Partial Occlusion in Facial Expression Recognition: A Systematic Literature Review

Ana Sofia Rodrigues
CeDRI, Instituto Politécnico
de Bragança, Portugal
Email: ana-rodrigues@ipb.pt

Júlio Castro Lopes
CeDRI, SusTEC
Instituto Politécnico de Bragança, Portugal
Email: juliolopes@ipb.pt

João Vieira
CeDRI, Instituto Politécnico
de Bragança, Portugal
Email: a41701@alunos.ipb.pt

Rui Pedro Lopes
CeDRI, SusTEC
Instituto Politécnico de Bragança, Portugal
Email: rlopes@ipb.pt

2024 IEEE 3rd International Conference on Intelligent Reality (ICIR) | 979-8-3315-3442-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICIR64558.2024.10976805

***Abstract*—This study presents a Systematic Literature Review (SLR) examining Facial Expression Recognition (FER) under partial occlusion conditions. Analyzing a diverse array of studies, it identifies prevalent methodologies and challenges within the field. The review reveals a significant use of Gabor filters for feature extraction and classification, highlighting their efficiency in FER research. Additionally, black patches emerge as the primary method for simulating occlusions, which reflects the scarcity of datasets containing such occlusions. The complexity of FER under occlusion conditions is highlighted by the inconclusive findings on the relative importance of facial regions like the mouth and eyes in expression recognition. The results point to the need for more research to comprehend the complex role of different facial regions in conveying emotions. Furthermore, the review emphasizes the importance of developing larger and more diverse datasets, that encompass a broader range of facial expressions and occlusion types, to advance the development and evaluation of FER systems in real-world scenarios.**

***Index Terms*—Computer Vision, Facial Expression Recognition, Occlusion, Partial Occlusion, Systematic Literature Review.**

I. Introduction

Facial expression is one of the most intuitive, all-encompassing, and effective ways to communicate inner emotions in emotional communication, which takes a fundamental role in daily activity for humans. Facial Expression Recognition (FER) is a computer vision research topic that aims to automatically analyze facial expressions from images, to identify the emotional states of people. With its developments over the years, FER is currently being applied in many different fields, such as human-computer interaction [1], healthcare [2], education [3] and many more.

In recent times, the domain of serious games has increasingly leveraged emotion recognition, presenting a challenge in Affective Computing [4]. Serious games aim to replicate real-life scenarios in the gaming environment, using industry approaches to create engaging simulations. These games are created to entertain, educate, and train, promoting the absorption of concepts and psychomotor skills through interactive experience [5]. In this context, FER emerges as a fundamental tool for understanding the emotional responses of the players, thereby enhancing the challenge of the game, improving interaction quality, and enriching the overall gaming experience.

The complexity of extracting distinctive features from occluded facial areas is compounded by the presence of occlusion, leading to difficulties such as inaccurate feature localization, imprecise face alignment, or errors in face registration. Some occlusion parameters that pose challenges for FER include eye or mouth occlusion, where expressions involving eye movements or changes, such as surprise or happiness, may be difficult to recognize. Similarly, partial or self-occlusion, such as by masks or sunglasses, can obscure relevant facial features, making expression recognition challenging. Addressing these occlusion parameters is essential for enhancing the effectiveness of FER systems.

II. Methodology

The Systematic Literature Review (SLR) is a crucial process for comprehending the State Of The Art (SOTA) in a field and identifying the gaps that call for additional research. The purpose of this research is to conduct an extensive SLR in order to present the main recent advances in algorithms and research, focused on FER with partial occlusion. A SLR aims to improve the precision in determining the applicability of papers and the review of the literature while lowering the subjectivity of researchers in the process. It facilitates researchers' work by providing a comprehensive analysis of what has already been accomplished in the field, as well as what has not been accomplished, suggesting future work.

The first step of an SLR is the planning phase, which consists of defining the bibliographic databases and the selection of query terms. To ensure the most relevant and accurate results, the data was extracted from Scopus and Web of Knowledge (WoK) databases.

After defining the databases, a query was performed to retrieve the most relevant works ("Facial Expression Analysis" OR "Facial Expression Recognition" AND "Partial Occlusion"). This query was used in Scopus and WoK and retrieves the results by matching the title, keywords, and also the

Authorized licensed use limited to: b-on: Instituto Politecnico de Braganca. Downloaded on June 06,2025 at 15:16:53 UTC from IEEE Xplore. Restrictions apply.

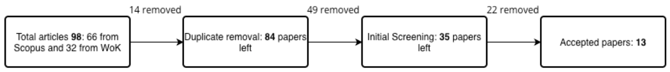

Fig. 1: Paper selection - flow chart.

abstract. The results were further limited to English-language papers written between 2002 and September 25, 2023.

In the first stage, a initial selection was performed, where the articles were classified as ‘accepted’, ‘rejected’, or ‘duplicated’. The aim is to curate scientific literature showcasing recent advancements within the subject under review. Furthermore, priority was given to applied research that clarifies the potential applications and beneficiaries of the technology. In this context, specific requirements were delineated, to ensure the reliability of documents for this systematic review. To ensure that the most relevant papers were kept in this review, several quality assessment questions were performed in each of the papers.

- Is there a clear objective of the research?
- Partial occlusion of the face is included in the research?
- Is the hidden places trully occluded for facial expression recognition?
- Is the paper based on research?

From the 98 extracted articles in the first stage (66 from Scopus and 32 from WoK), 14 articles were duplicated. From the remaining 84, based on a initial screening 49 papers were removed, resulting on a total of 35 papers left. From these 35, 13 papers were classified as ‘accepted’, according to the quality assessment questions (only papers with affirmative responses in all questions, were included). This procedure is graphically represented by Figure 1.

The definition of research questions is also an important step in the planning phase. These questions are designed to frame the research and offer specific and related information. The aim of this paper is to identify advances in FER when faced with partial occlusion conditions. This study systematically reviews the SOTA Artificial Intelligence (AI) methods, tailored to address the challenges posed by occluded facial environments, acknowledging the growing demand for robust FER solutions. Additionally, ongoing research is investigating AI implementations for FER in occluded environments and assessing the impact of occlusions on FER (RQ1). This seeks to understand which types of occlusions most significantly affect FER performance, shedding light on the challenges faced by AI methods in the recognition of facial expressions under various occlusion scenarios. As part of this exploration, the aim is to discover the algorithms used to simulate partial occlusion in FER datasets (RQ2), despite the absence of any datasets with partial occlusion. Understanding these algorithms can provide insights into the techniques used to generate realistic occluded facial data, crucial for training robust FER models in occlusion-rich environments. Furthermore, the research endeavors to identify the most popular dataset for FER with partial occlusion (RQ3). This analysis depicts the primary dataset(s) utilized in research focused on FER under occlusion conditions, offering valuable insights into benchmarking practices and informing future research directions. Lastly, this review looks to identify the main open lines of research or issues in this domain (RQ4). By exploring unresolved issues and gaps in the field of FER under occlusion environments, this paper contributes to the advancement of FER methodologies and address critical gaps in understanding occlusion impact on FER systems.

The review aims to uncover the most effective techniques, explore the methodologies employed, and delve into other pertinent aspects within the field. Based on the aforementioned objective, the following research questions were formulated (RQs): RQ1: Which type of occlusion most penalizes FER? RQ2: What algorithms are used to simulate the partial occlusion on FER datasets? RQ3: What is the most popular dataset for FER with partial occlusion? RQ4: What are the main open lines of research or issues in this domain?

In the subsequent stage, a Quality Assessment Question checklist was included, that enhances the research process by scrutinizing the reliability, validity, and credibility of gathered data. Four research questions were created, in order to evaluate the relevance of the papers retrieved by the string query: QA1: Is there a clear objective of the research? - This question attempts to discern the overarching goal or purpose of the research, seeking to understand the specific objective that the researchers aimed to achieve through their investigation. QA2: Partial occlusion of the face is included in the research? - This question tries to reveal whether the research incorporates the aspect of partial occlusion of the face in their study, questioning the inclusion or exclusion of this specific factor in the investigation. QA3: Are the hidden places truly occluded for FER? - This question looks to explore the validity of occlusion about FER. Essentially, it is intended to validate if the face is hidden, for example, through the use of surgical masks or VR glasses. QA4: Is the paper based on research? - This question aims to determine whether the research described is a research article, to exclude reviews or surveys.

The four questions were devised with binary ‘yes’ or ‘no’ responses, assigning a weight of 1.0 to yes and 0 to no. Only research papers scoring 4 points were accepted (meets all the requirements), while all others were rejected.

To delineate the datasets utilized (DE1), organize the methodologies employed for facial expression classification and occlusion techniques (DE2), assess the impact of various types of partial occlusions on recognition (DE3), identify challenges encountered when dealing with partially occluded faces (DE4), suggest future areas for improving partial occlusion in FER (DE5), evaluate whether the developed system classified images in real-time (DE6), determine which ML models were utilized (DE7), investigate the types of occlusions addressed in the research (DE8), and analyze the number and types of emotions evaluated (DE9), the following questions were formulated for data extraction: DE1: Which datasets were used and which accuracy was obtained? DE2: Which specific methods or techniques were used to perform the occlusion? DE3: How do different types of partial occlusions impact

Authorized licensed use limited to: b-on: Instituto Politecnico de Braganca. Downloaded on June 06,2025 at 15:16:53 UTC from IEEE Xplore. Restrictions apply.

the recognition? DE4: What challenges were found when dealing with partially occluded faces? DE5: Did the research paper suggest future areas of improvement for addressing partial occlusion in FER? DE6: Did the system developed for the research classify the images in real time? DE7: What ML models were used?DE8: What kind of occlusions are addressed in the investigation? DE9: How many emotions are evaluated? And which?

Following the outlined procedures, a total of 13 articles have been identified for consideration in this study.

III. Results

Understanding the intricacies of partial occlusion is crucial in elucidating its causes, impacts, and potential implications across several domains. This section, centers the attention on inquiries dedicated to comprehending the methodologies utilized in facial expression classification, elucidating the techniques employed for inducing partial occlusion in facial structures, examining the algorithms utilized for expression classification, and scrutinizing the datasets utilized by researchers. Additionally, this investigation involves investigating the variety of occlusions mentioned in the literature and evaluating their resulting effects. It aims to enhance our understanding of their effects on FER systems by highlighting the diverse amount of occlusions and their associated implications through research endeavors.

A. Collected Data Results

Occurrences of upper occlusions, lower occlusions, and other types of occlusions documented in the studies are shown in Tables I, II, and III, respectively. These tables include the title of the paper, the type of occlusion with a visual representation, the method used to simulate partial occlusion (if applicable), and the author's exploration of multiple methods, with the proposed method highlighted in bold for its superior performance compared to others, and the respective accuracy.

In the studies performed by Cheng et al. [6], Li et al. [7] and Liu et al. [8], it was revealed that striking eye features are crucial for FER, using the JAFFE dataset. The research suggests that the features surrounding the eyes are vital in distinguishing facial expressions, which is evidenced by the lower performance achieved, compared to lower occlusion. In this context, Cheng et al. stand out with the lowest accuracy value in eye occlusion (77.14%), compared to 82.86% in lower occlusion. At the same time, Li et al. [7] achieved the highest accuracy value in upper occlusion (86.97%), compared to 90.90% in lower occlusion, with the JAFFE dataset.

On the other hand, Li et al. [7], Poux et al. [9], Huang et al. [10], Buciu et al. [11], Zhi et al. [12] and Mushfieldt et al. [13] results show that the mouth region contains more decision-making information for emotional expression than the eyes. The absence of mouth region features results in a decrease in classifier accuracy when compared with upper occlusion supports this assertion. This scenario is emphasized in the study by Mushtfield et al. [13], which reported approximately 45% accuracy in lower occlusion and about 68% in upper occlusion, representing the lowest values. Conversely, Zhi et al. achieved the highest accuracy values, reaching 91.4% for lower occlusion and 93.3% for upper occlusion.

This incongruence can arise due to the similarity in mouth and eye features across various types of expressions, resulting in discrepancies in the obtained values. Also, it must be considered that different facial areas are involved in different emotions, leading to variations in facial expressions. For instance, when the eyes are covered, emotions like anger and fear rely more on the discriminative features of the mouth region than on the eyes. Happiness is more easily distinguishable by the eyes than by the mouth. When the mouth is obscured, the 'neutral' expression can be misinterpreted as 'fear' or 'sadness' [11].

Other experiments were also performed by Li et al. [7] (left and right occlusions), M.D and Rahiman [14] (left and right eyebrow, nose and lip corner occlusions), Zhi et al. [12] (nose occlusion) and Mushfieldt et al. [13] (left and right occlusions). Li et al. [7] obtained the best accuracy values for left occlusion with 89.69% and 87.79%, employing the JAFFE dataset. In contrast, the lowest performance was reported by Mushfieldt et al. [13] with 73.00% and 56.00%, respectively.

B. Research Questions Results

RQ1: Which type of occlusion most penalizes FER?

The findings of the studies are not clear regarding the indication of which type of occlusion has the most negative impact on FER. While occlusion in the eye region tended to result in lower performance in some studies, occlusions in other facial regions such as the mouth also showed significant impacts on FER accuracy. For example, Cheng et al. [6] achieved a lower accuracy value in eye occlusion (77.14%) compared to 82.86% in lower occlusion, while Mushtfield et al. [13] reported approximately 45% accuracy in lower occlusion and about 68% in upper occlusion, with Zhi et al. [12] achieving the highest accuracy values for both lower and upper occlusions. These results suggest that different types of occlusion may have varying degrees of impact on FER performance and further research may be needed to determine the most detrimental type of occlusion.

RQ2: What algorithms are used to simulate the partial occlusion on FER datasets?

In this investigation, it is clear that many researchers favor the use of Gabor filters for both feature extraction and facial expression classification. This method has proven effective across various occlusion scenarios, including lower, upper, and other types of occlusions. The adaptability of Gabor filters in handling occlusions strengthens their position in the field of FER. Regarding the methods used to simulate occlusions in datasets, the majority of studies opt for black patches. Notably, only one study, conducted by Yang et al. [19], utilizes automatic facial masks. This diversity in techniques underscores the importance of exploring different approaches for generating occluded datasets in FER research. Such methods become essential due to the absence of occlusions in default datasets and the lack of datasets specifically designed

Authorized licensed use limited to: b-on: Instituto Politecnico de Braganca. Downloaded on June 06,2025 at 15:16:53 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Upper Occlusions

Paper	Type of occlusion	Simulation of partial occlusion	Methods	Datasets	Accuracy
Cheng et.al [6]	 (a) Eyes occlusion  (b) Upper occlusion	Not-described	• Gabor • LGBPHS • M-LGBPHS • **Gabor + Deep nonlinear network with 3 layers (proposed method)**	JAFFE	• Figure Ia: 82.86% • Figure Ib: 77.14%
Li et al. [7]	 (c) Eyes occlusion  (d) Upper occlusion	Not-described	• Gabor • F-LGBPHS • **Gabor Filter + GLCM + KNN + 10-fold cross-validation (proposed method)**	• JAFFE • RAF-DB	Figure Ic • 86.97% • 84.55% Figure Id • 84.12% • 81.12%
Houshmand et al. [15]		• Detection of the face through grayscale images generated with modified HOG and linear SVM; • Estimation of 68 facial landmarks using the approach described in [16]; • Application of the VR patch following the dimensions of a Samsung Gear VR headset.	• VGG-Face (from scratch) + 5-fold cross-validation • ResNet50 (from scratch) + 5-fold cross-validation • **VGG-Face (transfer learning) + 5-fold cross-validation (proposed method)** • ResNet50 (transfer learning) + 5-fold cross-validation • VGG-Face + 5-fold cross-validation	• FER+ • AffectNet • RAF-DB	• 79.98% • 50.13% • 73.37%
Poux et al. [9]		Not-described	• Symmetric auto-encoder + optical flow + 10-fold cross-validation + MSE. • Symmetric auto-encoder + optical flow + 10-fold cross-validation + Wing. • **Symmetric auto-encoder + optical flow + 10-fold cross-validation + EndPoint (proposed method)**.	CK+	87.10%
Petrou et al. [17]		• Detection of five facial landmarks (two for the center of each eye, one for the center of the nose, and two for the right and left side of the mouth); • With the landmarks of the nose and the eyes, and the distances of the algorithm described in [18], a rectangle is drawn on top of each image; • Upper part of the face is hidden to simulate the inclusion of VR headsets.	• Mini-Xception (pre-trained) • Mini-Xception (pre-trained + unfreeze last layer) • **Mini-Xception (pre-trained + unfreeze all layers) (proposed method)** • Mini-Xception (trained from scratch)	The authors developed their own dataset by combining five online datasets (FER 2013, Jafar Hussain Human, Unsplash, Pexels and Pixabay).	69.00%
Liu et al. [8]		Addition of black masks with different positions of the expression region (eyes, the mouth, left side of the face and right side of the face)	• Gabor Histogram • LBP Histogram • **Gabor multi-orientation features fusion + LGBPHS (proposed method)**	JAFFE	85.53%
Huang et al. [10]		• To simulate occlusion, graphically generated eyeglasses, medical masks, and random region masks were superimposed on un-occluded facial expression sequences; • AAM locates the facial points in each frame, resulting in the same generation process for eye, mouth, and lower-face occlusions for the next frames; • The distance between frames at the top of the nose is defined for random occlusions: after determining the position in the first frame, the patch is placed after the computed distance is adjusted in the next frame.	• STLBP • EdgeMap • FSE • CFD • CFD-OD • **CFD-OD-WL (proposed method)**	CK+	93.00%
Buciu et al. [11]		Superimposed black rectangles around the eyes and mouth regions to occlude them partially.	• Gabor + CSM • **Gabor + MCC**	• JAFFE • CK	• 84.00% • 92.30%
Zhi et al. [12]		To simulate partial occlusion on the facial images, an eye mask, nose mask, and mouth mask were made.	• **GSNMF (proposed method)** • SNMF • Laplacianfaces	CK	93.30%
Mushfieldt et al. [13]		Black patches were applied to the eyes, mouth, and left and right sides of the face, superimposed.	• Gabor • DNMF • **Viola-Jones face detection algorithm + GMMs + LBP + SVM (proposed method)**	CK	68.00%
Rodrigues et al. [18]		• Obtaining 5 facial expression landmarks (two for the center of each eye, one for the center of the nose, and two for the right and left side of the mouth), using MTCNN; • Using the landmarks of the nose and eyes and the distances of the algorithm created by the authors, a rectangle is drawn on top of each image. This methodology simulates the presence of VR goggles.	• ResNet18 + MTCNN • VGG19 + MTCNN • **Combined (ResNet18 + VGG19) + MTCNN (proposed method)**	FER 2013	64.90%

Authorized licensed use limited to: b-on: Instituto Politecnico de Braganca. Downloaded on June 06,2025 at 15:16:53 UTC from IEEE Xplore. Restrictions apply.

TABLE II: Lower Occlusions

Paper	Type of occlusion	Simulation of partial occlusion	Methods	Datasets	Accuracy
Cheng et.al [6]	 (a) Mouth occlusion  (b) Lower occlusion	Not-described	• Gabor • LGBPHS • M-LGBPHS • **Gabor + Deep nonlinear network with 3 layers (proposed method)**	JAFFE	• Figure IIa: 82.86% • Figure IIb: 82.86%
Li et al. [7]	 (c) Mouth occlusion  (d) Lowew occlusion	Not-described	• Gabor • F-LGBPHS • **Gabor Filter + GLCM + KNN + 10-fold cross-validation (proposed method)**	• JAFFE • RAF-DB	Figure IIc • 90.90% • 81.73% Figure IId • 86.66% • 74.90%
Poux et al. [9]	 (e) Lower occlusion  (f) Mouth occlusion	Not-described	• Symmetric auto-encoder + optical flow + 10-fold cross-validation + MSE. • Symmetric auto-encoder + optical flow + 10-fold cross-validation + Wing. • **Symmetric auto-encoder + optical flow + 10-fold cross-validation + EndPoint** (proposed method).	CK+	• Figure IIe: 70.20% • Figure IIf: 80.10%
Yang et al. [19]		Automatic wearing of a face mask, which automatically adds face masks that are shaped according to the orientation.	• VGG • MobileNet • RAN • ACNN • OADN • **Two-stage attention model that consists of a binary deep classifier and a face-mask-aware FER deep classifier (proposed method)**.	• M-LFW-FER • M-KDDI-FER	• 87.92% • 90.01%
Liu et al. [8]		Addition of black masks with different positions of the expression region (eyes, the mouth, left side of the face and right side of the face).	• Gabor Histogram • LBP Histogram • **Gabor multi-orientation features fusion + LGBPHS (proposed method)**	JAFFE	92.11%
Huang et al. [10]	 (g) Mouth occlusion  (h) Mouth occlusion (mask simulation)	• To simulate occlusion, graphically generated eyeglasses, medical masks, and random region masks were superimposed on un-occluded facial expression sequences; • AAM locates the facial points in each frame, resulting in the same generation process for eye, mouth, and lower-face occlusions for the next frames; • The distance between frames at the top of the nose is defined for random occlusions: after determining the position in the first frame, the patch is placed after the computed distance is adjusted in the next frame.	• STLBP • EdgeMap • FSE • CFD • CFD-OD • **CFD-OD-WL (proposed method)**	CK+	• Figure IIg: 73.54% • Figure IIh: 79.08%
Buciu et al. [11]		Superimposed black rectangles around the eyes and mouth regions to occlude them partially.	• Gabor + CSM • **Gabor + MCC**	• JAFFE • CK	• 83.50% • 87.20%
Zhi et al. [12]		To simulate partial occlusion on the facial images, an eye mask, nose mask, and mouth mask were made.	• **GSNMF (proposed method)** • SNMF • Laplacianfaces	CK	91.40%
Mushfieldt et al. [13]		Black patches were applied to the eyes, mouth, and left and right sides of the face, superimposed.	• Gabor • DNMF • **Viola-Jones face detection algorithm + GMMs + LBP + SVM (proposed method)**	CK	45.00%

Authorized licensed use limited to: b-on: Instituto Politecnico de Braganca. Downloaded on June 06,2025 at 15:16:53 UTC from IEEE Xplore. Restrictions apply.

TABLE III: Other Occlusions

Paper	Type of occlusion	Simulation of partial occlusion	Methods	Datasets	Accuracy
Li et al. [7]	 (a) Left occlusion  (b) Right occlusion	Not-described	• Gabor • F-LGBPHS • **Gabor Filter + GLCM + KNN + 10-fold cross-validation (proposed method)**	• JAFFE • RAF-DB	Figure IIIa: • 89.69% • 87.73% Figure IIIb: • 89.45% • 81.83%
M.D and Rahiman [14]	 (c) Left eyebrow occlusion  (d) Right eyebrow occlusion  (e) Left nose occlusion  (f) Right nose occlusion  (g) Left lip occlusion  (h) Right lip occlusion	The occlusion was applied using MatLab code.	Combination of LBP and SAX feature extraction + Ensemble bag classifier (with supervised learning) (proposed method).	Fused database (JAFFE and YALE)	• Figure IIIc: 88.81% • Figure IIId: 93.47% • Figure IIIe: 93.25% • Figure IIIf: 88.81% • Figure IIIg: 90.93% • Figure IIIh: 93.25%
Zhi et al. [12]		To simulate partial occlusion on the facial images, an eye mask, nose mask, and mouth mask were made.	• **GSNMF (proposed method)** • SNMF • Laplacianfaces	CK	94.00%
Mushfieldt et al. [13]	 (i) Right occlusion  (j) Left occlusion	Black patches were applied to the eyes, mouth, and left and right sides of the face, superimposed.	• Gabor • DNMF • **Viola-Jones face detection algorithm + GMMs + LBP + SVM (proposed method)**	CK	• Figure IIIi: 73.00% • Figure IIIi: 56.00%

to incorporate such occlusions. While existing datasets may contain items like glasses or scarves, they often fall short of fully replicating the types of occlusions encountered in real-world scenarios.

RQ3: What is the most popular dataset for FER with partial occlusion?

To recognize facial expressions with partial occlusion, the Japanese Female Facial Expression (JAFFE) dataset is the most frequently used dataset. The JAFFE dataset is a well-known collection of 213 grayscale images of Japanese women expressing six basic emotions: neutral, happy, sad, angry, surprised, and disgusted. It should be noted that the JAFFE dataset is free from instances of occlusion, unlike other datasets presented. The lack of occluded data indicates a discrepancy in the current datasets, as there are currently no datasets that specifically address occlusions in FER tasks. As such, researchers tackle this limitation, by proposing methods to simulate occlusion, as mentioned in response to RQ2.

RQ4: What are the main open lines of research or issues in this domain?

Many important areas need further exploration in FER with partial occlusion. The first step is to explore how occlusion

Authorized licensed use limited to: b-on: Instituto Politecnico de Braganca. Downloaded on June 06,2025 at 15:16:53 UTC from IEEE Xplore. Restrictions apply.

influences different facial regions and comprehend its consequences for facial reconstruction accuracy. Examining the significance of specific facial features for accurate expression recognition when occluded, may be necessary. Another avenue for research is to develop novel algorithms or techniques that can improve FER performance in partial occlusion conditions. Investigating approaches that are resilient to occlusions and can effectively extract meaningful features from obscured facial images, may also be required. Moreover, there is a call to establish standardized benchmarks or datasets that are specifically designed to evaluate FER systems with occlusions. The lack of sufficient occluded data in current datasets can make it difficult to evaluate the performance of FER models under realistic conditions. A more comprehensive evaluation framework for researchers in this field can be achieved by developing dedicated datasets with diverse occlusion scenarios. Finally, it is necessary to investigate the generalization capabilities of FER models trained on datasets with occlusions, in real-world scenarios.

IV. Discussion

This section, emphasizes the most significant research findings, particularly those that lead to significant outcomes and insights. Furthermore, the prevalence of various types of occlusions was explored, as well as the methods used to simulate partial occlusion in various scenarios. In addition, the ML networks that are most frequently used for classification were examined, spotlighting also the type of occlusion that is most often emphasized by the reviewed literature. Regarding the most significant research studies in terms of performance accuracy, five studies have contributed significantly to advance the understanding of FER under occlusion:

- Zhi et al. [12] developed an algorithm, Graph-Preserving Sparse NMF (GSNM), with the capability to recognize facial expressions. The authors improved the traditional NMF approach by introducing sparsity and maintaining sample neighborhood information through graph embedding theory. Furthermore, GSNMF incorporated DNMF algorithm into its framework, allowing for supervised learning. To solve convergence problems in NMF algorithms, they also introduced a projected gradient framework with additional constraints to ensure stationarity;
- Huang et al. [10] proposed a methodology that includes three crucial components: facial representation, occlusion detection, and feature fusion. To ensure robust facial representation, the authors derived six feature vectors from facial components, incorporating temporal cues using dynamic texture and structural shape feature descriptors. Sparse representation and residual statistics were used to detect occlusions, which ensured independence from facial identity. To solve the issue of dimensionality caused by concatenating feature vectors, the authors presented multiple feature fusion, which consists of a fusion module and weight learning;
- Buciu et al. [11] developed a methodology that involved combining each image from the database with a variety of Gabor filters, each with different orientations and frequencies, to obtain localized facial expression features. The representation of a face model was based on automatically tracked focal points, and facial expression classification was achieved through decision-level fusion. Instead of manually selecting specific regions, features were extracted by applying Gabor filters to the entire face;
- Li et al. [7] created an algorithm to extract Gabor feature statistics by analyzing the spatial distribution of facial features. In order to compensate for the lack of block Gabor features, a gray-level co-occurrence matrix (GLCM) was introduced into the expression recognition domain, where pixel association was not present. The block Gabor feature statistics were finally combined with texture features extracted by GLCM to produce a linear superposition;
- Liu et al. [8] developed a novel method for FER under partial occlusion, utilizing Gabor multi-orientation features fusion and Local Gabor Binary Pattern Histogram Sequence (LGBPHS). Initially, Gabor filters were employed to extract multi-scale and multi-orientation features. Subsequently, the magnitudes of Gabor features from different orientations within the same scale were fused based on the fusion rule outlined in the paper. The fused features were then encoded using the Local Binary Pattern (LBP) operator. Finally, the fused image was partitioned into several non-overlapping rectangle units of equal size, and the histogram of each unit was computed and combined to form facial expression features.

Furthermore, it is notable that, regarding occlusions, the authors place more emphasis on occlusions affecting the eyes. This focus suggests a recognition of the critical role of the eyes in conveying facial expressions, emphasizing this kind of occlusion in relation to others.

V. Conclusion

This paper employs an SLR methodology to achieve a comprehensive understanding of the current SOTA and to present the primary advancements in algorithms and research from around the world, specifically focusing on FER under conditions of partial occlusion. The studies were divided into three types: upper occlusions, encompassing all occlusions in the eye region; lower occlusions, covering occlusions in the mouth region; and other types, which include occlusions that do not fit into the first two groups. The most commonly employed technique for feature extraction and classification was the utilization of Gabor filters, reflecting its prevalence and effectiveness in FER research. Using Gabor filters as feature extraction and classification techniques highlights the significance of texture-based approaches in FER research. In situations with partial occlusion, texture features play a crucial role in capturing facial expressions, as suggested by this. Regarding the partial occlusion simulation, the majority of authors utilized black patches to simulate occlusion, due to the absence of datasets containing such occlusions. Using this method for simulating partial occlusions highlights the

Authorized licensed use limited to: b-on: Instituto Politecnico de Braganca. Downloaded on June 06,2025 at 15:16:53 UTC from IEEE Xplore. Restrictions apply.

need for more diverse and realistic occlusion scenarios in dataset creation. Exploration of new techniques to create occlusions that better represent real-life conditions, such as facial accessories, lighting variations, or environmental factors, could be part of this.

JAFFE was the dataset most frequently used in the studies reviewed. However, it is essential to recognize that this dataset does not contain instances of occlusion, which underscores the limitations of available datasets for evaluating FER under occlusion conditions. Despite its use as a benchmark in FER research, the JAFFE limited size and absence of occluded data highlight the ongoing need for larger and more diverse datasets in this field. The development and evaluation of FER systems under real-world conditions could be significantly advanced by creating datasets that cover a wider range of facial expressions, occlusion types, and environmental factors.

The research on FER highlights an intriguing aspect by highlighting the importance of facial regions in discerning expressions. The mouth region is emphasized in some studies as crucial in conveying emotional expressions. In these studies, it has been found that subtle mouth movements and configurations, such as lip curvature or tension, can be valuable clues for recognizing specific emotions. In contrast, other research suggests that the eye region is more crucial in determining facial expressions. Eyes are renowned for their ability to express emotions through changes in eyelid position or eyebrow movement. According to these studies, the eyes may have more discriminative information to accurately identify emotions compared to other facial regions. The significance given to the mouth and eye regions in interpreting facial expressions remains unclear. The complexity of facial expressions, in which multiple facial regions are involved, may be the cause of this ambiguity. Cultural factors, individual differences, and the specific context of observed expressions may affect the importance of different facial regions. Furthermore, the analysis becomes more complicated due to the similarity in features across different expressions. It can be challenging to determine the relative importance of each region in expression recognition when certain expressions have common features in both the mouth and eye regions.

Acknowledgment

This work was supported by national funds through FCT/MCTES (PIDDAC): CeDRI, UIDB/05757/2020 (DOI: 10.54499/UIDB/05757/2020) and UIDP/05757/2020 (DOI: 10.54499/UIDP/05757/2020); and SusTEC, LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020).

References

[1] K. Vinutha, M. Kumar Niranjan, J. Makhijani, B. Natarajan, V. Nirmala, and T. R. Vijaya Lakshmi, "A Machine Learning based Facial Expression and Emotion Recognition for Human Computer Interaction through Fuzzy Logic System," in *2023 International Conference on Inventive Computation Technologies (ICICT)*, Apr. 2023, pp. 166–173, iSSN: 2767-7788.

[2] R. P. Lopes, B. Barroso, L. Deusdado, A. Novo, M. Guimarães, J. P. Teixeira, and P. Leitão, "Digital Technologies for Innovative Mental Health Rehabilitation," *Electronics*, vol. 10, no. 18, p. 2260, Jan. 2021, number: 18 Publisher: Multidisciplinary Digital Publishing Institute.

[3] L. Liu, "Application of facial expression recognition based on domain-adapted convolutional neural network in English smart teaching system," *Soft Computing*, vol. 27, no. 12, pp. 8437–8448, Jun. 2023.

[4] A. Psaltis, K. Kaza, K. Stefanidis, S. Thermos, K. C. Apostolakis, K. Dimitropoulos, and P. Daras, "Multimodal affective state recognition in serious games applications," in *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2016, pp. 435–439.

[5] L. S. Machado, R. M. Moraes, and F. Nunes, "Serious games para saúde e treinamento imersivo," *Abordagens práticas de realidade virtual e aumentada*, vol. 1, pp. 31–60, 2009, publisher: SBC Porto Alegre.

[6] Y. Cheng, B. Jiang, and K. Jia, "A Deep Structure for Facial Expression Recognition under Partial Occlusion," in *2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2014, pp. 211–214.

[7] R. Li, P. Liu, K. Jia, and Q. Wu, "Facial Expression Recognition under Partial Occlusion Based on Gabor Filter and Gray-Level Cooccurrence Matrix," in *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, 2015, pp. 347–351.

[8] S.-S. Liu, Y. Zhang, K.-P. Liu, and Y. Li, "Facial Expression Recognition under Partial Occlusion Based on Gabor Multi-orientation Features Fusion and Local Gabor Binary Pattern Histogram Sequence," in *2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Oct. 2013, pp. 218–222.

[9] D. Poux, B. Allaert, N. Ihaddadene, I. M. Bilasco, C. Djeraba, and M. Bennamoun, "Dynamic Facial Expression Recognition Under Partial Occlusion With Optical Flow Reconstruction," *IEEE Transactions on Image Processing*, vol. 31, pp. 446–457, 2022, publisher: Institute of Electrical and Electronics Engineers (IEEE).

[10] X. Huang, G. Zhao, W. Zheng, and M. Pietikäinen, "Towards a dynamic expression recognition system under facial occlusion," *Pattern Recognition Letters*, vol. 33, no. 16, pp. 2181–2191, Dec. 2012.

[11] I. Buciu, I. Kotsia, and I. Pitas, "Facial expression analysis under partial occlusion," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 5, Mar. 2005, pp. v/453–v/456 Vol. 5, iSSN: 2379-190X.

[12] R. Zhi, M. Flierl, Q. Ruan, and W. B. Kleijn, "Graph-Preserving Sparse Nonnegative Matrix Factorization With Application to Facial Expression Recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 1, pp. 38–52, Feb. 2011, conference Name: IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics).

[13] D. Mushfieldt, M. Ghaziasgar, and J. Connan, "Robust facial expression recognition in the presence of rotation and partial occlusion," in *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, ser. SAICSIT '13. New York, NY, USA: Association for Computing Machinery, Oct. 2013, pp. 186–193.

[14] S. M.d and M. A. Rahiman, "Symbolic Aggregate approXimation-Local Binary Pattern Feature Descriptor Combination for Automatic Facial Expression Recognition," *Journal of Computer Science*, vol. 15, no. 1, pp. 45–56, Jan. 2019, publisher: Science Publications.

[15] B. Houshmand and N. Mefraz Khan, "Facial Expression Recognition Under Partial Occlusion from Virtual Reality Headsets based on Transfer Learning," in *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, 2020, pp. 70–75.

[16] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[17] N. Petrou, G. Christodoulou, K. Avgerinakis, and P. Kosmides, "Lightweight Mood Estimation Algorithm For Faces Under Partial Occlusion," in *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, 2023, pp. 402–407.

[18] A. Rodrigues, J. Lopes, R. Lopes, and L. Teixeira, "Classification of Facial Expressions Under Partial Occlusion for VR Games," *Communications in Computer and Information Science*, vol. 1754 CCIS, pp. 804–819, 2022, iSBN: 9783031232350.

[19] B. Yang, W. Jianming, and G. Hattori, "Face Mask Aware Robust Facial Expression Recognition During The Covid-19 Pandemic," in *2021 IEEE International Conference on Image Processing (ICIP)*, Sep. 2021, pp. 240–244, iSSN: 2381-8549.

Authorized licensed use limited to: b-on: Instituto Politecnico de Braganca. Downloaded on June 06,2025 at 15:16:53 UTC from IEEE Xplore. Restrictions apply.