International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. 10 Issue 02 | 2026

Content-Based Image Retrieval Using Deep Learning

[1]Dr. Sachin A Vyawhare
[1]Assistant Professor, [1]Sanmati Engineering College, Washim, Maharashtra

DOI: 10.5281/zenodo.20605361

ABSTRACT –

A content-based image retrieval (CBIR) system's retrieval performance depends on its ability to learn efficient feature representations and similarity metrics. It is still one of the most difficult unsolved issues that significantly impedes the performance of practical CBIR systems, despite decades of intensive research efforts. The primary obstacle has been identified as the well-known "semantic gap" problem between machine-captured low-level picture pixels and human-perceived high-level semantic concepts. Machine learning is one approach that has been extensively researched as a potential long-term solution to close the semantic gap. Inspired by the recent successes of deep learning methods for computer vision and other applications, we try to solve an open problem in this paper: whether deep learning can help close the semantic gap in CBIR and to what extent the state-of-the-art deep learning methods for learning feature representations and similarity measures can improve CBIR tasks. In particular, we examine a state-of-the-art deep learning technique (Convolutional Neural Networks) for CBIR tasks under various conditions in order to research a deep learning framework with application to CBIR tasks with a wide range of empirical studies. We review some key findings for further research and identify some promising results from our empirical investigations.

***Keywords* - Deep Learning; Content-Based Image Retrieval; Convolutional Neural Networks; Feature Representation**

I. INTRODUCTION

Multimedia specialists have dedicated decades to the examination of feature representation and similarity evaluation, which are essential for the retrieval efficacy of a content-based image retrieval system. The prominent "semantic gap" issue, which arises between low-level image pixels captured by machines and high-level semantic concepts understood by humans, is the principal reason it remains one of the most challenging problems in contemporary content-based image retrieval (CBIR) research, despite numerous proposed approaches. This challenge can be attributed to the fundamental issue of artificial intelligence (AI): the creation and training of intelligent machines capable of performing tasks like to those of humans. A viable approach to address this significant challenge over time is machine learning.

Recent years have witnessed numerous notable advancements in machine learning methodologies. A notable invention is "deep learning," which consists of a category of machine learning algorithms employing deep architectures composed of several non-linear transformations to represent high-level abstractions in data. Deep learning mimics the human brain, characterized by a deep architecture that processes information through several levels of transformation and representation, unlike conventional machine learning methods that often employ "shallow" structures. Deep learning techniques allow a system to learn complex functions that convert raw sensory input data into output without relying on human-engineered features, by exploring deep architectures to autonomously learn features at multiple levels of abstraction from data. The application of deep learning techniques across various domains, including speech recognition, object recognition, and natural language processing, has yielded encouraging outcomes in multiple recent research.

The essence of a content-based picture retrieval system is in the image's color, texture, and shape attributes. Figure 2 illustrates the fundamental architecture of a content-based image retrieval system. The image is obtained based on the degree of similarity between the query image and the database photos. Every image comprises color, texture, shape, and both low-level and high-level components. Although resolving mathematical problems in high-level features is challenging, the attributes of image retrieval based on low-level features can be readily extracted from the image. Three primary techniques are employed for image retrieval: color retrieval, texture retrieval, and shape retrieval. This article provides a designated channel to view this characteristic and acquire the corresponding image.

A. What is Deep Learning?

Deep learning is an advanced branch of machine learning that utilizes artificial neural networks (ANNs) to perform complex computations on large volumes of data. These neural networks are inspired by the structure and functioning of the human brain, enabling machines to learn patterns, make decisions, and improve performance over time without explicit programming. By processing data through multiple layers, deep learning models can automatically extract meaningful features and representations.

Deep learning algorithms are trained using large datasets, allowing systems to learn from examples rather than relying on manually defined rules. This capability has made deep learning highly effective in solving real-world problems. As a result, it is widely used across various industries such as healthcare for disease diagnosis, eCommerce for recommendation systems, entertainment for content personalization, and advertising for targeted marketing.

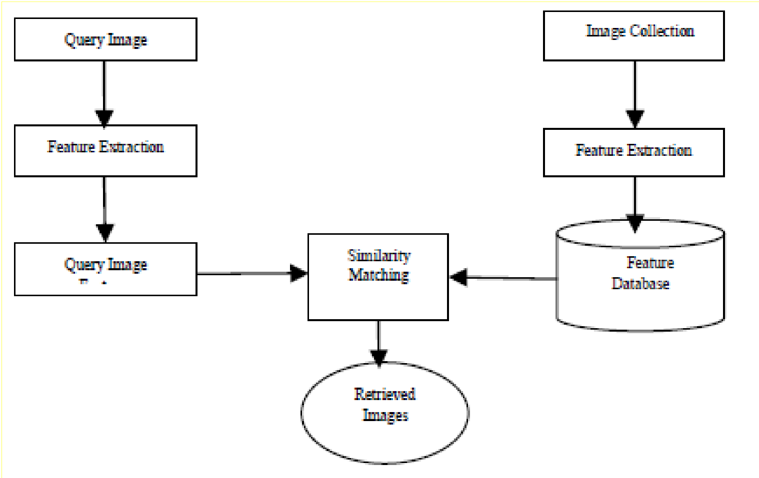

Fig.1 Representation of CBIR system

B. How Deep Learning Algorithms Work

Deep learning algorithms operate through artificial neural networks that mimic how the human brain processes information. These networks consist of multiple interconnected layers, where each layer transforms the input data into more abstract representations. During the training phase, the algorithm analyzes large datasets and identifies hidden patterns, relationships, and features within the data.

The learning process involves adjusting weights and biases within the network to minimize errors and improve prediction accuracy. Deep learning models build hierarchical representations, meaning simpler features are learned in earlier layers while more complex features are extracted in deeper layers. Since different tasks require different approaches, multiple deep learning algorithms are available, each suited for specific applications. Therefore, selecting the appropriate algorithm depends on the nature of the problem and the type of data being used.

C. Types of Deep Learning Algorithms

1) Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs), also known as ConvNets, are specialized deep learning models primarily used for image processing and object detection tasks. These networks consist of multiple layers designed to automatically and adaptively learn spatial hierarchies of features from input data. The first CNN, known as LeNet, was developed by Yann LeCun in 1988 and was initially used for recognizing handwritten digits and ZIP codes.

CNNs are widely applied in various domains such as medical image analysis, satellite image interpretation, anomaly detection, and time-series forecasting. Their ability to capture spatial and visual patterns makes them highly effective for image-related tasks.

CNNs work through a sequence of layers. The convolution layer applies filters to extract important features from the input data. This is followed by the Rectified Linear Unit (ReLU) layer, which introduces non-linearity by transforming the output into a rectified feature map. Next, the pooling layer reduces the dimensionality of the feature map, making the computation more efficient while preserving important information. Finally, the fully connected layer processes the flattened data and performs classification or prediction based on the learned features.

2) Long Short-Term Memory Networks (LSTMs)

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network designed to capture and retain long-term dependencies in sequential data. Unlike traditional neural networks, LSTMs are capable of remembering information for extended periods, making them highly suitable for tasks that involve time-dependent patterns.

LSTMs are particularly useful in time-series prediction, speech recognition, music generation, and pharmaceutical research. Their architecture includes a chain-like structure with specialized components called gates, which regulate the flow of information. These gates enable the model to selectively forget irrelevant data, update important information, and produce meaningful outputs based on both past and current inputs. This ability to manage long-term dependencies is what makes LSTMs powerful for sequential data analysis.

3) Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are designed to handle sequential data by maintaining an internal memory of previous inputs. Unlike traditional neural networks, RNNs have connections that form directed cycles,

allowing information to persist across different time steps. This structure enables the output from a previous step to be used as input for the current step, making RNNs effective for tasks where context and sequence matter.
RNNs are commonly used in applications such as natural language processing, machine translation, handwriting recognition, image captioning, and time-series analysis. By remembering past inputs, they can identify patterns over time and make more informed predictions. However, traditional RNNs may struggle with long-term dependencies, which is why advanced variants like LSTMs are often preferred for more complex sequential tasks.

II. RELATED WORK

The connection between content-based photo retrieval, distance metric learning, and deep neural network learning is the primary focus of the study that we are currently pursuing. Following this, we will provide a brief summary of each collection of works that are related together. Image retrieval based on the content of the images received While the content-based image retrieval (CBIR) technique is one of the core research challenges that has been extensively researched in the multimedia field for decades, it is also one of the most challenging research problems. The objective of the CBIR technique is to search for images by analyzing the visual contents of those images; hence, the CBIR technique is centered on the representation of images. There have been a great number of low-level feature descriptors proposed for the purpose of picture representation throughout the course of the past few decades. These feature descriptors include global features like color features, edge features, texture feature, GIST, and CENTRIST, as well as more recent local feature representations like the bag-of-words (BoW) models that make use of local feature descriptors. There are also global features like color features, edge features, and texture feature. It was discovered, during the process of conducting a review of similar work in the field of agriculture, that earlier research had been carried out with the purpose of detecting and classifying plant illnesses (including those that affect leaves, flowers, fruits, and other plant parts).
This research had been carried out using image processing and computer vision technologies. Only a small amount of research has been carried out in the field of CBIR, which was designed expressly for the purpose of diagnosing diseases that affect plants. Meunkaewjinda et al. (2008)[7] developed an automatic plant disease detection system that makes use of a variety of different artificial intelligence approaches in order to diagnose grape leaf illnesses. This method was developed in order to detect diseases that affect grape leaves. The employment of a self-organizing feature map in conjunction with a back propagation neural network is what allows for the classification of grape leaf hues. A modified version of the self-organizing feature map is utilized for the goal of segmentation, and support vector machine is applied for classification. Both algorithms are utilized for the purpose of segmentation. It is possible to acquire an average of 86.0330 percent of correct diagnoses by the utilization of this approach. Self-organizing feature maps are also applied for the aim of diagnosing diseases that damage cotton leaves (Gulhane and Gurjar, 2011)[8]. Cotton leaves are susceptible to such diseases. Kebapci et al. (2011)[9] constructed the CBIR system with the intention of obtaining images of plants by making use of color, shape, and texture qualities. This was done during the process of developing the system. The color histogram, the color co-occurrence matrix, and a modified version of the Gabor method that is based on a patch-based approach are the methods that have been developed. For the purpose of capturing the local characteristics of the plant, the SIFT technique is applied, whereas the global shape descriptor technique is utilized for the purpose of capturing the global circumstances of the plant. By means of the experiment that is now being carried out, it is possible to reach an accuracy rate of 73% in the identification of residential plants. The authors of the study (Li et al., 2010)[10] present a method for the automatic detection of diseases that are frequently seen in wheat. Through the utilization of the Otsu method, the lesion area can be extracted from the image for the purpose of utilization. Through the use of principal component analysis, fourteen distinct morphological traits are extracted from the segmented region and filtered. It has been observed that the disease detection rate accounts for roughly 85 percent.
Using photographs of plant leaves, the authors of the study (SathyaBama et al., 2011)[11] describe a method for the recovery of plant images. This method includes the use of photographs. Certain aspects of the leaf, like its color, shape, and texture, are taken into consideration during the retrieval process. SIFT, on the other hand, is utilized for the extraction of form features, SIFT in the saturation band of the HSV color space is utilized for the extraction of color characteristics, and log Gabor wavelet in SIFT is employed for the extraction of texture data. All of these applications are applied in order to extract both form and color features. The photos of the leaves are extracted by combining these qualities in order to accomplish this job. In order to get a retrieval efficacy of around 97.9%, it is possible to achieve this. (Kulkarni and Patil, 2012)[12] Anand and Ashwini came up with a method that may be used to identify diseases that affect plants. Through the utilization of the Gabor filter and an artificial neural network, this treatment approach is carried out. By employing this approach, it is possible to achieve a level of performance that comprises 91%. Kailey and Sahdra (2012)[13] make a suggestion for an automatic plant disease detection technique that is based on histogram matching. This technique is intended to detect diseases in plants. In order to search for similarities in multimedia content, standard CBIR techniques

often select stiff distance functions on some extracted low-level features. This is done with the intention of locating similarity. Distance functions such as Euclidean distance and cosine similarity are included in this category. Because of the enormous difficulties of the semantic gap that exists between low-level visual data recovered by computers and high-level human perceptions, the fixed rigid similarity/distance function may not always be perfect to the intricate visual picture retrieval tasks. This is because the semantic gap is a significant and difficult problem to solve. The semantic gap offers a significant barrier, which is the reason for this conclusion. This has led to an increase in the number of active research efforts that have been aimed towards the formulation of various distance/similarity measures on some low-level features through the investigation of machine learning processes over the course of the previous few years. This has been the result of the aforementioned situation. A number of the works that have been done have focused on learning how to hash or condense codes, which is one of the tactics that are included in these strategies. The use of distance metric learning, also known as DML, is yet another approach that can be implemented in order to enhance the feature representation.

The unit of measurement for distance Learning distance metrics for the purpose of picture retrieval is the process that is being involved. There has been a substantial amount of time and effort invested in the investigation of this topic by the communities of machine learning and multimedia retrieval. In the next lines, a brief review of the numerous areas of work that are currently being done for distance metric learning is offered. These groups are structured in accordance with learning environments and ideologies that are distinct from one another. When it comes to the formats of training data, the majority of the DML studies that are currently available typically operate with two distinct types of data, which are also referred to as side information. These two types of data are paired constraints, which include must-link constraints and cannot-link constraints, and triplet constraints, which include a similar pair and a dissimilar pair. Additional research has been conducted that makes direct use of the class labels for DML by adhering to a standard machine learning scheme. One example of this is the Large Margin Nearest Neighbor (LMNN) algorithm, which is not fundamentally different from the other algorithms. In terms of the various learning approaches, distance metric learning techniques are typically divided into two groups: the global supervised approaches, which learn a metric on a global setting by simultaneously satisfying all of the constraints, and the local supervised approaches, which learn a metric on the local sense by only satisfying the given local constraints from neighboring information. Both of these groups are referred to as "distance metric learning."

Both of these are instances of alternative ways of approaching the learning process. The majority of the DML research that is now available generally makes use of batch learning methods. This is because batch learning is the most effective learning methodology. It is common for these methods to operate under the assumption that the whole collection of training data must be provided before the learning assignment and that a model must be trained from the absolute beginning. Online DML algorithms, on the other hand, have been the focus of extensive research in recent years in order to effectively manage huge amounts of data. This is in contrast to the batch learning methods. The basic goal of distance metric learning is to acquire an ideal metric that simultaneously decreases the distance between images that are similar to one another while simultaneously increasing the distance between images that are not similar to one another. Both of these outcomes are desirable. There is a significant connection between the idea of distance metric learning and a separate method that is known as similarity learning. This strategy is applicable to this particular situation. An example of this would be Chechik et al. [7], who proposed an online method for scaled image similarity (OASIS) with the intention of improving the efficiency of photo retrieval.

III. ANALYSIS OF PROBLEM

Inspired by the remarkable success of deep learning in various domains, this work focuses on exploring its application in Content-Based Image Retrieval (CBIR) tasks. While deep learning has gained significant attention in areas such as image classification and object recognition within computer vision, its application in CBIR remains relatively underexplored. This study aims to bridge that gap by investigating how deep learning techniques can be effectively utilized to learn meaningful feature representations from images and improve similarity measurement for retrieval tasks.

The primary objective of this research is to examine the potential of deep learning methods in addressing key challenges associated with CBIR systems. Specifically, the study seeks to determine whether deep learning models can effectively learn robust and discriminative image features that enhance retrieval performance. Additionally, it evaluates whether these learned features can outperform traditional handcrafted features developed by experts in multimedia and computer vision. Another important aspect of the research is to explore the adaptability of deep learning models, particularly how a model trained in one domain can be transferred and applied efficiently to CBIR tasks in a different domain.

To address these research questions, a comprehensive deep learning framework for CBIR is proposed. The framework leverages Convolutional Neural Networks (CNNs) to automatically learn feature representations directly from image data. CNNs are chosen due to their proven capability in capturing spatial hierarchies and visual patterns in images. The study includes extensive empirical evaluations across a variety of CBIR tasks, using different datasets and experimental settings to assess the effectiveness and robustness of the proposed approach.

The results obtained from these experiments are promising and provide valuable insights into the role of deep learning in CBIR systems. The findings indicate that deep learning-based feature representations can significantly enhance retrieval accuracy and performance compared to traditional approaches. Furthermore, the study highlights the potential of transfer learning in adapting pre-trained models to new CBIR domains, thereby reducing computational cost and improving efficiency.

In summary, this work makes several important contributions to the field of CBIR. First, it introduces a deep learning-based framework that utilizes large-scale convolutional neural networks to learn effective and discriminative image features. Second, it presents an extensive set of empirical studies that provide a comprehensive evaluation of CNN-based feature learning across different CBIR scenarios and configurations. These contributions collectively demonstrate the effectiveness of deep learning techniques in advancing the capabilities of content-based image retrieval systems.

IV. PROPOSED METHOD

In the proposed approach, a multi-feature Content-Based Image Retrieval (CBIR) method is developed by integrating several visual features such as color histogram, edge information, edge direction, edge histogram, and texture characteristics. This combination of multiple features enhances the system's ability to capture comprehensive image content. Initially, images are collected into a structured database consisting of different image groups. Before feature extraction, preprocessing steps are applied, including noise removal and irrelevant data elimination, to improve the quality of the input images.

After preprocessing, important visual features are extracted from each image and stored in the form of compact signature files. These signatures serve as unique representations of the images, where similar images are expected to have similar signatures. When a query image is provided, its features are extracted in the same way and compared with the stored signatures in the database. The similarity between images is measured by calculating the distances between their respective feature vectors.

To improve the accuracy of similarity measurement, appropriate weights are assigned to different features, ensuring normalization of the distance coefficients. These normalized values are then sorted and indexed based on their similarity scores, enabling efficient retrieval of the most relevant images. This optimized framework improves both the performance and precision of the CBIR system.

V. APPLICATIONS

Content-Based Image Retrieval (CBIR) has a wide range of real-world applications across different domains. Some of the major application areas are described below:

A. Medical Applications

CBIR plays a significant role in the medical field, particularly in the analysis and retrieval of medical images such as MRI scans. One important application involves retrieving 2D MRI slices from 3D brain volumes. In this system, a user provides a 2D MRI query image, and the system identifies the corresponding 3D brain volume that matches the query. Once the relevant volume is identified, further search is conducted within that volume to retrieve the most similar slices related to the specific brain region. This CBIR application typically uses machine learning techniques such as the Support Vector Machine (SVM) to classify and retrieve relevant images accurately, assisting doctors in diagnosis and research.

B. Remote Sensing Image Retrieval

Remote sensing image retrieval is another critical application of CBIR, especially for environmental monitoring and territorial management. High-resolution satellite images are used as input, and advanced models are applied to analyze their structural properties. In this approach, the anisotropic characteristics of images are modeled using Local Structure Tensor (LST), which helps in capturing directional information effectively. The use of structure tensor-based Riemannian statistical models enhances the retrieval performance by accurately representing image patterns and textures. This application is highly useful in land-use analysis, disaster monitoring, and urban planning.

C. Natural Image Retrieval

Natural image retrieval deals with images captured in everyday environments such as streets, homes, and natural landscapes. These images often contain complex and diverse visual patterns, making retrieval a challenging task. To improve performance, advanced similarity measures are used. In this context, researchers such as Guo-Dong Guo have proposed combining multiple similarity measures to enhance retrieval accuracy. One such approach is the Constrained Similarity Measure (CSM), which integrates Euclidean distance with machine learning techniques like Support Vector Machines. This method considers both positive and negative samples to define clear decision boundaries, leading to improved retrieval results.

VI. CONCLUSION

This study presents a robust Content-Based Image Retrieval (CBIR) system by integrating many visual aspects, including color histogram, edge information, edge direction, edge histogram, and texture. The integration of these features allows the system to record detailed and distinctive visual attributes, resulting in enhanced retrieval precision. Employing preprocessing techniques guarantees superior quality input, whereas the creation of compact signature files facilitates efficient storage and comparison of picture data. The resemblance between images is accurately assessed by calculating weighted distances between feature vectors, with normalization and suitable weighting being essential for improving performance. The suggested multi-feature methodology exhibits distinct advantages over conventional single-feature methods by delivering more accurate and dependable retrieval outcomes. Moreover, optimization via indexing and sorting algorithms enhances the system's overall efficiency. The research underscores the extensive applicability of CBIR across many fields, including medical imaging, remote sensing, and natural image retrieval. Methods such as Support Vector Machines enhance classification and retrieval precision in these applications. The suggested CBIR system demonstrates efficacy, scalability, and adaptability for practical image retrieval problems. Future research may concentrate on using deep learning methodologies, such as Convolutional Neural Networks, to augment feature extraction and similarity learning, thus attaining superior performance and automation in CBIR systems.

VII. REFERENCES

[1] Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: State of the art and challenges. ACM Trans. Multimed. Comput. Commun. Appl. **2**(1), 1–19 (2006)

[2] Bringer, J., Chabanne, H., Patey, A.: Privacy-preserving biometric identification using secure multiparty computation: an overview and recent trends. IEEE Signal Process. Mag. **30**(2), 42–52 (2013)

[3] Aghasaryan, A., Bouzid, M., Kostadinov, D., Kothari, M., Nandi,A.: On the use of LSH for privacy preserving personalization. In: Proceedings of the 12th IEEE International Conference Trust,Security, Privacy in Computing and Communications (TrustCom). pp. 362–371 (2013)

[4] Fanti, G., Finiasz, M., Ramchandran, K.: One-way private media search on public databases: the role of signal processing. IEEE Signal Process. Mag. **30**(2), 53–61 (2013)

[5] Acar, G., et al.: FPDetective: dusting the web for fingerprinters. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security (CCS), pp. 1129–1140 (2013)

[6] Balsa, E., Troncoso, C., Diaz, C.: OB-PWS: Obfuscation-based private web search. In: Proceedings of the IEEE Symposium on Security and Privacy, pp. 491–505 (2012)

[7] Meunkaewjinda, A., Kumsawat, P., Attakitmongcol, K., Srikaew, A., 2008. Grape leaf disease detection from color imagery using hybrid intelligent system. In: Proceedingsof ECTI-CON 2008. IEEE, pp. 513e516.

[8] Gulhane, Viraj A., Gurjar, Ajay, 2011. Detection of diseases on cotton leaves and its possible diagnosis. Int. J. Image Process. 5 (5), 590e598.

[9] Kebapci, Hanife, Yanikoglu, Berrin, Unal, Gozde, 2011. Plant image retrieval using color, shape and texture features. Comput. J. 54 (9), 1475e1490.

[10] Li, Jinghui, Gao, Lingwang, Shen, Zuorui, 2010. Extraction and analysis of digital images feature of three kinds of wheat diseases. In: International Congress on Image and Signal Processing. IEEE, pp. 2543e2548.

[11] SathyaBama, B., MohanVali, S., Raju, S., Abhai Kumar, V., 2011. Content based leaf image retrieval (CBLIR) using shape, color and texture features. Indian J. Comput.Sci. Eng. 2 (2), 202e211.

[12] Kulkarni, Anand H., Patil, Ashwin, 2012. Applying image processing technique to detect plant diseases. Int. J. Mod. Eng. Res. 2 (5), 3661e3664.

[13] Kailey, Kamaljot Singh, Sahdra, Gurjinder Singh, 2012. Content Based Image Retrieval (CBIR) for identifying image based plant disease. Comput. Technol. Appl. 3 (3), 1099e1104. Conference on Machine Learning and Cybernetics (ICMLC), pp. 719e724.

[14] Guo-Dong Guo, Anil K.Jain, Wei-Ying Ma and HongJiang, ”Learning similarity measure for natural image retrieval with relevance feedback”, IEEE transactions on neural networks,Vol-13, No.4, 2002

[15] Erkin, Z.: Protection and retrieval of encrypted multimedia content: when cryptography meets signal processing. EURASIP J. Inf. Secur. **2007**, 20 (2007)

[16] Lagendijk, R.L., Erkin, Z., Barni, M.: Encrypted signal processing for privacy protection: conveying the utility of homomorphic encryption and multiparty computation. IEEE Signal Process. Mag. **30**(1), 82–105 (2013)

[17] Shashank, J., Kowshik, P., Srinathan, K., Jawahar, C.V.: Private content based image retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2008)
[18] Sabbu, P.R.,Ganugula, U.,Kannan, S., Bezawada,B.: An oblivious image retrieval protocol. In: Proceedings of the IEEE International Workshops ofAdvanced InformationNetworking and Applications (WAINA), pp. 349–354 (2011)
[19] Erkin, Z., Franz,M., Guajardo, J., Katzenbeisser, S., Lagendijk, I., Toft, T.: Privacy-preserving face recognition. In: Proceedings of the 9th International Symposium on Privacy Enhancing Technologies (PETS), pp. 235–253 (2009)
[20] Sadeghi, A.-R, Schneider, T., Wehrenberg, I.: Efficient privacypreserving face recognition. In: Proceedings of the 12th International Conference on Information Security and Cryptology (ICISC), pp. 229–244 (2009)