



**HAL**  
open science

## Optimal marker set assessment for motion capture of 3D mimic facial movements

Nicole Dagnes, Federica Marcolin, Enrico Vezzetti, François-Régis Sarhan, Stéphanie Dakpé, Frédéric Marin, Francesca Nonis, Khalil Ben Mansour

### ► To cite this version:

Nicole Dagnes, Federica Marcolin, Enrico Vezzetti, François-Régis Sarhan, Stéphanie Dakpé, et al.. Optimal marker set assessment for motion capture of 3D mimic facial movements. *Journal of Biomechanics*, 2019, 93, pp.86-93. 10.1016/j.jbiomech.2019.06.012 . hal-02406863

**HAL Id: hal-02406863**

**<https://hal.utc.fr/hal-02406863>**

Submitted on 20 Dec 2021

**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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## Title Page & Abstract

### Title

# Optimal Marker Set Assessment for Motion Capture of 3D Mimic Facial Movements

### Autors :

Nicole DAGNES<sup>1, 2</sup>, Federica MARCOLIN<sup>2</sup>, Enrico Vezzetti<sup>2</sup>, François-Régis SARHAN<sup>3,5</sup>,  
Stéphanie DAKPE<sup>3, 4, 5</sup>, Frédéric MARIN<sup>1</sup>, Francesca NONIS<sup>2</sup>, Khalil BEN MANSOUR<sup>1\*</sup>.

<sup>1</sup>Sorbonne Universités, Université de Technologie de Compiègne, UMR CNRS 7338, France

<sup>2</sup>Department of Management and Production Engineering, Politecnico di Torino, Torino, Italy

<sup>3</sup>Service de Chirurgie Maxillo-faciale, Centre Hospitalier Universitaire d'Amiens, France

<sup>4</sup>Institut Faire Faces, Amiens, France

<sup>5</sup>EA 7516 Chimère, Université de Picardie Jules Verne, Amiens, France

### \* Corresponding author at :

Université de Technologie de Compiègne

Centre d'innovation CS60319,

60203 Compiègne Cedex, France

**E-mail address:** [khalil.ben-mansour@utc.fr](mailto:khalil.ben-mansour@utc.fr)

# Optimal Marker Set Assessment for Motion Capture of 3D Mimic Facial Movements

## Abstract

Nowadays, facial mimicry studies have acquired a great importance in the clinical domain and 3D motion capture systems are becoming valid tools for analysing facial muscles movements, thanks to the remarkable developments achieved in the 1990s.

However, the face analysis domain suffers from a lack of valid motion capture protocol, due to the complexity of the human face. Indeed, a framework for defining the optimal marker set layout does not exist yet and, up to date, researchers still use their traditional facial point sets with manually allocated markers.

Therefore, the study proposes an automatic approach to compute a minimum optimized marker layout to be exploited in facial motion capture, able to simplify the marker allocation without decreasing the significance level. Specifically, the algorithm identifies the optimal facial marker layouts selecting the subsets of linear distances among markers that allow to automatically recognizing with the highest performances, through a k-nearest neighbours classification technique, the acted facial movements. The marker layouts are extracted from them. Various validation and testing phases have demonstrated the accuracy, robustness and usefulness of the custom approach.

## Keywords

3D Face; Face Analysis; Motion Capture; Marker Optimization; Feature Extraction.

## 1 Introduction

The face is one of the most important parts of the human anatomy, as it is a huge source of information, and plays an essential role in social interaction. In fact, facial expression is one of the ways for conveying emotional messages, creating interpersonal communication and establishing links between individuals (Bargiela-Chiappini & Haugh, 2010). Muscle movements, in particular, are the key element which facial expressions rely on. Indeed, their role in the social environment of the individual is core for communication activities. However, accidents or musculoskeletal face disorders may lead to alterations in facial muscle functions, bringing to unsynchronized facial movements or even to facial paralysis; these conditions not only cause a loss of physical function but also affect the patient's social communications and interaction, damaging his wellbeing.

Hence, in these decades, the objective quantification of facial movements has acquired great importance in the clinical domain, and facial mimicry studies have been carried out embracing many applications, such as for helping maxillofacial surgery (Sforza, et al., 2010) and facial motion rehabilitation (Bajaj-Luthra, et al., 1998) (Byrne, 2004).

Currently, different methods are employed for evaluating facial muscle function. One way consists of observing their contraction, which causes local facial skin displacement and, subsequently, visible facial appearance changes, to record the clinical observations in writing and to regularly photograph facial expressions. Then, these data are analysed according to biopsychosocial aspects. In other cases, scales are used to quantify deficits, as numerous scales have been proposed (House, 1983) (Reitzen, et al., 2009) (Henstrom, et al., 2011) during the last years. However, it stills a difficult task to understand and correctly evaluate facial deficits, as it requires specific experience (Di Stadio, 2015).

55 Another approach to quantify facial movements is to use facial electromyography (EMG), which has been  
56 found to be a useful tool since it is sensitive even to small facial muscle changes that no visual coding  
57 technique can capture (Van Boxtel, 2010) (Gupta, et al., 2017) (Chandu, et al., 2005). However, EMG has  
58 some limitations that can reduce its effectiveness in the field of facial analysis. Firstly, the initial setup  
59 requires a significant amount of time; then, it is an invasive technique, as the connection between skin  
60 electrodes and the control device can interfere with the natural and spontaneous behaviour of the patient.

61  
62 These evaluations brought the researchers to investigate other approaches such as 3D scan techniques, which  
63 have resulted to be a valid tool for evaluating facial muscle movements. In fact, these systems can be used  
64 for planning future maxillofacial surgery (Adolphs, et al., 2012), as well as for quantifying soft tissue  
65 changes (Bianchi, et al., 2012) and facial mimics' variations in patients before and after the treatment  
66 (Bianchi, et al., 2012). The facial surface data are acquired using three-dimensional scanners and the facial  
67 movements' information is described in terms of surface and landmark displacements (Sjogreen, et al.,  
68 2010). However, the main drawback of scanner systems is that they do not measure facial movements in  
69 motion (Ju, et al., 2012).

70  
71 3D motion capture systems solve the previous problems. Relying on multiple external sensors (i.e. calibrated  
72 video cameras), it tracks the movements performed by a subject, equipped with a set of tiny markers placed  
73 on his face. The configuration can involve a relatively large number of markers, but it can easily change,  
74 according to the researchers' needs and goals, always permitting to the skin to move freely. The position,  
75 velocity and displacement information of the markers in the three dimensions are determined by cameras, by  
76 capturing the light reflected or emitted by the markers and using triangulation.

77  
78 Several studies in the clinical field have already highlighted the potential of this technology, but, up to date,  
79 most of the researchers use their own manually placed marker configurations, which significantly vary in the  
80 total number of facial points and in their location.

81  
82 Presumably, the main reason is that finding a standardized and optimal marker layout is challenging due to  
83 the anatomical complexity, large shape variation, and no rigid deformability of the human face. Moreover,  
84 choosing the correct marker displacement is not a simple task, as their number and position are directly  
85 related to the chosen application and they affect the usefulness of motion capture data. Indeed, the reliability  
86 of a motion capture analysis is directly connected to the number of available markers. Thus, dozens of  
87 markers are needed for adequately tracking the complex changes due to facial motions and skin deformations  
88 (Furukawa & Ponce, 2010). However, having an over-densely set placed in a small region also increases the  
89 redundancy in the recorded data, making it harder to capture individual markers, and unnecessarily waste  
90 technicians' time to dispose dozens of them on the facial surface.

91  
92 This paper proposes an automatic approach to compute a minimum optimized facial marker layout for facial  
93 motion acquisition, choosing those that best of all permit to automatically recognise the type of movement  
94 performed. In fact, marker location and number have to be conceived in order to avoid significant loss of  
95 quality (reliability and accuracy), especially for their application in the clinical domain. Moreover, having  
96 optimized facial marker layouts will help to improve the efficiency and practicality of facial motion data in  
97 clinical applications.

98  
99 The article is structured as follows. Section 2 gives an overview on the materials used. Then, it faces the  
100 developed method, used for computing novel optimized marker layouts. Section 3 provides all the layouts  
101 obtained, whereas Section 4 contain the discussion part with the satisfactory results and the currently  
102 limitation of the developed procedure. Finally, Section 5 concludes the work, providing some clues about  
103 future works and researches.

104  
105

## 106 **2 Materials and Methods**

### 107 **2.1 Protocol and Instrumentation**

108  
109

110 Fifteen healthy young adult volunteers aged from 20 to 30 years participated in this study. This study was  
111 approved by the local ethics committee (n°2011-A00532-39), was registered in clinicaltrials.gov  
112 (NCT02002572) and was performed in accordance with the ethical standards of the 1964 declaration  
113 of Helsinki. Informed consent was obtained from each subject.

114  
115 As shown in Figure 1, each participant was equipped with 109 markers ( $\emptyset$  1.5mm) fixed by a trained  
116 physiotherapist on the anatomical facial surface (Hontanilla & Aubá, 2008), and 3 markers pasted on a rigid  
117 structure attached on the maxilla (Rm), which represent the most reliable reference for an accurate estimation  
118 of the facial movements (Ben Mansour, et al., 2014).

119  
120  
121 <Figure 1 near here >  
122

123 Every subject was asked to perform six facial expressions, later called MOUV1, MOUV2, MOUV3,  
124 MOUV4, MOUV5 and MOUV 6. These movements, shown in Figure 2, are chosen due to their great  
125 importance in the analysis of facial expression in the healthy, pathological or rehabilitative subject (Sarhan,  
126 2017). Moreover, they take place in different zones of the face, involving both the soft tissues of the frontal  
127 and orbicular zones and the zones of the lips and the chin.

128  
129 <Figure 2 near here>  
130

131 Every capture, acquired through 17 optoelectronic cameras T160 and two Bonita video cameras (ViconLtd,  
132 Oxford, UK) at a recording frequency of 100Hz (i.e. at 100 frames per second (fps)), contains the 3D  
133 coordinates of facial markers. The six movements are captured separately and each movement is performed  
134 three times.

135  
136 Finally, data are imported into Matlab (Mathworks, R2016a) where a custom algorithm processes and  
137 analyses the marker 3D positions.

138  
139 The computer used is an HP PC, with a 2.5 GHz Intel Core i5 processor and 16 GB of RAM. The  
140 operating system is the Windows 8 version.

## 141 142 143 **2.2 Problem Analysis** 144

145 Since the fundamental goal of clinical facial motion analysis is to record the motion information quickly and  
146 accurately, the proposed approach identifies the optimum facial marker layout by choosing the marker setups  
147 that best permit an automatic recognition of the type of movement performed. In other words, the capacity of  
148 a set of markers to discriminate the different movements is the criterion used to quantify the goodness of a  
149 layout.

150  
151 The algorithm identifies the optimal setups selecting the subsets of linear distances among markers that best  
152 allow the automatic recognition, through a k-nearest neighbours (k-NN) classification technique, of the  
153 performed facial movements. The marker layouts are extracted from them. The general structure of the  
154 algorithm is shown in Figure 3, which also underlines how each step regressively reduces the number of  
155 useful markers, up to few dozen.

156  
157  
158 < Figure 3 near here>  
159

## 160 **2.3. Data Input** 161

162 Let  $F$  be the number of the frame in a facial motion capture, and  $N$  be the number of markers. Assuming  
163  $m_i \in \mathbb{IR}^3$  is the 3D coordinate of the marker  $i$ , expressed in the Maxilla reference, and  $m_i^{(t)} \in \mathbb{IR}^3$  is the 3D  
164 coordinate of marker  $i$  at the  $t$  th frame. Each facial motion capture can be represented as a matrix,  $X \in \mathbb{IR}^{3 \times N \times F}$ ,

165 where for each marker three consecutive rows are needed for storing its position in the x, y, z reference and  
 166 each column represents the displacement of one point over the time, as shown below:  
 167

$$168 \quad X = \begin{bmatrix} m_{1,x}^1 & m_{1,x}^2 & m_{1,x}^3 & \dots & m_{1,x}^F \\ m_{1,y}^1 & m_{1,y}^2 & m_{1,y}^3 & \dots & \vdots \\ m_{1,z}^1 & m_{1,z}^2 & m_{1,z}^3 & \dots & \vdots \\ m_{2,x}^1 & m_{2,x}^2 & m_{2,x}^3 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{N,z}^1 & m_{N,z}^2 & m_{N,z}^3 & \dots & m_{N,z}^F \end{bmatrix}$$

169  
 170  
 171

## 172 2.4. Pre-processing step

173

174 As the number of markers is very high, a first selection is undertaken to reduce the dimensionality of the  
 175 problem, and hence the time necessary for features' computation and subsequently selection. This reduction  
 176 step aims to discard those markers that do not move during the six facial movements, and that can be  
 177 considered meaningless for the movement analysis. Therefore, the trajectories study over the time (i.e. from  
 178 the starting frame until the last time stamp F) is undertaken.

179

180 Examples of trajectories are shown in Figure 4.

181

< Figure 4 near here >

182

183

184

185 Let  $DispVect \in \mathbb{R}^N$  be the vector that records the maximum displacement performed by each marker over

186

187

188

$$DispVect(i) = \max_{f=\{2,\dots,F\}} [dist(m_i^1, m_i^f)] \quad i = \{1, \dots, N\}$$

189

190 where  $dist$  is the Euclidean distances between the marker  $i$  at the frame 1 and the same marker over time.  
 191 This operation is done for every recording session. Then, the vectors are grouped in order to identify for each  
 192 movement MOUV1, MOUV2..., MOUV6 a subset of meaningful markers.

193

194

Figure 5 is a graphical representation of the results obtained through this study.

195

196

< Figure 5 near here >

197

198

199 The value, in terms of mean and variation, used as a threshold (equal to 1.39 grid point, experimentally set)  
 200 for identifying the points whose motion can be considered negligible, is calculated using the movement  
 201 information of the markers pasted on the rigid structure, attached on the maxilla. These markers, which are  
 202 considered fixed and represented the most reliable reference for an accurate estimation of the facial  
 203 movements, have some natural negligible vibrations due to the global movement of the head and  
 204 disentangled from the specific facial actions. Hence, the markers placed on the face, which have comparable  
 205 movements, may be considered meaningless and are eliminated, in order to keep only the facial deformation  
 206 for the synthesis of the rehabilitation movement.

207

208 Then, the subsets obtained for each type of movement are merged together, obtaining a set of 70 markers,  
 209 shown in Figure 6.

210

211

212

< Figure 6 near here >

213

214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266

These 70 markers will be used in the subsequent phases of the algorithm, while the markers discarded during this pre-processing step will no longer be used.

Let  $RadVect \in \mathbb{R}^N$  be the array that records the maximum radial distances among those calculated over time between each off the 70 markers and the centre of the face (the marker on the dorsum of the nose, coloured in green in Figure 9). The  $i$ -th element is calculated as

$$RadVect(i) = \max_{f=\{1,\dots,F\}} [dist(m_i^f, c^f)] \quad i = \{1, \dots, N\}$$

Let  $DistVect \in \mathbb{R}^N$  be the array that records the maximum linear distances among those calculated over time between each off the 70 markers and its nearest neighbours, automatically identified exploiting the Delaunay Triangulation technique. The  $i$ -th element is calculated as

$$DistVect(i) = \max_{f=\{1,\dots,F\}} [dist(m_i^f, m_j^f)] \quad i = \{1, \dots, N\}, j = \{1, \dots, M\}$$

with M that scans the nearest neighbours for the  $i$ -th marker.

## 2.5. Clustering step

Then, a clustering is performed, as the direct application of classification algorithm on data may not produce satisfactory results (Alapati & Sindhu, 2016).

In detail, clustering is a useful technique for finding subgroups within observations, so that objects in the same group (here called a cluster) are more similar to each other than to those in other groups. The number of expected clusters is fixed to 3, according to the structure of the data. Indeed, MOUV1-MOUV2 take place in the frontal area, whereas both MOUV3-MOUV4 and MOUV5-MOUV6 concern the mouth zone, but lips move in opposite directions for the two last movements. The three clusters are called C1, C2, and C3.

The k-means clustering algorithm is used for this task. It iteratively assigns each observation to one of the 3 groups, based on the similarity existing between the features provided. In particular, the features used for this purpose are the values recorded in  $RadVect$ , i.e. the radial distances between each marker and the centre of the face. As a result, the k-means clustering algorithm provides each observation with a data label (namely a clustering id), with value 1, 2 or 3 which will be used as a feature in the classification subsequent phase.

However, due to the high dimensionality of the problem, a Feature Selection Procedure is undertaken on  $RadVect$  before clustering. The algorithm used is the Sequential Backward Selection (SBS) algorithm, which consists of sequentially removing the least useful distances from the full set, one-at-a-time. In other words, it is a search algorithm, that is used to reduce an initial d-dimensional feature space to a m-dimensional feature subspace where  $m < d$ . According to the SBS algorithm, only 4 markers are needed for correctly performing the clustering phase. These markers are: the centre of the face, the marker located in the middle of the right eyebrow and the two markers on the sides of the mouth.

The motivation behind feature selection algorithms is to automatically select a subset of features that is most relevant to the problem, improving the computational efficiency by removing irrelevant features based on the cluster performance.

## 2.6. Classification step

267 Next, a classification technique is used to accomplish the task of “facial expression automatic recognition”  
268 on the six types of movement available. A classification is a form of data analysis which aims to find the  
269 model that better describes and distinguishes different data classes. The classes are six as the movements that  
270 the algorithm try to automatically recognize (MOUV1, MOUV2 MOUV3, MOUV4, MOUV5 and MOUV6).  
271 Among the many classification methods proposed by researchers, the k-nearest neighbours (KNN) algorithm  
272 is chosen due to its low calculation time and highly competitive results.

273  
274 Its implemented strategy may be summarized with the sentence: “tell me who your neighbours are, and I’ll  
275 tell you who you are”. In more technical words, an unknown sample is classified with the most common  
276 class among k closest samples. KNN classification is in fact a two-step learning process, consisting of a  
277 training phase (where the classification model is constructed by the classification algorithm from a set of  
278 labelled samples) and testing phase (where the model is used to predict the class label for a different testing  
279 dataset, composed of unlabelled data).

280  
281 Hence, the total number of the recorded motion capture sessions is split into two groups: a training dataset  
282 (80%) is used for the construction of the classification model during the training phase; a testing set (20%) is  
283 used for testing the classification accuracy of the model, when it is used with unlabelled data for predicting  
284 the class label (i.e. a value between 1 and 6 that represents the corresponding movement). Both sets are  
285 balanced according to the number of elements of each class, i.e. 80% of motion captures referring to the  
286 movement M1 stay in the training set, the other 20% in the testing set, and so on for all the 6 movements.

287  
288 The input parameter k, which indicates the number of neighbours to consider for labelling the unlabelled  
289 data, has been chosen equal to 7, to avoid overfitting of the model to the training set, and this value has been  
290 used for all predictions.

291  
292 The features used in this part of the procedure are the data label (namely the clustering id obtained during the  
293 clustering phase) and the values recorded in DistVect, i.e. the linear distances between each marker and its  
294 nearest neighbours.

295  
296 Again, a Feature Selection Procedure is used for reducing the dimensionality problem. In this case, the  
297 Sequential Forward Selection (SFS) algorithm is applied on DistVect in order to find the subset that allows  
298 an automatic recognition of the performed movement at the fixed rate of 95%. This algorithm, unlike the  
299 SBS method, consists of adding the distances from an empty candidate subset until the addition of further  
300 distances does not decrease the criterion (i.e. the classification rate).

301  
302 In addition, the procedure has been made recursive; at the beginning, the candidate subset is empty. Then, at  
303 each iteration, a new linear distance is added and a quality control is performed. If the current set of distances  
304 does not fulfil the imposed quality criterion (i.e. classification rate higher than 95%), the code continues to  
305 add new linear distances to the candidate subset. Otherwise, the solution is recorded and the procedure starts  
306 backtracking for finding other possible solutions.

307  
308 Moreover, several testing/training sets are randomly created starting from the initial data, following a 20/80  
309 partition, in order to investigate if the optimum subsets of linear distances automatically determined always  
310 lead to accurate classification results, making the procedure sounder.

311  
312 Finally, the marker sets are extracted from the subsets of linear distances defined previously during the  
313 classification step.

314  
315

### 316 **3 Results and discussion**

317  
318 The results of the clustering phase are shown in Figure 7 and 8, whereas Figure 9 shows the first six  
319 optimum marker set layouts, extracted from the sets of linear distances selected by our recursive approach  
320 during the classification step after 37 hours of processing.

321  
322

< Figure 7 near here >



323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378

< Figure 8 near here >

< Figure 9 near here >

Various experiments have been conducted to test and validate the optimized subsets of linear distances obtained by the automated procedure. Overall, the classification rates obtained are higher than 95% (recognition rate). These results validate the hypothesis that it is possible to select the optimal marker layouts by choosing the features that best allow for automatic recognition of the type of performed movement.

Other layouts can be found by running the algorithm longer. For the moment, these first 6 outputted marker sets have been considered adequate, since they are sufficiently different from one another, allowing clinicians to choose one layout rather than another, depending on the pathology of their patient.

The marker groups proposed contain less than 20 markers, which is a big improvement considering that the initial number of markers has been reduced by 4/5. These layouts can be used as practical guidelines for positioning facial markers, for facial movements' acquisition, especially for clinical applications.

The points' distribution on the facial surface is not homogeneous or symmetrical, as might be expected given the symmetry of the facial stimuli considered. However, this result is due to how the algorithm works. In fact, it considers each marker separately, evaluating how each point, considered alone, can improve the automatic facial expression recognition. In addition, previous studies have shown that facial expression recognition algorithms are able to function even in the presence of facial occlusions, thanks to the facial expression symmetry, which makes specular information not meaningful, or even, as shown by our algorithm, redundant.

Moreover, due to a large number of markers (109), a qualified clinician previously needed no less than 25 minutes to set up the subject. Now, the use of these new layouts considerably reduces the initial time required for the experiment setting. Moreover, tests prove that the reduction of markers still ensures an accurate quantification of the facial mimic's movement. In addition, providing different layouts, doctors have the possibility to choose the set of markers that best suits their case.

Actually, the different marker layouts provided are six. However, in the future the algorithm will be run longer in order to find more layouts, different from each other, able to adapt to a larger number of facial disorders. Indeed, different marker locations may be useful for studying different facial pathology or checking the rehabilitation degree of the specific part of the face, damaged by an injury. This way, the study will be even more focused and effective.

However, there are some limitations in the current approach. Firstly, like several other data-driven approaches, the marker layouts optimized by our algorithm rely upon the linear distances of the training facial acquisitions. Hence, if the training set is composed of a scarce variety of facial expressions and movements, the marker layouts found may not be perfectly uniformly distributed throughout the face.

At present, our training set considers a reduced number of movements in the upper and lower part of the face (natural and forced closure of the eyes, pronunciation of the sound "o" and "pu", and two types of smiles). These stimuli have been evaluated as indicative also for facial surgery by some sources (Sarhan, 2017).

Given that the purpose of the present methodology lies within the field of rehabilitation, only the stimuli related to the medical context are analysed here. However, to make these marker layouts more versatile also in other contexts, other stimuli should be considered and introduced in the training data. In a perspective of generalizability, this method with a larger training set could be used to determine the optimal position of markers able to recognize various facial expressions or action units.

Secondly, another potential issue of this approach may be the over-fitting of the feature sets to the training data such that they focus on nuances of this training set but that are not found in future samples. Finally, because the approach is designed as a recursive algorithm, finding the global optimum is challenging and the subsets found do not always guarantee the best results. However, the extracted solutions always allow a recognition rate higher than 95%, demonstrating the effectiveness of this trade-off between robustness and computational time.

379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427

## 4 Conclusion

An automatic technique for finding optimized marker layouts for marker-based facial motion capture is proposed in this paper. Specifically, by using 124 motion captures of 15 different subjects, each one containing the 3D position of 112 markers (109 markers fixed on the facial surface + 3 pasted on a rigid structure attached on the maxilla and used only as a reliable reference), the algorithm identifies the optimal facial marker layouts selecting the subset of measured features (i.e. distances among the 109 facial markers) that best allows for automatic recognition of the facial movements performed. To accomplish this task of “automatic recognition”, a KNN classification technique, preceded by the application of a K-means clustering algorithm, is applied on a different subset of features, containing the distances among each marker and its neighbours. The subsets that best allows for the classification algorithm to automatically recognise the types of movement performed are chosen as the bests, and the marker layouts are extracted from them. Various validations and testing phases have demonstrated the accuracy, robustness and usefulness of the custom approach.

The layouts extracted can be directly used for facial motion acquisitions for guiding the marker application on the facial surface of the patients. Having standardized and optimal marker layouts was a long-standing problem remaining to be resolved in marker-based facial motion capture.

In the future, a deep look at the problem should be taken, in order to overcome the above limitations and improve the soundness and accuracy of automated extraction of marker layouts. For example, the training data may be enhanced by the introduction of others types of movements.

## Acknowledgements

This work was supported by the “Conseil Régional de Picardie” and the “Fondation des Gueules Cassées”. This work uses the Equipex-Figure platform.

## Conflict of Interest Statement

All authors have approved the manuscript, agree with its submission to the journal, and certify that they have NO affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

## 6 References

Adolphs, N., Haberl, H., Liu, W., Keeve, E., Menneking, H., Hoffmeister, B., 2012. Virtual planning of craniomaxillofacial surgery – *Is there still a need for "real" models? Experiences of one centre using commercially available software.* In *XXI Congress EACMFS*, Dubrovnik, Croatia.

Alapati, Y. K., Sindhu, K., 2016. Combining Clustering with Classification: A Technique to Improve Classification Accuracy. *International Journal of Computer Science Engineering (IJCSE)*, 5(6).

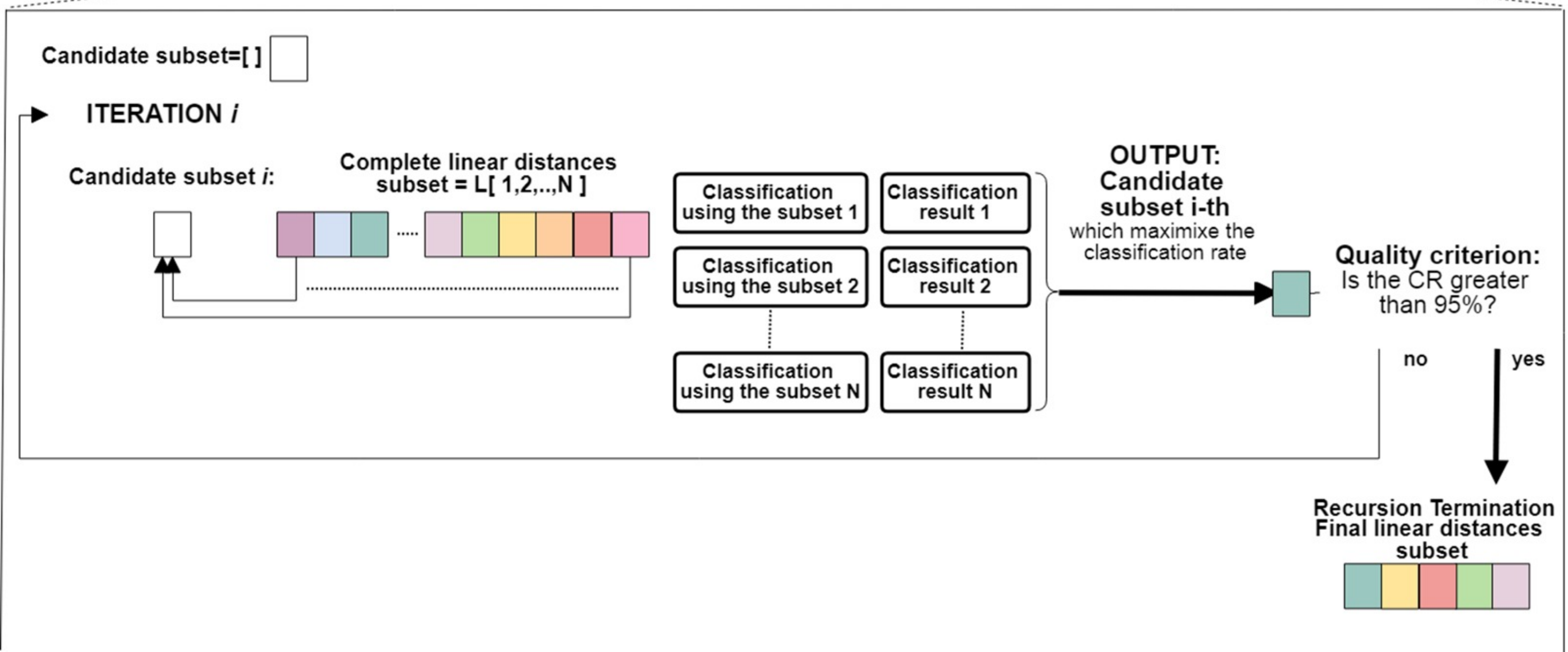
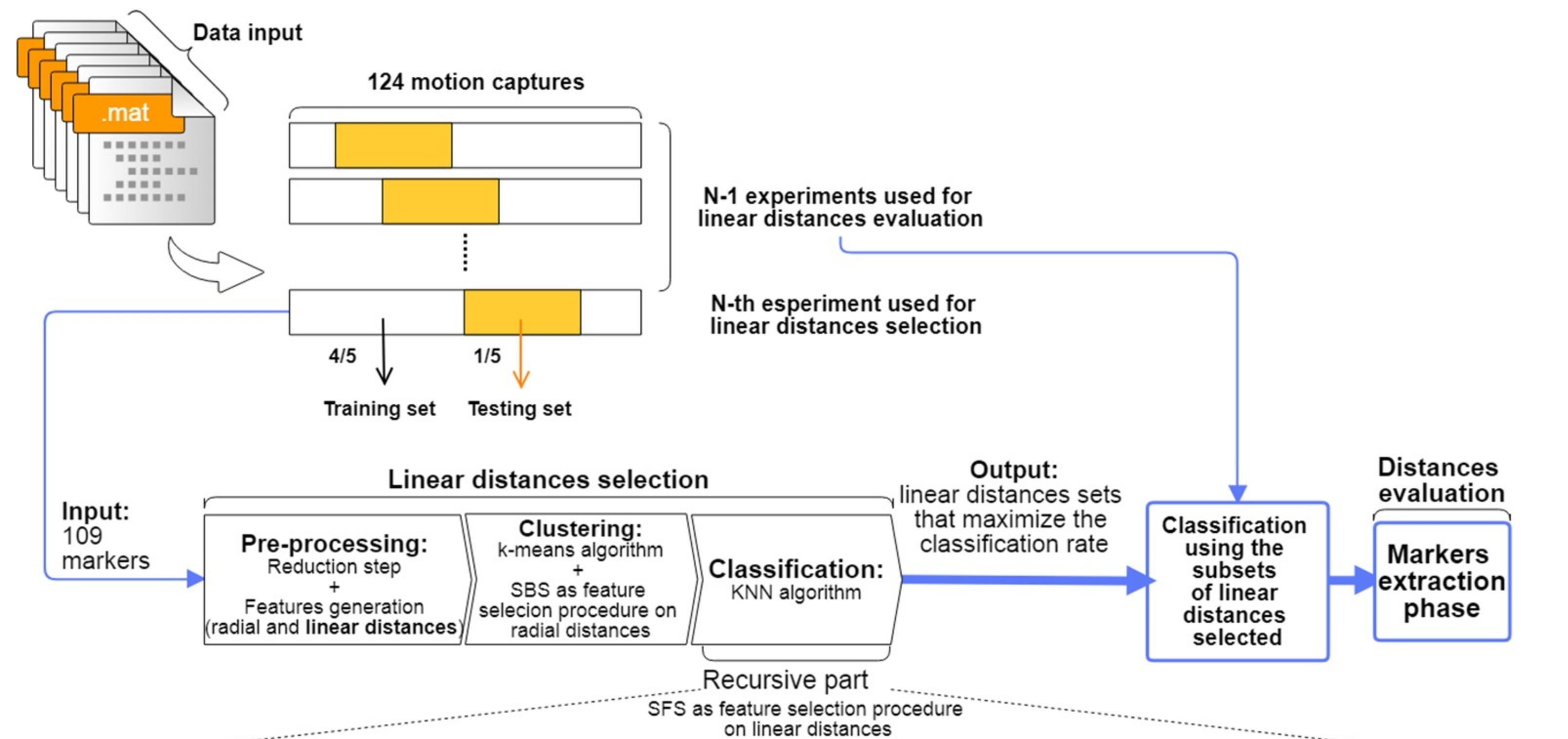
Bajaj-Luthra, A., VanSwearingen, J., Thornton, R. H., Johnson, P. C., 1998. Quantitation of Patterns of Facial Movement in Patients with Ocular to Oral Synkinesis. *Plastic and Reconstructive Surgery*, 101(6), pp. 1473–1480.

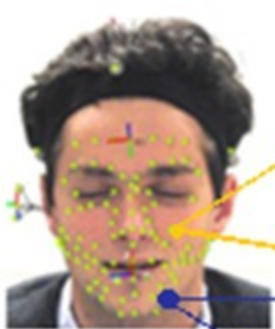
Bargiela-Chiappini, F., Haugh, M., 2010. *Face, Communication and Social Interaction*. Sheffield: Equinox.

- 428 Ben Mansour, K., Sarhan, F.-R., Neiva, C., Godard, C., Devauchelle, B., Marin, F., Dakpé, S., 2014. Analysis  
429 of mimic facial movements based on motion capture. *Computer Methods in Biomechanics and*  
430 *Biomedical Engineering*, 17, pp. 78-79.
- 431 Bianchi, A. F., Ramieri, G. A., Gerbino, G., Verzè, L., 2012. Soft tissue changes after orthodonticsurgical  
432 correction of jaws asymmetry evaluated by 3D surface laser scanner. In *XXI Congress EACMFS*.
- 433 Bianchi, F. B., Verzè, L. & Ramieri, G. A., 2012. Facial mobility after bimaxillary surgery in class III  
434 patients: a three-dimensional study. In *XXI Congress EACMFS*.
- 435 Byrne, P. J., 2004. Importance of facial expression in facial nerve rehabilitation. *Current Opinion in*  
436 *Otolaryngology & Head and Neck Surgery*, 12(4 ), pp. 332-335.
- 437 Chandu, A., Suvinen, T. I., Reade, P. C., Borromeo, G. L., 2005. Electromyographic activity of frontalis  
438 and sternocleidomastoid muscles in patients with temporomandibular disorders. *Oral Rehabilitation*,  
439 32(8), pp. 571-576.
- 440 Di Stadio, A., 2015. Another Scale for the Assessment of Facial Paralysis? ADS Scale: Our Proposition,  
441 How to Use It. *Journal of Clinical and Diagnostic Research*, 9(12).
- 442 Furukawa, Y., Ponce, J., 2010. Dense 3D Motion Capture from Synchronized Video Streams. *Image and*  
443 *Geometry Processing for 3-D Cinematography*, 5, pp. 193-211.
- 444 Gupta, A., Singh, K., Tandon, R., Chauhan, A., Kulshrestha, R., Trivedi, H., 2017. Functional  
445 Electromyographic Activities of Facial Muscles in Different Growth Patterns. *SM Dent Oral Disord*, 1(1),  
446 pp. 1004-1009.
- 447 Henstrom, D. K., Skilbeck, C. J., Weinberg, J., Knox, C., Cheney, M. L., Hadlock, T. A., 2011. Good  
448 correlation between original and modified House Brackmann facial grading systems. *The*  
449 *Laryngoscope*, 121, pp. 47-50.
- 450 Hontanilla, B., Aubá, C., 2008. Automatic three-dimensional quantitative analysis for evaluation of  
451 facial movement. *Jurnal of Plastic, Reconstructive & Aesthetic Surgery*, 61(1), pp. 18-30.
- 452 House, J. W., 1983. Facial nerve grading systems. *The Laryngoscope*, 93, pp. 1056-1069.
- 453 Ju, X., Khambay, B., O'Leary, E., Al-Anezi, T., Ayoub, A., 2012. Evaluation of the Reproducibility of Non-  
454 verbal Facial Animations. *Lecture Notes in Computer Science*, 7378, pp. 184-193.
- 455 Le, B., Zhu, M., Deng, Z., 2012. Marker optimization for facial motion acquisition and deformation. *IEEE*  
456 *Trans. Visualization and Computer Graphics*, 19(11), pp. 1859-1871.
- 457 Reitzen, S. D., Babb, J. S., Lalwani, A. K., 2009. Significance and rehabilitee of the House-Brackmann  
458 grading system for regional facial nerve function. *Otolaryngol Head Neck Surg.*, 140, pp. 154-158.
- 459 Sarhan, F. R., 2017. *Quantification des mouvements de la mimique faciale par motion capture sur une*  
460 *population de volontaires sains*, Université de Technologie de Compiègne.
- 461 Sforza, C., Galante, D., Shirai, Y. F., Ferrario, V. F., 2010. A three-dimensional study of facial mimicry in  
462 healthy young adults. *Journal of Cranio-Maxillofacial Surgery*, 38(6), pp. 409-415.
- 463 Sjogreen, L., Lohmander, A., Killiaridis, S., 2010. Exploring quantitative methods for evaluation of lip  
464 function. *Oral Rehabilitation*.
- 465 Van Boxtel, A., 2010. *Facial EMG as a Tool for Inferring Affective States*. Eindhoven, The Netherlands.

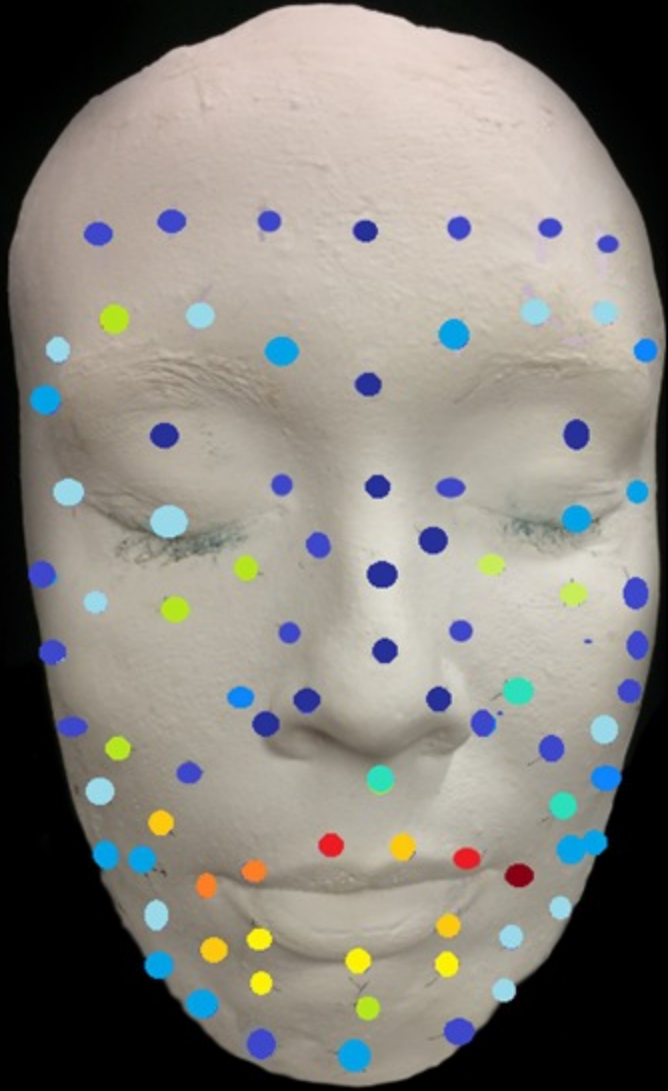






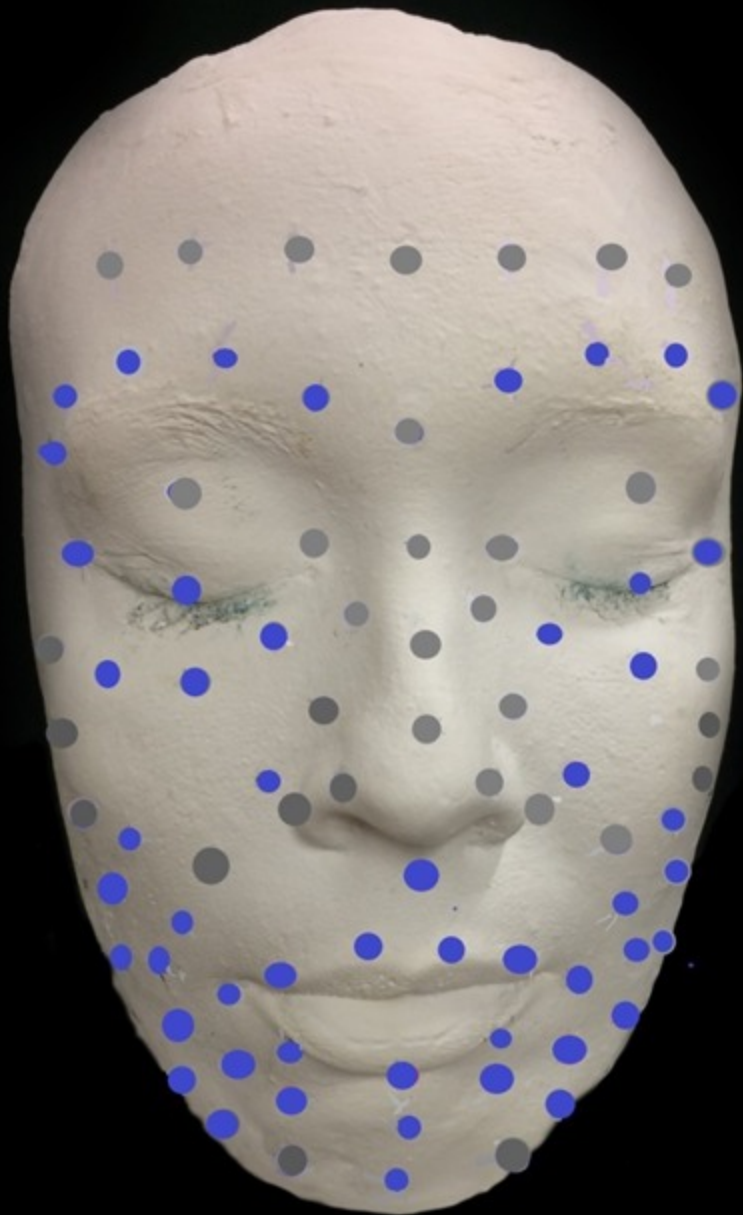


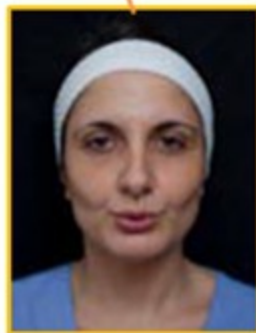
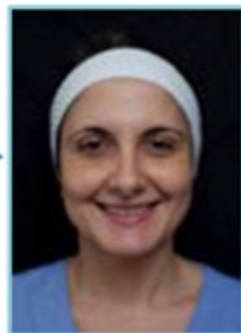
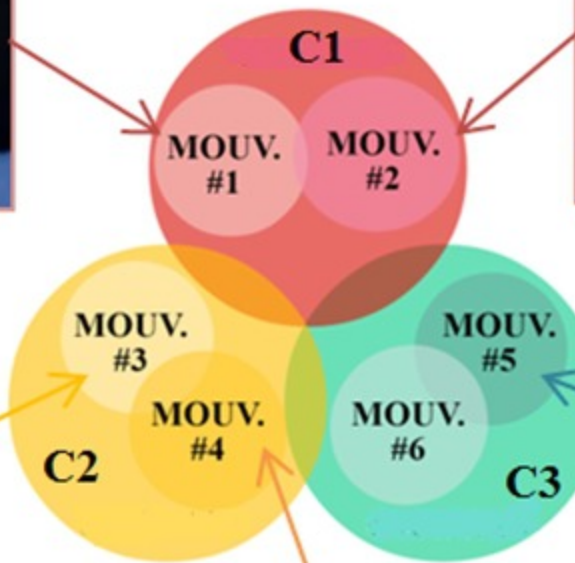
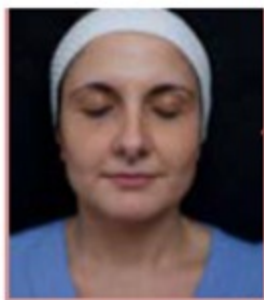
Max



Min







# Plot of k-means clusters

