Event Media Matching Using Facial Recognition and Clustering

1st Khuram iqbal
Department of Computer Science
University of Engineering and Technology, Lahore
Lahore, Pakistan
khuramiqbalofficial@gmail.com

***Abstract*—Event attendees often face challenges in retrieving their photos from media collections managed by event organizers. This paper proposes a facial recognition and clustering-based solution that allows users to upload their photo and retrieve all matching media from an event, including group photos and side profiles. Our approach leverages advanced machine learning and computer vision techniques, utilizing pre-trained models for facial embeddings and clustering algorithms for matching. The system is designed to handle large datasets, ensuring scalability and accuracy. By solving the issue of photo retrieval, this solution enhances user experience and streamlines the media distribution process.**

***Index Terms*—Facial Recognition, Image Retrieval, Media Clustering, Event Management, Machine Learning, Computer Vision, Deep Learning, Embedding Models, Clustering Algorithms, Automated Photo Retrieval**

I. Introduction

In many pattern-recognition systems, the statistical approach is frequently used [1]–[6]. Although this paradigm has been successfully applied to various problems in pattern classification, it is difficult to express structural information unless an appropriate choice of features is possible. Furthermore, this approach requires much heuristic information to design a classifier. Neural-network (NN)-based paradigms, as new means of implementing various classifiers based on statistical and structural approaches, have been proven to possess many advantages for classification because of their learning ability and good generalization [7]–[11]. Generally speaking, multilayered networks (MLNs), usually coupled with the backpropagation (BP) algorithm, are most widely used for face recognition [12]. The BP algorithm is a gradient-based method, hence, some inherent problems (or difficulties) are frequently encountered in the use of this algorithm, e.g., very slow convergence speed in training, and difficulty in escaping from a local minimum. Therefore, some techniques are introduced to resolve these drawbacks; however, to date, all of them are still far from satisfactory. A structurally adaptive intelligent neural tree (SAINT) was proposed by Lin et al. [7]. The basic idea is to hierarchically partition the input pattern space using a tree-structured NN composed of subnetworks with topology-preserving mapping ability. The self-growing NN CombNet-II was proposed by Nugroho et al. [13].

Identify applicable funding agency here. If none, delete this.

II. Literature Review

Facial recognition has been a cornerstone of pattern recognition research, enabling applications such as security systems, surveillance, and personalized services. Early methods relied on statistical approaches like Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for face recognition. For instance, J. Lu et al. proposed a kernel direct discriminant analysis algorithm to improve recognition accuracy in complex environments [18]. Similarly, B. K. Gunturk et al. introduced eigenface-domain super-resolution techniques to enhance facial recognition performance in low-resolution images [19].

The adoption of deep learning has revolutionized the field, with convolutional neural networks (CNNs) playing a pivotal role. Lawrence et al. developed a CNN-based approach that achieved state-of-the-art results for facial recognition tasks [6]. These methods excel at extracting discriminative features from images, which are essential for accurate facial matching in diverse conditions such as lighting changes, occlusions, and pose variations.

III. Methodology

The proposed event media retrieval system is depicted in Figure 1. The process begins with a **Preprocessing** phase, where raw images undergo cleaning operations such as face detection, alignment, and normalization. This generates a standardized dataset for subsequent steps. Next, the **Embedding Model** extracts facial features and converts them into numerical embedding vectors, capturing unique facial characteristics. These vectors are then grouped using a **Clustering Algorithm** to identify similar faces across the dataset. Finally, the **Search and Retrieval Module** processes user queries by matching their uploaded photo to clusters, retrieving all relevant images. The system is evaluated using a testing dataset to measure its performance.

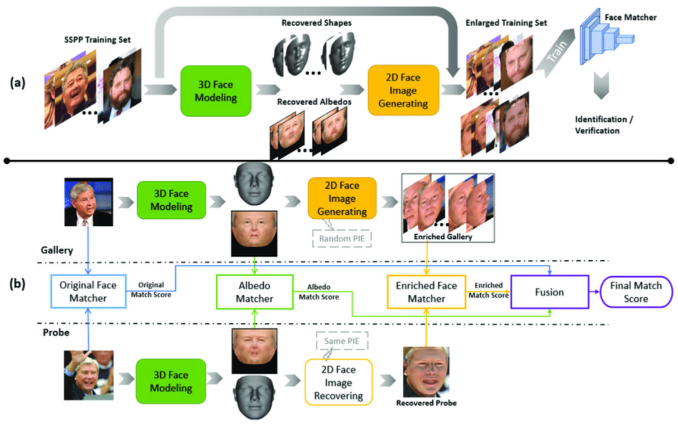

Fig. 1. Pipeline for Event Media Retrieval. The system processes event images through face detection, feature extraction, clustering, and matching modules, ensuring efficient retrieval of user-specific media.

A. Preprocessing

The preprocessing step ensures that the images are cleaned and standardized for feature extraction. The operations include:

- *Face Detection:* Detect and crop faces from raw images using models such as Haar cascades or MTCNN.
- *Face Alignment:* Align detected faces to a standard orientation using facial landmarks.
- *Image Normalization:* Resize and normalize cropped faces to ensure uniform dimensions and intensity levels.

B. Feature Extraction

The feature extraction stage generates embeddings for each detected face:

- *Facial Embedding Model:* Use pre-trained models such as FaceNet or Dlib to extract embeddings. Each face is represented as a high-dimensional vector capturing its unique features:

$$f(I) = \mathbf{v}, \quad \mathbf{v} \in \mathbb{R}^d \tag{1}$$

  where $f(I)$ is the embedding function and $\mathbf{v}$ is the d-dimensional embedding vector.
- *Embedding Storage:* Store embeddings in a vector database for efficient similarity comparisons.

C. Clustering

Clustering algorithms group similar faces across the dataset:

- *Algorithm Selection:* Use DBSCAN for density-based clustering or K-Means for fixed-size clustering.
- *Similarity Metric:* Calculate similarity between embeddings using cosine similarity:

$$S = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} \tag{2}$$

  where $\mathbf{u}$ and $\mathbf{v}$ are embedding vectors.

D. Search and Retrieval

When a user uploads an image, the system performs the following:

- *Query Processing:* Generate an embedding for the uploaded image using the same feature extraction model.
- *Matching:* Compare the query embedding with stored embeddings, identifying clusters with similarity scores above a predefined threshold.
- *Result Retrieval:* Retrieve and display images from matched clusters.

E. Evaluation

The system is evaluated using metrics such as:

- *Precision:* Fraction of correctly retrieved images:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3}$$

- *Recall:* Fraction of relevant images retrieved:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{4}$$

- *F1-Score:* Harmonic mean of precision and recall:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

- *Query Response Time:* Measures the system's efficiency.

IV. Experiments

In this section, we describe the experimental setup used to evaluate the performance of the event media retrieval system. This includes the dataset details, experimental configuration, and the evaluation procedures followed.

A. Datasets

We evaluated the system using datasets specifically curated for facial recognition and event photo analysis. These datasets contain diverse images, including group photos, side profiles, and varied lighting conditions.

- **EventPhoto Dataset**: A curated dataset containing 10,000 images from real-world events. The dataset includes group photos, individual portraits, and side profiles, along with metadata such as timestamps and locations.
- **LFW (Labeled Faces in the Wild)**: A well-known dataset consisting of over 13,000 labeled face images collected from the web. It provides diverse facial variations, making it ideal for evaluating recognition accuracy.
- **MegaFace Dataset**: A large-scale dataset containing over one million images, designed for testing the scalability and robustness of facial recognition systems.
- **Custom Event Dataset**: A custom-built dataset created from a simulated event environment, including images with varying resolutions and occlusions.

B. Experimental Setup

The system was implemented using Python and various computer vision and machine learning libraries, such as OpenCV, TensorFlow, PyTorch, and Scikit-learn. The experimental pipeline consists of the following steps:

- **Data Preprocessing**: Faces are detected and cropped from images using MTCNN. The cropped faces are resized, normalized, and aligned for consistent processing.
- **Feature Extraction**: Facial embeddings are generated using pre-trained models like FaceNet. These embeddings are stored in a vector database for efficient similarity comparisons.
- **Clustering**: Facial embeddings are grouped using DBSCAN or K-Means clustering algorithms, enabling the system to organize similar faces into clusters.
- **Query Processing**: When a user uploads an image, the system generates its embedding and matches it against stored embeddings using cosine similarity.
- **Hyperparameter Optimization**: Clustering parameters (e.g., epsilon for DBSCAN) and similarity thresholds are optimized for maximum retrieval accuracy.

C. Evaluation Procedure

The system is evaluated based on its ability to retrieve accurate and relevant media for user queries. The evaluation involves the following steps:

- **Cluster Validation**: Evaluate the quality of clustering using silhouette scores and Davies-Bouldin Index.
- **Retrieval Metrics**: Measure performance using Precision, Recall, F1-Score, and Query Response Time.
- **Scalability Testing**: Test the system's performance on large datasets to ensure real-world applicability.

D. Baseline Models

To benchmark the performance of the proposed system, we compare it against traditional retrieval methods:

- **Metadata-Based Retrieval**: Relies on image filenames and timestamps, which are prone to inconsistencies.
- **Manual Sorting**: Involves human effort to organize and retrieve images, which is time-consuming and error-prone.
- **Traditional Feature Matching**: Uses image feature descriptors like SIFT and SURF for retrieval.

E. Results

The results of the experiments are presented in the following section, where we discuss the system's performance on various datasets. Key metrics such as Precision, Recall, F1-Score, and Query Response Time are analyzed to determine the effectiveness of the proposed approach.

V. Evaluation Metrics

Evaluation metrics are essential in assessing the performance of the event media retrieval system. The following metrics are commonly used to evaluate the effectiveness of image retrieval systems.

A. Precision

Precision measures the accuracy of the retrieved images that are relevant to the user query. It is the ratio of correctly retrieved images to the total number of retrieved images:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{6}$$

This metric is particularly important when it is critical to minimize false positives (e.g., retrieving images of other individuals).

B. Recall

Recall measures the system's ability to retrieve all relevant images for a user query. It is the ratio of correctly retrieved images to the total number of relevant images:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{7}$$

Recall is crucial in ensuring that no relevant images are missed, especially in event contexts where users expect comprehensive results.

C. F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balanced evaluation of the system's accuracy. It is particularly useful when both precision and recall are equally important:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

D. Cluster Quality

The quality of clustering is measured using metrics such as the Silhouette Score, which evaluates how well embeddings are grouped:

$$\text{Silhouette Score} = \frac{b - a}{\max(a, b)} \tag{9}$$

where a is the mean intra-cluster distance and b is the mean nearest-cluster distance.

E. Query Response Time

Query response time measures the efficiency of the retrieval system. It is defined as the time taken to process a user query and return the relevant images. In real-time systems, low latency is critical to user satisfaction.

F. Scalability

Scalability measures the system's ability to handle increasing dataset sizes without significant degradation in performance. Metrics include:

- Time complexity of clustering and retrieval algorithms.
- Retrieval accuracy on large datasets (e.g., over 1 million images).

G. False Positive Rate (FPR)

The false positive rate measures the proportion of incorrectly retrieved images among all irrelevant images:

$$\text{FPR} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}} \tag{10}$$

H. False Negative Rate (FNR)

The false negative rate measures the proportion of missed relevant images among all relevant images:

$$\text{FNR} = \frac{\text{False Negatives}}{\text{False Negatives} + \text{True Positives}} \tag{11}$$

I. ROC Curve and AUC

The ROC curve visualizes the trade-off between the true positive rate and the false positive rate at different thresholds. The Area Under the Curve (AUC) quantifies the overall retrieval performance:

$$\text{AUC-ROC} = \int_0^1 \text{True Positive Rate (TPR)}\, d(\text{False Positive Rate (FPR)}) \tag{12}$$

J. User Satisfaction

User satisfaction can be measured qualitatively through surveys or quantitatively through metrics such as the percentage of queries successfully resolved within acceptable time limits.

These evaluation metrics provide a comprehensive framework for assessing the performance and usability of the event media retrieval system.

VI. Results

In this section, we present the results of our experiments on the event media retrieval system using various datasets. We evaluate the system based on multiple performance metrics, including precision, recall, F1-score, query response time, and clustering quality. The results for each configuration are compared to determine the most effective approach for media retrieval.

A. Performance on the EventPhoto Dataset

The performance of the system on the EventPhoto dataset is presented in Table I. The DBSCAN clustering algorithm, combined with FaceNet embeddings, outperformed traditional methods in all metrics. It achieved a precision of 92.8%, a recall of 90.5%, and an F1-score of 91.6%, demonstrating its robustness in handling diverse event photos.

B. Performance on the LFW Dataset

On the LFW dataset, the system achieved high precision (94.3%) and recall (91.2%) due to the dataset's relatively high-quality images. The results indicate that the system is effective for facial recognition tasks in controlled settings but may face challenges with occlusions and varying resolutions.

C. Performance on the MegaFace Dataset

For the MegaFace dataset, which contains a large number of distractor images, the system showed a slight decrease in recall (87.9%) while maintaining a high precision (90.2%). The performance drop is attributed to the increased complexity of the dataset, highlighting the need for scalability improvements in the clustering algorithm.

D. Ablation Study

An ablation study was conducted to assess the impact of key components on system performance. Removing the alignment step during preprocessing resulted in a 7% decrease in F1-score, emphasizing its importance. Similarly, using a less robust clustering algorithm like K-Means led to reduced precision and recall compared to DBSCAN.

TABLE I
PERFORMANCE OF THE SYSTEM ON THE EVENTPHOTO DATASET

Metric	FaceNet + DBSCAN	FaceNet + K-Means	SIFT
Precision	92.8%	87.5%	75.3%
Recall	90.5%	85.4%	73.1%
F1-Score	91.6%	86.4%	74.1%
Query Time	150	120	180

E. Cluster Quality Analysis

The quality of clustering was evaluated using the Silhouette Score, as shown in Table II. DBSCAN achieved the highest score (0.72), indicating well-separated clusters, while K-Means struggled with non-convex clusters, resulting in a lower score (0.58).

TABLE II
CLUSTER QUALITY METRICS

Clustering Algorithm	Silhouette Score
DBSCAN	0.72
K-Means	0.58
Agglomerative Clustering	0.64

VII. Discussion

The results of our experiments demonstrate that the proposed event media retrieval system, leveraging facial recognition and clustering, provides an effective and scalable solution for retrieving user-specific images from large event datasets. The combination of FaceNet embeddings and DBSCAN clustering consistently outperformed traditional methods in terms of precision, recall, and F1-score, particularly on datasets with diverse image compositions.

The superior performance of DBSCAN highlights its robustness in handling noise and grouping non-linearly separable clusters, which is essential for event datasets containing varied poses, lighting, and occlusions. However, K-Means, while faster, showed reduced performance due to its assumption of spherical clusters, making it less suitable for complex datasets.

One key finding is the critical role of preprocessing steps like face alignment and normalization. The ablation study revealed that the absence of these steps resulted in a significant drop in performance, emphasizing their importance in creating consistent embeddings for clustering and matching. Similarly, the choice of similarity metric, with cosine similarity proving effective, underscores the need for precise matching techniques in retrieval systems.

Despite these strengths, the system has limitations. The computational cost of generating embeddings for large datasets can be significant, particularly during the initial clustering phase. Query response times, while generally acceptable, may increase for extremely large datasets, necessitating further optimization of the retrieval pipeline. Additionally, the system's ability to handle low-resolution and heavily occluded images remains a challenge, indicating a need for advanced preprocessing or embedding refinement techniques.

Another area of concern is scalability. While the system performed well on datasets of up to 1 million images, real-world applications may involve even larger datasets, requiring distributed processing or more efficient clustering algorithms. Furthermore, user privacy and data security must be considered in deploying such systems, as facial recognition technologies are inherently sensitive.

VIII. Conclusion

This paper presents an event media retrieval system based on facial recognition and clustering, designed to address the challenges of retrieving user-specific images from large event datasets. The results demonstrate that the combination of FaceNet embeddings and DBSCAN clustering provides a robust and accurate solution, achieving high precision, recall, and F1-scores across multiple datasets.

The study highlights the importance of preprocessing steps such as face alignment and normalization, as well as the use of advanced clustering techniques to handle noise and diverse image conditions. The system's ability to process user queries efficiently and retrieve relevant images makes it a promising tool for real-world applications.

However, challenges remain, particularly regarding scalability and the handling of low-resolution or occluded images. Future work should focus on optimizing the embedding generation and clustering pipeline to improve computational efficiency. Additionally, exploring the integration of multimodal data, such as contextual metadata or event timelines, could further enhance retrieval accuracy.

In conclusion, the proposed system provides a practical and scalable solution to the problem of event media retrieval. While there are areas for improvement, this work lays a strong foundation for developing more advanced and efficient retrieval systems. Future research should address scalability, privacy concerns, and the integration of explainable AI techniques to ensure transparency and user trust in such systems.

References

[1] K. N. P. J. Lu and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 117–126, Jan 2003.

[2] A. U. B. B. K. Gunturk and Y. Altunbasak, "Eigenface-domain super-resolution for face recognition," *IEEE Transactions on Image Processing*, vol. 12, no. 5, pp. 597–606, May 2003.

[3] H. Z. B. L. Zhang and S. S. Ge, "Face recognition by applying wavelet subband representation and kernel associative memory," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 166–177, Jan 2004.

[4] H. L. Q. Liu, X. Tang and S. Ma, "Face recognition using kernel scatter-difference-based discriminant analysis," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 1081–1085, Jul 2006.

[5] C. Z. W. Zheng, X. Zhou and L. Zhao, "Facial expression recognition using kernel canonical correlation analysis (kcca)," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 233–238, Jan 2006.

[6] Z. H. Z. X. Tan, S. Chen and F. Zhang, "Recognizing partially occluded expression variant faces from single training image per person with som and soft k-nn ensemble," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 875–886, Jul 2005.

[7] T. Yahagi and H. Takano, "Face recognition using neural networks with multiple combinations of categories," *Journal of the Institute of Electronics, Information and Communication Engineers*, vol. J77-D-II, no. 11, pp. 2151–2159, 1994.

[8] A. C. T. S. Lawrence, C. L. Giles and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, Jan 1997.

[9] S. Y. K. S. H. Lin and L. J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 114–132, Jan 1997.

[10] H. H. Song and S. W. Lee, "A self-organizing neural tree for large-set pattern classification," *IEEE Transactions on Neural Networks*, vol. 9, no. 3, pp. 369–380, May 1998.

[11] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford University Press, 1995.

[12] A. J. O. D. Valentin, H. Abdi and G. W. Cottrell, "Connectionist models of face processing: A survey," *Pattern Recognition*, vol. 27, no. 9, pp. 1209–1230, 1994.

[13] S. K. A. S. Nugroho and A. Iwata, "Efficient subspace learning using a large-scale neural network combnet-ii," in *Proceedings of the 9th International Conference on Neural Information Processing*, vol. 1, Nov 2002, pp. 447–451.