

Framework for reliable, real-time facial expression recognition for low resolution images

Rizwan Ahmed Khan, Alexandre Meyer, Hubert Konik, Saïda Bouakaz

► **To cite this version:**

Rizwan Ahmed Khan, Alexandre Meyer, Hubert Konik, Saïda Bouakaz. Framework for reliable, real-time facial expression recognition for low resolution images. Pattern Recognition Letters, Elsevier, 2013, 34 (10), pp.1159-1168. 10.1016/j.patrec.2013.03.022 . hal-00817293

HAL Id: hal-00817293

<https://hal.archives-ouvertes.fr/hal-00817293>

Submitted on 25 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Framework for reliable, real-time facial expression recognition for low resolution images

Rizwan Ahmed Khan^{a,b,*}, Alexandre Meyer^{a,b}, Hubert Konik^{a,c}, Saida Bouakaz^{a,b}

^a *Université de Lyon, CNRS*

^b *Université Lyon 1, LIRIS, UMR5205, F-69622, France*

^c *Université Jean Monnet, Laboratoire Hubert Curien, UMR5516, 42000 Saint-Etienne, France*

Abstract

Automatic recognition of facial expressions is a challenging problem specially for low spatial resolution facial images. It has many potential applications in human-computer interactions, social robots, deceit detection, interactive video and behavior monitoring. In this study we present a novel framework that can recognize facial expressions very efficiently and with high accuracy even for very low resolution facial images. The proposed framework is memory and time efficient as it extracts texture features in a pyramidal fashion only from the perceptual salient regions of the face. We tested the framework on different databases, which includes Cohn-Kanade (CK+) posed facial expression database, spontaneous expressions of MMI facial expression database and FG-NET facial expressions and emotions database (FEED)

*corresponding author for this article is Rizwan A Khan, Phone: (+33) (0)4 72 43 19 75, Cell: (+33) (0) 63 86 54 278, Fax: (+33) (0)4 72 43 15 36

Email addresses: Rizwan-ahmed.khan@liris.cnrs.fr (Rizwan Ahmed Khan), Alexandre.meyer@liris.cnrs.fr (Alexandre Meyer), Hubert.Konik@univ-st-etienne.fr (Hubert Konik), Saida.bouakaz@liris.cnrs.fr (Saida Bouakaz)

and obtained very good results. Moreover, our proposed framework exceeds state-of-the-art methods for expression recognition on low resolution images.

Keywords:

Facial expression recognition, Low Resolution Images, Local Binary Pattern, Image pyramid, Salient facial regions

1. Introduction

Communication in any form i.e. verbal or non-verbal is vital to complete various routine tasks and plays a significant role in daily life. Facial expression is the most effective form of non-verbal communication and it provides a clue about emotional state, mindset and intention [7]. Human visual system (HVS) decodes and analyzes facial expressions in real time despite having limited neural resources. As an explanation for such performance, it has been proposed that only some visual inputs are selected by considering “salient regions” [36], where “salient” means most noticeable or most important.

For computer vision community it is a difficult task to automatically recognize facial expressions in real-time with high reliability. Variability in pose, illumination and the way people show expressions across cultures are some of the parameters that make this task difficult. Low resolution input images makes this task even harder. Smart meeting, video conferencing and visual surveillance are some of the real world applications that require facial expression recognition system that works adequately on low resolution images. Another problem that hinders the development of such system for real world application is the lack of databases with natural displays of expres-

20 sions [27]. There are number of publicly available benchmark databases with
21 posed displays of the six basic emotions [6] exist but there is no equivalent of
22 this for spontaneous basic emotions. While, It has been proved that Spon-
23 taneous facial expressions differ substantially from posed expressions [2]. In
24 this work, we propose a facial expression recognition system that caters for
25 illumination changes and works equally well for low resolution as well as for
26 good quality / high resolution images. We have tested our proposed system
27 on spontaneous facial expressions as well and recorded encouraging results.

28 We propose a novel descriptor for facial features analysis, Pyramid of
29 Local Binary Pattern (PLBP) (refer Section 3). PLBP is a spatial represen-
30 tation of local binary pattern (LBP) [19] and it represents stimuli by its local
31 texture (LBP) and the spatial layout of the texture. We combined pyramidal
32 approach with LBP descriptor for facial feature analysis as this approach has
33 already been proved to be very effective in a variety of image processing tasks
34 [10]. Thus, the proposed descriptor is a simple and computationally efficient
35 extension of LBP image representation, and it shows significantly improved
36 performance for facial expression recognition tasks for low resolution images.
37 We base our framework for automatic facial expression recognition (FER)
38 on human visual system (HVS) (refer Section 5), so it extracts PLBP fea-
39 tures only from the salient regions of the face. To determine which facial
40 region(s) is the most important or salient according to HVS, we conducted
41 a psycho-visual experiment using an eye-tracker (refer Section 4). We con-
42 sidered six universal facial expressions for psycho-visual experimental study
43 as these expressions are proved to be consistent across cultures [6]. These
44 six expressions are anger, disgust, fear, happiness, sadness and surprise. The

45 novelty of the proposed framework is that, it is illumination invariant, reli-
 46 able on low resolution images and works adequately for both i.e. posed and
 47 spontaneous expressions.

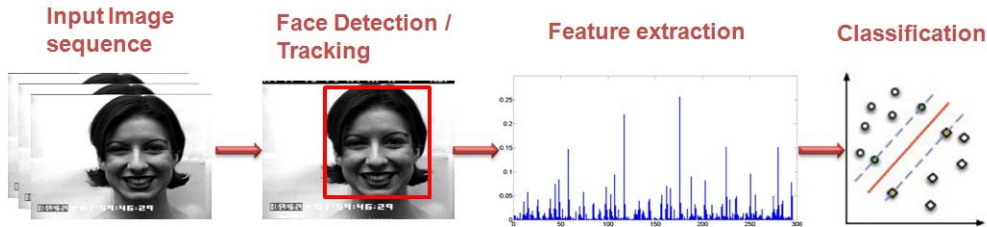


Figure 1: Basic structure of facial expression recognition system pipeline.

48 Generally, facial expression recognition system consists of three steps:
 49 face detection, feature extraction and expression classification. The same
 50 has been shown in Figure 1. In our framework we tracked face / salient
 51 facial regions using Viola-Jones object detection algorithm [30] as it is the
 52 most cited and considered the fastest and most accurate pattern recognition
 53 method for face detection [13]. The second step in the framework is feature
 54 extraction, which is the area where this study contributes. The optimal fea-
 55 tures should minimize within-class variations of expressions, while maximize
 56 between class variations. If inadequate features are used, even the best clas-
 57 sifier could fail to achieve accurate recognition [25]. Section 3 presents the
 58 novel method for facial features extraction which is based on human visual
 59 system (HVS). To study and understand HVS we performed psycho-visual
 60 experiment. Psycho-visual experimental study is briefly described in Section
 61 4. Expression classification or recognition is the last step in the pipeline. In
 62 literature two different ways are prevalent to recognize expressions i.e. direct

63 recognition of prototypic expressions or recognition of expressions through
64 facial action coding system (FACS) action units (AUs) [8]. In our proposed
65 framework, which is described in Section 5 we directly classify six universal
66 prototypic expressions [6]. The performance of the framework is evaluated
67 for five different classifiers (from different families i.e. classification Tree,
68 Instance Based Learning, SVM etc) and results are presented in Section 6.
69 Next section presents the brief literature review for facial features extraction
70 methods.

71 2. Related work

72 In the literature, various methods are employed to extract facial features
73 and these methods can be categorized either as appearance-based methods
74 or geometric feature-based methods.

75 **Appearance-based methods.** One of the widely studied method to
76 extract appearance information is based on Gabor wavelets [15, 26, 5]. Gen-
77 erally, the drawback of using Gabor filters is that it produces extremely
78 large number of features and it is both time and memory intensive to con-
79 volve face images with a bank of Gabor filters to extract multi-scale and
80 multi-orientational coefficients. Another promising approach to extract ap-
81 pearance information is by using Haar-like features, see Yang et al. [33].
82 Recently, texture descriptors and classification methods i.e. Local Binary
83 Pattern (LBP) [19] and Local Phase Quantization (LPQ) [21] are also stud-
84 ied to extract appearance-based facial features. Zhao et al. [35] proposed to
85 model texture using volume local binary patterns (VLBP) an extension to
86 LBP, for expression recognition.

87 **Geometric-based methods.** Geometric feature-based methods [34, 22,
88 28, 1] extracts shapes and locations of facial components information to form
89 a feature vector. The problem with using geometric feature-based methods
90 is that they usually require accurate and reliable facial feature detection
91 and tracking which is difficult to achieve in many real world applications
92 where illumination changes with time and images are recorded in very low
93 resolution.

94 Generally, we have found that all the reviewed methods for automatic
95 facial expression recognition are computationally expensive and usually re-
96 quires dimensionally large feature vector to complete the task. This explains
97 their inability for real-time applications. Secondly, in literature, very few
98 studies exist that tackles the issue of expressions recognition from low res-
99 olution images, this adds to lack of applicability of expression recognition
100 system for real world applications. Lastly, all of the reviewed methods, spend
101 computational time on whole face image or divides the facial image based
102 on some mathematical or geometrical heuristic for features extraction. We
103 argue that the task of expression analysis and recognition could be done in
104 more conducive manner, if only some regions are selected for further process-
105 ing (i.e. salient regions) as it happens in human visual system. Thus, our
106 contributions in this study are:

- 107 1. We propose a novel descriptor for facial expression analysis i.e. Pyra-
108 mid of Local Binary Pattern (PLBP), which outperforms state-of-the-
109 art methods for expression recognition on low resolution images (spa-
110 tially degraded images). It also performs better than other state-of-
111 the-art methods for good resolution images (with no degradation).

112 2. As the proposed framework is based on human visual system it algorithmically
113 processes only salient facial regions which reduces the length of
114 feature vector. This reduction in feature vector length makes the proposed
115 framework suitable for real-time applications due to minimized
116 computational complexity.

117 3. Pyramid of Local Binary Pattern

118 The proposed framework creates a novel feature space by extracting proposed
119 PLBP (pyramid of local binary pattern) features only from the visually
120 salient facial region (see Section 4 for psycho-visual experiment). PLBP is a
121 *pyramidal-based spatial* representation of local binary pattern (LBP) descriptor.
122 PLBP represents stimuli by their local texture (LBP) and the spatial
123 layout of the texture. The spatial layout is acquired by tiling the image
124 into regions at multiple resolutions. The idea is illustrated in Figure 2. If
125 only the coarsest level is used, then the descriptor reduces to a global LBP
126 histogram. Comparing to the multi-resolution LBP of Ojala et al.[20], our
127 descriptor selects samples in a more uniformly distributed manner, whereas
128 Ojala’s LBP takes samples centered around a point leading to missing some
129 information in the case of face (which is different than a repetitive texture).

130 LBP features were initially proposed for texture analysis [19], but recently
131 they have been successfully used for facial expression analysis [35, 25]. The
132 most important property of LBP features are their tolerance against illumination
133 changes and their computational simplicity [18, 19, 20]. The operator
134 labels the pixels of an image by thresholding the 3 x 3 neighbourhood of
135 each pixel with the center value and considering the result as a binary num-

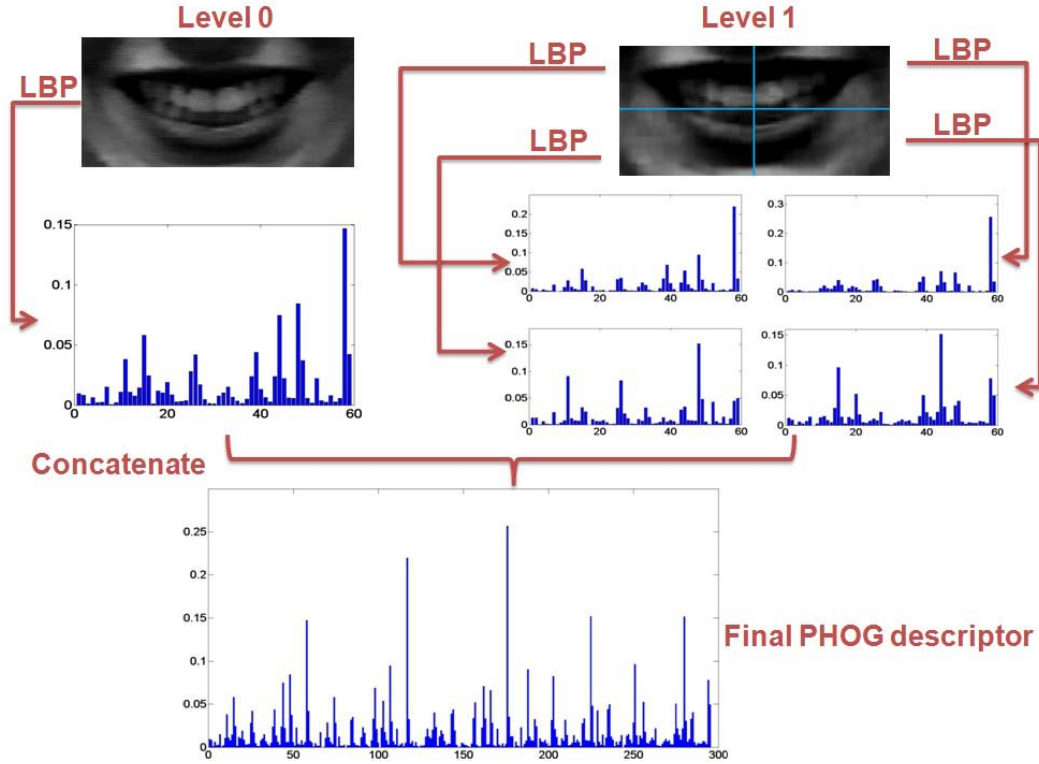


Figure 2: Pyramid of Local Binary Pattern. First row: stimuli at two different pyramid levels, second row: histograms of LBP at two respective levels, third row: final descriptor.

136 ber. Then the histogram of the labels can be used as a texture descriptor.
 137 Formally, LBP operator takes the form:

$$LBP(x_c, y_c) = \sum_{n=0}^7 s(i_n - i_c) 2^n \quad (1)$$

138 where in this case n runs over the 8 neighbours of the central pixel c ,
 139 i_c and i_n are the grey level values at c and n and $s(u)$ is 1 if $u \geq 0$ or 0
 140 otherwise.

141 Later, the LBP operator is extended to use neighborhood of different sizes

142 [20] as the original operator uses 3 x 3 neighbourhood. Using circular neigh-
 143 borhoods and bilinearly interpolating the pixel values allow any radius and
 144 number of pixels in the neighborhood. The LBP operator with P sampling
 145 points on a circular neighborhood of radius R is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \quad (2)$$

146 Another extension to the original operator is the definition of *uniform*
 147 *patterns*, which can be used to reduce the length of the feature vector and
 148 implement a simple rotation-invariant descriptor. A local binary pattern is
 149 called uniform if the binary pattern contains at most two bitwise transitions
 150 from 0 to 1 or vice versa when the bit pattern is traversed circularly. Accu-
 151 mulating the patterns which have more than 2 transitions into a single bin
 152 yields an LBP operator, denoted $LBP_{P,R}^u$. patterns. These binary patterns
 153 can be used to represent texture primitives such as spot, flat area, edge and
 154 corner.

155 We extend LBP operator so that the stimuli can be represented by its
 156 local texture and the spatial layout of the texture. We call this extended
 157 LBP operator as pyramid of local binary pattern or PLBP. PLBP creates
 158 the spatial pyramid by dividing the stimuli into finer spatial sub-regions by
 159 iteratively doubling the number of divisions in each dimension. It can be
 160 observed from the Figure 2 that the pyramid at level l has 2^l sub-regions
 161 along each dimension $(R_0, \dots R_m)$. Histograms of LBP features at the same
 162 levels are concatenated. Then, their concatenation at different pyramid levels

163 gives final PHOG descriptor (as shown in Figure 2). It can be defined as:

$$H_{i,j} = \sum_l \sum_{xy} I\{f_l(x,y) = i\} I\{(x,y) \in R_l\} \quad (3)$$

164 where $l = 0 \dots m - 1$, $i = 0 \dots n - 1$. n is the number of different labels
 165 produced by the LBP operator and

$$I(A) = \begin{cases} 1 & \text{if A is true ,} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

166 While, the dimensionality of the descriptor can be calculated by:

$$N \sum_l 4^l \quad (5)$$

167 Where, in our experiment (see Section 6) $l=1$ and $N= 59$ as we created
 168 pyramid up to level 1 and extracted 59 LBP features using $LBP_{8,2}^{u2}$ operator,
 169 which denotes a uniform LBP operator with 8 sampling pixels in a local
 170 neighborhood region of radius 2. This pattern reduces the histogram from
 171 256 to 59 bins. In our experiment we obtained 295 dimensional feature vector
 172 from one facial region i.e. mouth region (59 dimensions / sub-region), since
 173 we executed the experiment with the pyramid of level 1 (the same is shown
 174 in Figure 2).

175 3.1. Novelty of the proposed descriptor

176 There exist some methods in literature that uses Pyramid of LBP for
 177 different applications and they look similar to our proposed descriptor i.e.
 178 [32, 9, 17]. Our proposition is novel and there exist differences in the method-
 179 ology that creates differences in the extracted information. Method for face

180 recognition proposed in [32] creates pyramids before applying LBP operator
181 by down sampling original image i.e. scale-space representation, whereas we
182 propose to create the spatial pyramid by dividing the stimuli into finer spatial
183 sub-regions by iteratively doubling the number of divisions in each dimen-
184 sion. Secondly, our approach reduces memory consumption (do not requires
185 to store same image in different resolutions) and is computationally more
186 efficient. Guo et al. [9] proposed approach for face and palmprint recogni-
187 tion based on multiscale LBP. Their proposed method seems similar to our
188 method for expression recognition but how multiscale analysis is achieved de-
189 viates our approach. Approach proposed in [9] achieves multiscale analysis
190 using different values of P and R , where $LBP(P, R)$ denotes a neighborhood
191 of P equally spaced sampling points on a circle of radius R (discussed ear-
192 lier). Same approach has been applied by Moore et al. [17] for facial features
193 analysis. Generally the drawback of using such approach is that it increases
194 the size of the feature histogram and increases the computational cost. [17]
195 reports dimensionality of feature vector as high as 30,208 for multiscale face
196 expression analysis as compared to our proposition which creates 590 dimen-
197 sional feature vector (see Section 5) for the same task. We achieve the task
198 of multiscale analysis much more efficiently than any other earlier proposed
199 methods. By the virtue of efficient multiscale analysis our framework can
200 be used for real time applications (see Table 1 for the time and memory
201 consumption comparison) which is not the case with other methods.

202 As mentioned earlier, we base our framework for facial expression recog-
203 nition on human visual system (HVS), which selects only few facial regions
204 (salient) to extract information. In order to determine the saliency of facial

205 region(s) for a particular expression, we conducted psycho-visual experiment
206 with the help of an eye-tracker. Next section briefly explains the psycho-
207 visual experimental study.

208 **4. Psycho-Visual experiment**

209 The aim of our experiment was to record the eye movement data of human
210 observers in free viewing conditions. The data were analyzed in order to find
211 which components of face are salient for specific displayed expression.

212 *4.1. Participants, apparatus and stimuli*

213 Eye movements of fifteen human observers were recorded using video
214 based eye-tracker (EyelinkII system, SR Research), as the subjects watched
215 the collection of 54 videos selected from the extended Cohn-Kanade (CK+)
216 database [16], showing one of the six universal facial expressions [6]. Ob-
217 servers include both male and female aging from 20 to 45 years with normal
218 or corrected to normal vision. All the observers were naïve to the purpose of
219 an experiment.

220 *4.2. Eye movement recording*

221 Eye position was tracked at 500 Hz with an average noise less than 0.01° .
222 Head mounted eye-tracker allows flexibility to perform the experiment in free
223 viewing conditions as the system is designed to compensate for small head
224 movements.

225 *4.3. Psycho-Visual experiment Results*

226 In order to statistically quantify which region is perceptually more attrac-
227 tive for specific expression, we have calculated the average percentage of trial



Figure 3: Summary of the facial regions that emerged as salient for six universal expressions. Salient regions are mentioned according to their importance (for example facial expression of "fear" has two salient regions but mouth is the most important region according to HVS).

228 time observers have fixated their gazes at specific region(s) in a particular
 229 time period. As the stimuli used for the experiment is dynamic i.e. video
 230 sequences, it would have been incorrect to average all the fixations recorded
 231 during trial time (run length of the video) for the data analysis as this could
 232 lead to biased analysis of the data. To meaningfully observe and analyze the
 233 gaze trend across one video sequence we have divided each video sequence in
 234 three mutually exclusive time periods. The first time period correspond to
 235 initial frames of the video sequence i.e. neutral face. The last time period en-

236 capsulates the frames where the expression is shown with full intensity (apex
237 frames). The second time period is a encapsulation of the frames which has
238 a transition of facial expression i.e. transition from neutral face to the be-
239 ginning of the desired expression (i.e neutral to the onset of the expression).
240 Then the fixations recorded for a particular time period are averaged across
241 fifteen observers. For drawing the conclusions we considered second and third
242 time periods as they have the most significant information in terms of specific
243 displayed expression. Conclusions drawn are summarized in Figure 3. Refer
244 [11] for the detailed explanation of the psycho-visual experimental study.

245 5. Expression Recognition Framework

246 Feature selection along with the region(s) from where these features are
247 going to be extracted is one of the most important step to recognize ex-
248 pressions. As the proposed framework draws its inspiration from the human
249 visual system (HVS), it extracts proposed features i.e. PLBP, only from the
250 perceptual salient facial region(s) which were determined through Psycho-
251 Visual experiment. Schematic overview of the framework is presented in
252 Figure. 4. Steps of the proposed framework are as follows:

- 253 1. First, the framework extracts PLBP features from the mouth region,
254 feature vector of 295 dimensions (f_1, \dots, f_{295}). The classification
255 (“Classifier-a” in the Figure. 4) is carried out on the basis of extracted
256 features in order to make two groups of facial expressions. First group
257 comprises of those expressions that has one perceptual salient region i.e.
258 happiness, sadness and surprise while the second group is composed of
259 those expressions that have two or more perceptual salient regions i.e.

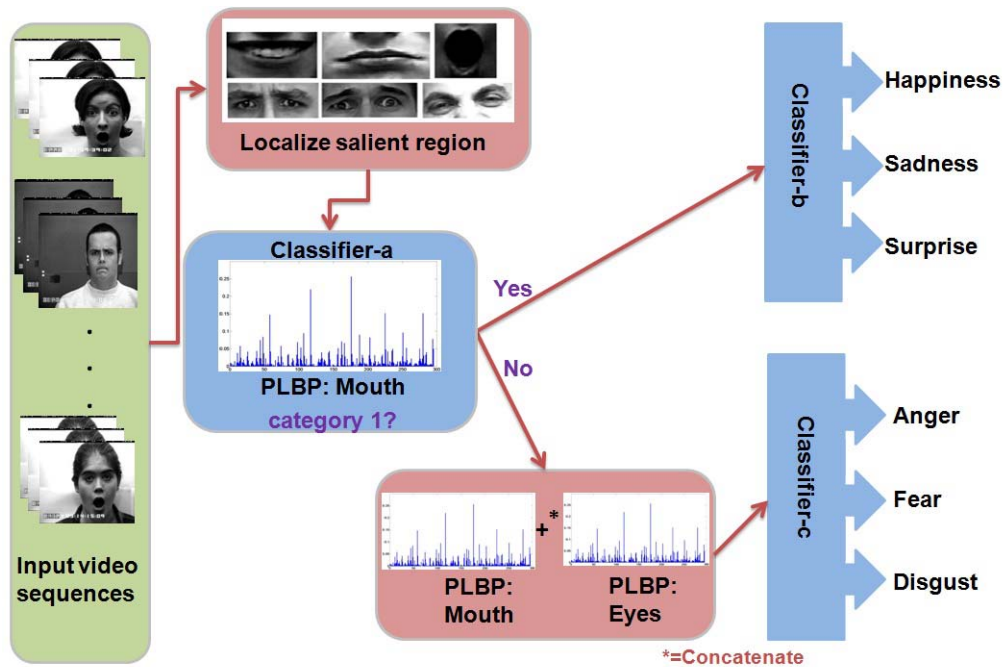


Figure 4: Schematic overview of the framework.

- 260 anger, fear and disgust (see Section 4.3). Purpose of making two groups
 261 of expressions is to reduce feature extraction computational time.
- 262 2. If the stimuli is classified in the first group, then it is classified either
 263 as happiness, sadness or surprise by the “Classifier-b” using already
 264 extracted PLBP features from the mouth region.
- 265 3. If the stimuli is classified in the second group, then the framework ex-
 266 tracts PLBP features from the eyes region (it is worth mentioning here
 267 that for the expression of ”disgust” nose region emerged as the salient
 268 but the framework do not explicitly extracts features from the nose
 269 region as the region of nose that emrged as salient is the upper nose
 270 wrinkle area which is connected and already included in the localiza-

271 tion of the eyes region, refer Figure 3) and concatenates them with the
272 already extracted PLBP features from the mouth region, feature vector
273 of 590 dimensions ($f_1, \dots, f_{295} + f_1, \dots, f_{295}$). Then, the concate-
274 nated feature vector is fed to the classifier (“Classifier-c”) for the final
275 classification.

276 6. Experiment and results

277 We performed person-independent facial expression recognition using pro-
278 posed PLBP features ¹. We performed four experiments to test different
279 scenarios.

- 280 1. First experiment was performed on the extended Cohn-Kanade (CK+)
281 database [16]. This database contains 593 sequences of posed universal
282 expressions.
- 283 2. Second experiment was performed to test the performance of the pro-
284 posed framework on low resolution image sequences.
- 285 3. Third experiment tests the robustness of the proposed framework when
286 generalizing on the new dataset.
- 287 4. Fourth experiment was performed on the MMI facial expression database
288 (Part IV and V of the database) [27] which contains spontaneous/natural
289 expressions.

290 For the first two experiments we used all the 309 sequences from the
291 CK+ database which have FACS coded expression label [8]. The experiment

¹video showing the result of the proposed framework on good quality image sequences
is available at: http://liris.cnrs.fr/~rakhan/FER_Demo.wmv

292 was carried out on the frames which covers the status of onset to apex of the
293 expression, as done by Yang et al. [33]. Region of interest was obtained auto-
294 matically by using Viola-Jones object detection algorithm [30] and processed
295 to obtain PLBP feature vector. We extracted LBP features only from the
296 salient region(s) using $LBP_{8,2}^{u2}$ operator which denotes a uniform LBP oper-
297 ator with 8 sampling pixels in a local neighborhood region of radius 2. Only
298 exception was in the second experiment, when we adopted $LBP_{4,1}^{u2}$ operator
299 when the spatial facial resolution gets smaller than 36 x 48.

300 In our framework we created image pyramid up to level 1, so in turn got
301 five sub-regions from one facial region i.e. mouth region (see Figure. 2).
302 In total we obtained 295 dimensional feature vector (59 dimensions / sub-
303 region). As mentioned earlier we adopted $LBP_{4,1}^{u2}$ operator when the spatial
304 facial resolution was 18 x 24. In this case we obtained 75 dimensional feature
305 vector (15 dimensions / sub-region).

306 We recorded correct classification accuracy in the range of 95% for image
307 pyramid level 1. We decided not to test framework with further image pyra-
308 mid levels as it would double the size of feature vector and thus increase the
309 feature extraction time and likely would add few percents in the accuracy of
310 the framework which will be insignificant for a framework holistically.

311 6.1. First experiment: posed expressions

312 This experiment measures the performance of the proposed framework
313 on the classical database i.e. extended Cohn-Kanade (CK+) database [16].
314 Most of the methods in literature report their performance on this database,
315 so this experiment could be considered as the benchmark experiment for
316 facial expression recognition framework.

317 The performance of the framework was evaluated for five different classi-
318 fiers:

- 319 1. Support Vector Machine (SVM) with χ^2 kernel and $\gamma=1$
- 320 2. C4.5 Decision Tree (DT) with reduced-error pruning
- 321 3. Random Forest (RF) of 10 trees
- 322 4. 2 Nearest Neighbor (2NN) based on Euclidean distance
- 323 5. Naive Bayes (NB) classifier

324 Above mentioned classifiers are briefly described below.

325 *Support vector machine (SVM)*. SVM performs an implicit mapping of
326 data into a higher dimensional feature space, and then finds a linear sepa-
327 rating hyperplane with the maximal margin to separate data in this higher
328 dimensional space [29]. Given a training set of labeled examples $\{ (x_i, y_i) , i$
329 $= 1 \dots l \}$ where $x_i \in \mathfrak{R}^n$ and $y_i \in \{-1, 1\}$, a new test example x is classified
330 by the following function:

$$f(x) = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + b\right) \quad (6)$$

331 where α_i are Langrange multipliers of a dual optimization problem that
332 describe the separating hyperplane, $K(., .)$ is a kernel function, and b is the
333 threshold parameter of the hyperplane. We used Chi-Square kernel as it is
334 best suited for histograms. It is given by:

$$K(x, y) = 1 - \sum_i \frac{2 \times (x_i - y_i)^2}{(x_i + y_i)} \quad (7)$$

335 *Classification Trees*. A Classification Tree is a classifier composed by
336 nodes and branches which break the set of samples into a set of covering

337 decision rules. In each node, a single test is made to obtain the partition.
 338 The starting node is called the root of the tree. In the final nodes or leaves,
 339 a decision about the classification of the case is made. In this work, we have
 340 used C4.5 paradigm [24]. Random Forest (RFs) are collections of Decision
 341 Trees (DTs) that have been constructed randomly. RFs generally performs
 342 better than DT on unseen data.

343 *Instance Based Learning.* k -NN classifiers are instance-based algorithms
 344 taking a conceptually straightforward approach to approximating real or dis-
 345 crete valued target functions. The learning process consists in simply storing
 346 the presented data. All instances correspond to points in an n -dimensional
 347 space and the nearest neighbors of a given query are defined in terms of the
 348 standard Euclidean distance. The probability of a query q belonging to a
 349 class c can be calculated as follows:

$$p(c | q) = \frac{\sum_{k \in K} W_k \cdot 1_{(kc=c)}}{\sum_{k \in K} W_k} \quad (8)$$

$$W_k = \frac{1}{d(k, q)} \quad (9)$$

350 K is the set of nearest neighbors, kc the class of k and $d(k, q)$ the Eu-
 351 clidean distance of k from q .

352 *Naive Bayes Classifiers.* The Naive-Bayes (NB) classifier uses the Bayes
 353 theorem to predict the class for each case, assuming that the predictive genes
 354 are independent given the category. To classify a new sample characterized

355 by d genes $X = (X_1, X_2, \dots, X_d)$, the NB classifier applies the following rule:

$$C_N - B = \arg \max_{c_j \in C} p(c_j) \prod_{i=1}^d p(x_i | c_j) \quad (10)$$

356 where $C_N - B$ denotes the class label predicted by the Naive-Bayes clas-
357 sifier and the possible classes of the problem are grouped in $C = \{c_1, \dots, c_l\}$.

358 6.1.1. Results

359 The framework achieved average recognition rate of 96.7%, 97.9%, 96.2%,
360 94.7 % and 90.2 % for SVM, 2Nearest Neighbor (2NN), Random Forest
361 (RF), C4.5 Decision Tree (DT) and Naive Bayes (NB) respectively using 10-
362 fold cross validation technique. One of the most interesting aspects of our
363 approach is that it gives excellent results for a simple 2NN classifier which is a
364 non-parametric method. This points to the fact that framework do not need
365 computationally expensive methods such as SVM, random forests or decision
366 trees to obtain good results. In general, the proposed framework achieved
367 high expression recognition accuracies irrespective of the classifiers, proves
368 the descriptive strength of the extracted features (features minimizes within-
369 class variations of expressions, while maximizes between class variations).
370 For comparison and reporting results, we have used the classification results
371 obtained by the SVM as it is the most cited method for classification in the
372 literature.

373 6.1.2. Comparisons

374 We chose to compare average recognition performance of our framework
375 with the framework proposed by Shan et.al [25] with different SVM kernels.

Table 1: Comparison of time and memory consumption.

	LBP [25]	Gabor [25]	Gabor [3]	PLBP
Memory (feature dimension)	2,478	42,650	92,160	590
Time(feature extraction time)	0.03s	30s	-	0.01s

376 Our choice was based on the fact that both have common underlying descrip-
 377 tor i.e. local binary pattern (LBP), secondly framework proposed by Shan
 378 et.al [25] is highly cited in the literature. Our framework obtained average
 379 recognition percentage of 93.5% for SVM linear kernel while for the same
 380 kernel Shan et.al [25] have reported 91.5%. For SVM with polynomial kernel
 381 and SVM with RBF kernel our framework achieved recognition accuracy of
 382 94.7% and 94.9% respectively, as compared to 91.5% and 92.6%.

383 In terms of time and memory costs of feature extraction process, we have
 384 measured and compared our descriptor with the LBP and Gabor-wavelet
 385 features in Table 1. Table 1 shows the effectiveness of the proposed descrip-
 386 tor for facial feature analysis i.e. PLBP, for real-time applications as it is
 387 memory efficient and its extraction time is much lower than other compared
 388 descriptor (see Section 5 for the dimensionality calculation). In Table 1 fea-
 389 ture dimension reported are stored in a data type "float" and float occupies
 390 four bytes. The proposed framework is compared with other state-of-the-art
 391 frameworks using same database (i.e Cohn-Kanade database) and the results
 392 are presented in Table 2.

393 Table 2 shows the comparison of the achieved average recognition rate
 394 of the proposed framework with the state-of-the-art methods using same
 395 database (i.e Cohn-Kanade database). Results from [33] are presented for the

Table 2: Comparison with the state-of-the-art methods for posed expressions.

	Sequence	Class	Performance	Recog.
	Num	Num	Measure	Rate (%)
[15]	313	7	leave-one-out	93.3
[35]	374	6	2-fold	95.19
[35]	374	6	10-fold	96.26
[14]	374	6	5-fold	94.5
[26]	375	6	-	93.8
[33]a	352	6	66% split	92.3
[33]b	352	6	66% split	80
Ours	309	6	10-fold	96.7
Ours	309	6	2-fold	95.2

396 two configurations. “[33]a” shows the result when the method was evaluated
 397 for the last three frames from the sequence while “[33]b” presents the reported
 398 result for the frames which encompasses the status from onset to apex of the
 399 expression. It can be observed from the Table 2 that the proposed framework
 400 is comparable to any other state-of-the-art method in terms of expression
 401 recognition accuracy. The method discussed in “[33]b” is directly comparable
 402 to our method, as we also evaluated the framework on similar frames. In
 403 this configuration, our framework is better in terms of average recognition
 404 accuracy.

405 In general, Table 1 and 2 show that the framework is better than the
 406 state-of-the-art frameworks in terms of average expression recognition per-
 407 formance, time and memory costs of feature extraction processes. These

408 results show that the system could be used with the high degree of confi-
 409 dence for real-time applications as its unoptimized Matlab implementation
 410 runs at more than 30 frames/second (30 fps).

411 *6.2. Second experiment: low resolution image sequences*

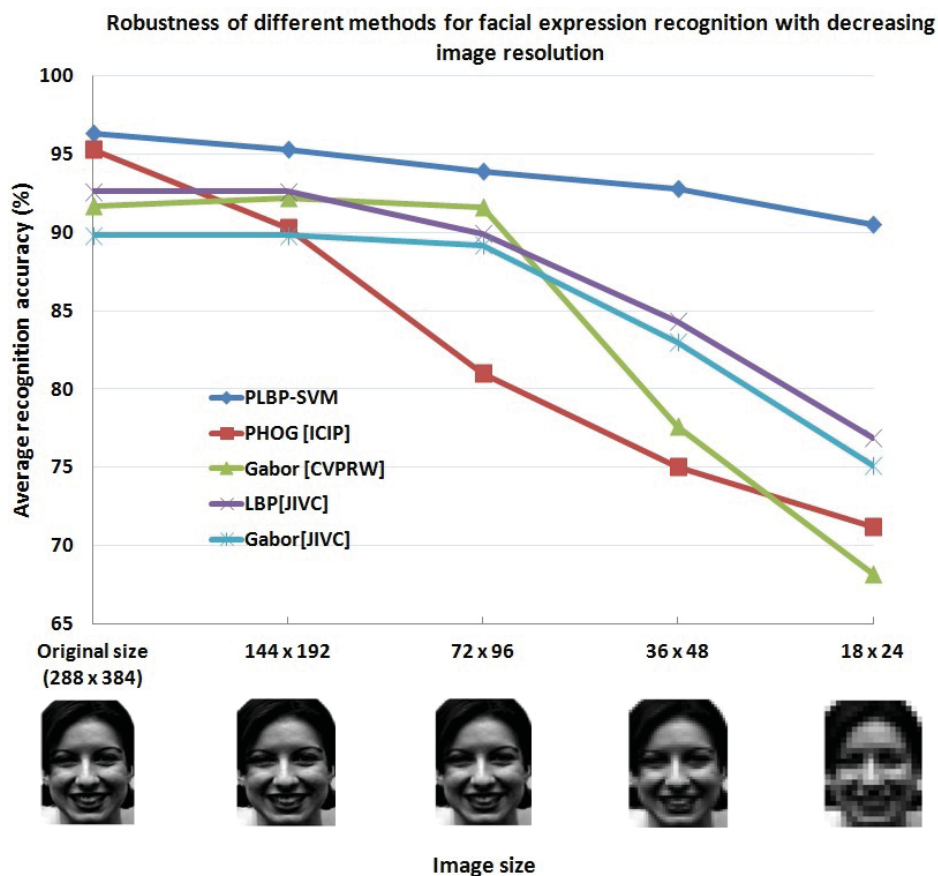


Figure 5: Robustness of different methods for facial expression recognition with decreasing image resolution. PHOG[ICIP] corresponds to framework proposed by Khan et. al [12], Gabor [CVPRW] corresponds to Tian’s work [26], LBP[JIVC] and Gabor[JIVC] corresponds to results reported by Shan et. al [25]

412 Most of the existing state-of-the-art systems for expressions recognition

413 report their results on high resolution images with out reporting results on
 414 low resolution images. As mentioned earlier there are many real world ap-
 415 plications that require expression recognition system to work amicably on
 416 low resolution images. Smart meeting, video conferencing and visual surveil-
 417 lance are some examples of such applications. To compare with Tian’s work
 418 [26], we tested our proposed framework on low resolution images of four dif-
 419 ferent facial resolutions (144 x 192, 72 x 96, 36 x 48, 18 x 24) based on
 420 Cohn-Kanade database. Tian’s work can be considered as the pioneering
 421 work for low resolution image facial expression recognition. Figure 5 shows
 422 the images at different spatial resolution along with the average recognition
 423 accuracy achieved by the different methods. Low resolution image sequences
 424 were obtained by down sampling the original sequences. All the other ex-
 425 perimental parameters i.e. descriptor, number of sequences and region of
 426 interest, were same as mentioned earlier in the Section 6.

427 Figure 5 reports the recognition results of the proposed framework with
 428 the state-of-the-art methods on four different low facial resolution images.
 429 Reported results of our proposed method i.e. are obtained using support
 430 vector machine (SVM)” with χ^2 kernel and $\gamma=1$. In Figure 5 recognition
 431 curve for our proposed method is shown as *PLBP-SVM*, recognition curves
 432 of LBP [25] and Gabor [25] are shown as *LBP[JIVC]* and *Gabor[JIVC]* re-
 433 spectively, curve for Tian’s work [26] is shown as *Gabor[CVPRW]* while Khan
 434 et al. [12] proposed system’s curve is shown as *PHOG[ICIP]*. Results reports
 435 in LBP [25] and Gabor [25], the different facial image resolution are 110 x
 436 150, 55 x 75, 27 x 37 and 14 x 19 which are comparable to the resolutions
 437 of 144 x 192, 72 x 96, 36 x 48, 18 x 24 pixels in our experiment. Referenced

438 figure shows the supremacy of the proposed framework for low resolution im-
439 ages. Specially for the smallest tested facial image resolution (18 x 24) our
440 framework performs much better than any other compared state-of-the-art
441 method.

442 Results from the first and second experiment show that the proposed
443 framework for facial expression recognition works amicably on classical dataset
444 (CK dataset) and its performance is not effected significantly for low reso-
445 lution images. Secondly, the framework has a very low memory requirement
446 and thus it can be utilized for real-time applications.

447 *6.3. Third experiment: generalization on the new dataset*

448 The aim of this experiment is to study how well the proposed frame-
449 work generalizes on the new dataset. We used image sequences from CK+
450 dataset and FG-NET FEED (Facial Expressions and Emotion Database) [31].
451 FG-NET FEED contains 399 video sequences across 18 different individuals
452 showing seven facial expressions i.e. six universal expression [6] plus one
453 neutral. In this dataset individuals were not asked to act rather expressions
454 were captured while showing them video clips or still images.

455 The experiment was carried out on the frames which covers the status
456 of onset to apex of the expression as done in the previous experiment. This
457 experiment was performed in two different scenarios, with the same classifier
458 parameters as the first experiment:

- 459 a. In the first scenario samples from the CK+ database were used for the
460 training of different classifiers and samples from FG-NET FEED [31]
461 were used for the testing. Obtained results are presented in Table 3.

462 b. In the second scenario we used samples from the FG-NET FEED for the
 463 training and testing was carried out with the CK+ database samples.
 464 Results obtained are presented in last two rows of Table 3.

465 This experiment simulates the real life situation when the framework
 466 would be employed to recognize facial expressions on the unseen data. Ob-
 467 tained results are presented in Table 3. Reported average recognition per-
 468 centages for training phase were calculated using 10-fold cross validation
 469 method. Obtained results are encouraging and they can be further improved
 470 by training classifiers on more than one dataset before using in real life sce-
 471 nario.

Table 3: Average recognition accuracy (%)

Training on CK+ database and testing it with FG-NET FEED				
	SVM	C4.5 DT	RF	2NN
Training samples	96.7	94.7	96.2	97.9
Test samples	81.9	74.8	79.5	83.1
Training on FG-NET FEED and testing it with CK+ database				
Training samples	92.3	91.2	90.5	93.3
Test samples	80.5	77.3	79	84.7

472 *6.4. Fourth experiment: spontaneous expressions*

473 Spontaneous/natural facial expressions differ substantially from posed ex-
 474 pressions [2]. The same has also been proved by psychophysical work [7]. To

475 test the performance of the proposed framework on the spontaneous facial
476 expressions we used 392 video segments from part IV and V of the MMI
477 facial expression database [27]. Part IV and V of the database contains
478 spontaneous/naturalistic expressions recorded from 25 participants aged be-
479 tween 20 and 32 years in two different settings. Due to ethical concerns the
480 database contains only the video recording of the expressions of happiness,
481 surprise and disgust [27].

482 The framework achieved average recognition rate of 91%, 91.4%, 90.3%
483 and 88% for SVM, 2-nearest neighbor, Random forest and C4.5 decision tree
484 respectively using 10-fold cross validation technique. Algorithm of Park et
485 al.[23] for spontaneous expression recognition achieved results for three ex-
486 pressions in the range of 56% to 88% for four different configurations which is
487 less than recognition rate of our proposed algorithm, although results cannot
488 be compared directly as they used different database.

489 7. Conclusions and future work

490 We presented a novel descriptor and framework for automatic and reliable
491 facial expression recognition. Framework is based on initial study of human
492 vision and works adequately on posed as well as on spontaneous expressions.
493 The key conclusion drawn from the study are:

- 494 1. Facial expressions can be analyzed automatically by mimicking human
495 visual system i.e. extracting features only from the salient facial re-
496 gions.
- 497 2. Features extracted using proposed pyramidal local binary pattern (PLBP)
498 operator have strong discriminative ability as the recognition result for

- 499 six universal expressions is not effected by the choice of classifier.
- 500 3. The proposed framework is robust for low resolution images, sponta-
501 neous expressions and generalizes well on unseen data.
 - 502 4. The proposed framework can be used for real-time applications since
503 its unoptimized Matlab implementation run at more than 30 frames /
504 second (30 fps) on a Windows 64 bit machine with i7 processor running
505 at 2.4 GHz having 6 GB of RAM.

506 In future we plan to investigate the effect of occlusion as this parameter
507 could significantly impact the performance of the framework for real world
508 applications. Secondly, the notion of movement could improve the perfor-
509 mance of the proposed framework for real world applications as the exper-
510 imental study conducted by Bassili [4] suggested that dynamic information
511 is important for facial expression recognition. Another parameter that needs
512 to be investigated is the variations of camera angle as for many applications
513 frontal facial pose is difficult to record.

514 **References**

- 515 [1] Bai, Y., Guo, L., Jin, L., Huang, Q., 2009. A novel feature extrac-
516 tion method using pyramid histogram of orientation gradients for smile
517 recognition, in: International Conference on Image Processing.
- 518 [2] Bartlett, M.S., Littlewort, G., Braathen, B., Sejnowski, T.J., Movellan,
519 J.R., 2002. A prototype for automatic recognition of spontaneous facial
520 actions, in: Advances in Neural Information Processing Systems.

- 521 [3] Bartlett, M.S., Littlewort, G., Fasel, I., Movellan, J.R., 2003. Real time
522 face detection and facial expression recognition: Development and ap-
523 plications to human computer interaction., in: Conference on Computer
524 Vision and Pattern Recognition Workshop, 2003.
- 525 [4] Bassili, J.N., 1979. Emotion recognition: The role of facial movement
526 and the relative importance of upper and lower areas of the face. *Journal*
527 *of Personality and Social Psychology* 37, 2049–2058.
- 528 [5] Donato, G., Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J.,
529 1999. Classifying facial actions. *IEEE Transaction on Pattern Analysis*
530 *and Machine Intelligence* 21, 974–989.
- 531 [6] Ekman, P., 1971. Universals and cultural differences in facial expressions
532 of emotion, in: *Nebraska Symposium on Motivation*, Lincoln University
533 of Nebraska Press. pp. 207–283.
- 534 [7] Ekman, P., 2001. *Telling Lies: Clues to Deceit in the Marketplace,*
535 *Politics, and Marriage.* W. W. Norton & Company, New York. 3rd
536 edition.
- 537 [8] Ekman, P., Friesen, W., 1978. The facial action coding system: A tech-
538 nique for the measurement of facial movements. *Consulting Psychologist*
539 .
- 540 [9] Guo, Z., Zhang, L., Zhang, D., Mou, X., 2010. Hierarchical multiscale
541 lbp for face and palmprint recognition, in: *IEEE International Confer-*
542 *ence on Image Processing*, pp. 4521–4524.

- 543 [10] Hadjidemetriou, E., Grossberg, M., Nayar, S., 2004. Multiresolution
544 histograms and their use for recognition. *IEEE Transactions on Pattern*
545 *Analysis and Machine Intelligence* 26, 831 –847.
- 546 [11] Khan, R.A., Meyer, A., Konik, H., Bouakaz, S., 2012a. Exploring hu-
547 man visual system: study to aid the development of automatic facial
548 expression recognition framework, in: *Computer Vision and Pattern*
549 *Recognition Workshop*.
- 550 [12] Khan, R.A., Meyer, A., Konik, H., Bouakaz, S., 2012b. Human vision
551 inspired framework for facial expressions recognition, in: *IEEE (Ed.),*
552 *IEEE International Conference on Image Processing*.
- 553 [13] Kolsch, M., Turk, M., 2004. Analysis of rotational robustness of hand
554 detection with a viola-jones detector, in: *17th International Conference*
555 *on Pattern Recognition*, pp. 107 – 110 Vol.3.
- 556 [14] Kotsia, I., Zafeiriou, S., Pitas, I., 2008. Texture and shape information
557 fusion for facial expression and facial action unit recognition. *Pattern*
558 *Recognition* 41, 833–851.
- 559 [15] Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J., Movellan, J., 2006.
560 Dynamics of facial expression extracted automatically from video. *Image*
561 *and Vision Computing* 24, 615–625.
- 562 [16] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews,
563 I., 2010. The extended cohn-kande dataset (CK+): A complete facial
564 expression dataset for action unit and emotion-specified expression., in:

- 565 IEEE Conference on Computer Vision and Pattern Recognition Work-
566 shops.
- 567 [17] Moore, S., Bowden, R., 2011. Local binary patterns for multi-view facial
568 expression recognition. *Computer Vision and Image Understanding* 115,
569 541 – 558.
- 570 [18] Ojala, T., Pietikäinen, M., 1999. Unsupervised texture segmentation
571 using feature distributions. *Pattern Recognition* 32, 477– 486.
- 572 [19] Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study
573 of texture measures with classification based on featured distribution.
574 *Pattern Recognition* 29, 51–59.
- 575 [20] Ojala, T., Pietikäinen, M., Mäenpää, T., 2002. Multiresolution gray-
576 scale and rotation invariant texture classification with local binary pat-
577 terns. *IEEE Transaction on Pattern Analysis and Machine Intelligence*
578 24, 971–987.
- 579 [21] Ojansivu, V., Heikkilä, J., 2008. Blur insensitive texture classification
580 using local phase quantization, in: *International conference on Image
581 and Signal Processing*.
- 582 [22] Pantic, M., Patras, I., 2006. Dynamics of facial expression: recognition
583 of facial actions and their temporal segments from face profile image
584 sequences. *IEEE Transactions on Systems, Man, and Cybernetics* 36,
585 433–449.
- 586 [23] Park, S., Kim, D., 2008. Spontaneous facial expression classification

- 587 with facial motion vector, in: IEEE Conference on Automatic Face and
588 Gesture Recognition.
- 589 [24] Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan
590 Kaufmann.
- 591 [25] Shan, C., Gong, S., McOwan, P.W., 2009. Facial expression recognition
592 based on local binary patterns: A comprehensive study. Image and
593 Vision Computing 27, 803–816.
- 594 [26] Tian, Y., 2004. Evaluation of face resolution for expression analysis, in:
595 Computer Vision and Pattern Recognition Workshop.
- 596 [27] Valstar, M., Pantic, M., 2010. Induced disgust, happiness and surprise:
597 an addition to the MMI facial expression database, in: International
598 Language Resources and Evaluation Conference.
- 599 [28] Valstar, M., Patras, I., Pantic, M., 2005. Facial action unit detection
600 using probabilistic actively learned support vector machines on tracked
601 facial point data, in: IEEE Conference on Computer Vision and Pattern
602 Recognition Workshop, pp. 76–84.
- 603 [29] Vapnik, V.N., 1995. The nature of statistical learning theory. New York:
604 Springer-Verlag.
- 605 [30] Viola, P., Jones, M., 2001. Rapid object detection using a boosted
606 cascade of simple features, in: IEEE Conference on Computer Vision
607 and Pattern Recognition.

- 608 [31] Wallhoff, F., 2006. Facial expressions and emotion database.
609 www.mmk.ei.tum.de/~waf/fgnet/feedtum.html.
- 610 [32] Wang, W., Chen, W., Xu, D., 2011. Pyramid-based multi-scale lbp
611 features for face recognition, in: International Conference on Multimedia
612 and Signal Processing (CMSP), pp. 151–155.
- 613 [33] Yang, P., Liu, Q., Metaxas, D.N., 2010. Exploring facial expressions
614 with compositional features, in: IEEE Conference on Computer Vision
615 and Pattern Recognition.
- 616 [34] Zhang, Y., Ji, Q., 2005. Active and dynamic information fusion for facial
617 expression understanding from image sequences. *IEEE Transactions on*
618 *Pattern Analysis and Machine Intelligence* 27, 699–714.
- 619 [35] Zhao, G., Pietikäinen, M., 2007. Dynamic texture recognition using
620 local binary patterns with an application to facial expressions. *IEEE*
621 *Transaction on Pattern Analysis and Machine Intelligence* 29, 915–928.
- 622 [36] Zhaoping, L., 2006. Theoretical understanding of the early visual pro-
623 cesses by data compression and data selection. *Network: computation*
624 *in neural systems* 17, 301–334.