madasu Hanmandlu 2013a A non-extensive entropy feature and its application to texture classification. *Neurocomputing* 120: 214–225
- [48] Seba Susan and Madasu Hanmandlu 2015 Unsupervised detection of nonlinearity in motion using weighted average of non-extensive entropies. *SIViP* 9(3): 511–525
- [49] Susan S and Dwivedi M 2014 Dynamic growth of hidden-layer neurons using the non-extensive entropy. In *2014 Fourth international conference on communication systems and network technologies (CSNT)*, 7 Apr 2014, pp. 491–495. IEEE
- [50] Seba Susan and Ankit Kumar 2016 Auto-segmentation using mean-shift and entropy analysis. In: *2016 3rd International conference on computing for sustainable global development (INDIACom)*, pp. 292–296. IEEE
- [51] Susan Seba and Madasu Hanmandlu 2013 cDifference theoretic feature set for scale-, illumination-and rotation-invariant texture classification. *IET Image Process.* 7(8): 725–732
- [52] Bezdek J C 1981 *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press
- [53] Raheja Jagdish Lal, Radhey Shyam, Jatin Gupta, Umesh Kumar and Bhanu Prasad P 2010 Facial gesture identification using lip contours. In: *Second International Conference on Machine Learning and Computing (ICMLC), 2010*, pp. 3–7. IEEE
- [54] Engin Avci 2007 An expert system based on wavelet neural network adaptive norm entropy for scale invariant texture classification. *Expert Syst. Appl.* 32: 919–926
- [55] Filko Damir and Goran Martinovic 2013 Emotion recognition system by a neural network based facial expression analysis. *Automatika–J. Control Meas. Electron. Comput. Commun.* 54(2): 263–272
- [56] Shan Caifeng, Shaogang Gong and Peter W McOwan 2009 Facial expression recognition based on local binary patterns: a comprehensive study. *Image Vis. Comput.* 27(6): 803–816