Harnessing Social Context for Improved Face Recognition

Romil Bhardwaj, Gaurav Goswami, Richa Singh, and Mayank Vatsa
IIIT-Delhi, India
{romil11092, gauravgs, rsingh, mayank}@iiitd.ac.in

Abstract

Face recognition is traditionally based on features extracted from face images which capture various intrinsic characteristics of faces to distinguish between individuals. However, humans do not perform face recognition in isolation and instead utilize a wide variety of contextual cues as well in order to perform accurate recognition. Social context or co-occurrence of individuals is one such cue that humans utilize to reinforce face recognition output. A social graph can adequately model social-relationships between different individuals and this can be utilized to augment traditional face recognition methods. In this research, we propose a novel method to generate a social-graph based on a collection of group photographs and learn the social context information. We also propose a novel algorithm to combine results from a commercial face recognition system and social context information to perform face identification. Experimental results on two publicly available datasets show that social context information can improve face recognition and help bridge the gap between humans and machines in face recognition.

1. Introduction

Automated face recognition is an active research area and has a wide variety of potential applications. However, unlike existing face recognition techniques which do not perform reliably in the presence of various covariates, face recognition by humans is highly accurate especially in the case of familiar faces [11]. Therefore, researchers have strived to incorporate human-like characteristics in their automated algorithms in order to bridge this gap. Several metrics borrowed from human cognition such as memorability [6] and visual cortex inspired [10] algorithms have been proposed. Another important cue which humans utilize for recognizing individuals, which is very less studied is *social context*.

Social context plays an important role in identity recognition and is readily available for a human observer. Humans tend to form social groups which become a part of their identity. People associate each other with these groups and it is easier for them to recognize individuals within known social context. As shown in Figure 1, it is rudimentary for humans to recognize people even with substantially altered appearance (beard, spectacles, hairstyle, or attire) if they are seen together with their known social group. However, since most of the existing automatic algorithms rely solely on the content of isolated face images, it is difficult for them to handle covariates which can alter facial appearances [5], [8]. It is our hypothesis that if social context information can be included in a face recognition system, then the performance can be improved.

Social media has become increasingly popular over the past few years, leading to an increase in the availability of tagged images which has allowed the study of social relationships and the creation of social graphs. A social graph connects people based on their co-occurrence in media such as videos and photos uploaded by users. Chen *et al.* [4] have proposed a method to determine the social relationship of a pair of people from a collection of group photographs. Bloess *et al.* [3] have also utilized social context for face recognition. However, their algorithm assumes that additional labels pertaining to social structure are available at the time of training which may or may not be true in different scenarios. Barr *et al.* [2] have utilized face clusters to build a social network from video clips where the individuals are unknown. However, they have not applied the generated social graph to improve the face recognition performance. Recently, Bharadwaj *et al.* [9] have incorporated social context in the form of association rules mined from labelled data pertaining to 4675 identities. The strength of a social relation between two individuals in the gallery is specified by support and confidence measures of the concerned association rule. They report 5% improvement in the rank 25 recognition accuracy when social information is used to augment an existing commercial face recognition system. Their results suggest that face recognition in group photographs can be highly useful in social media by providing accurate automated tagging, photo organization, and search functionalities.

In this research, we propose a novel algorithm to com-

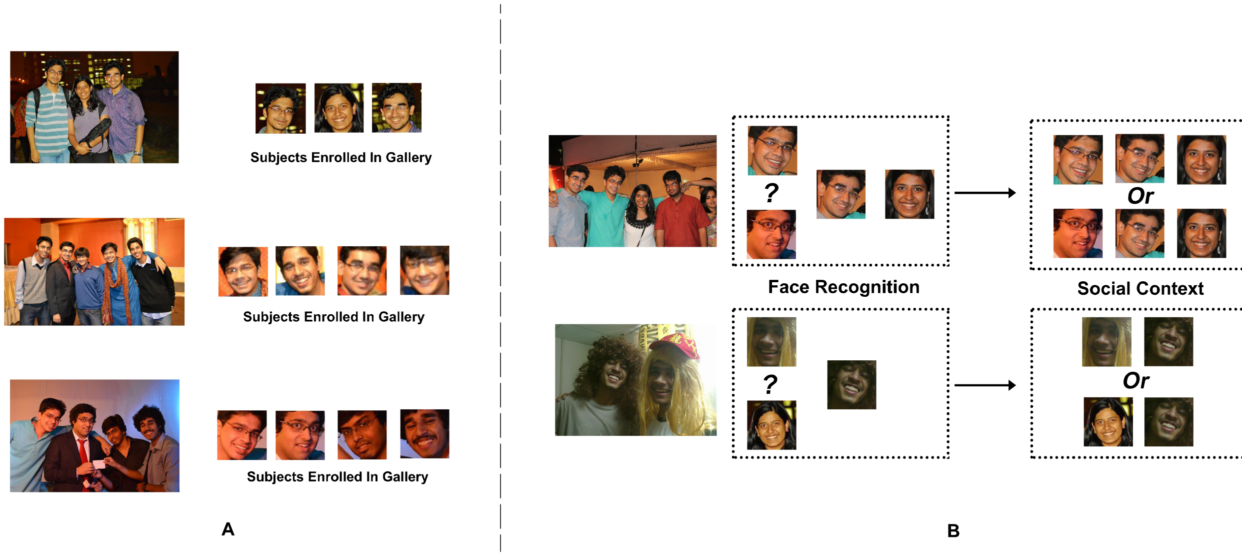


Figure 1: An example of how social context can improve face recognition in the case of group photographs.

bine ranked lists obtained after face recognition with social context information to improve the recognition performance. While similar concept has been explored by Bharadwaj *et al.* [9], we propose a different perspective by using a social graph to represent the social context. The edge weights and network structure of the social graph are incorporated into a face recognition algorithm which can combine social context with traditional methods. Social graphs and networks are better equipped to represent social relationships since people can be connected via mutual friends even if they are not directly connected in a social graph. These relationships and their strengths are intuitive to model and utilize when represented in the form of a graph. Another benefit of using such an approach is that the existing and future techniques for learning such graphs from social media data can be adapted into the algorithm for further robustness and scalability. Further, the proposed algorithm only requires subject identifiers for each training image, which makes it computationally efficient during training. It is evaluated on two publicly available datasets to enable comparison with different social context based face recognition algorithms. The major contributions of this research can be summarized as follows:

- A novel algorithm to create a social graph from a labeled collection of face photographs is proposed.

- A novel algorithm to combine social context information with face recognition ranked lists is proposed. The proposed algorithm is evaluated on two publicly available datasets and different methods of interpreting and incorporating social context are assessed.

2. Proposed Algorithm

A brief overview of the proposed algorithm is presented in Figure 2. Given a set of gallery images/videos pertaining to groups of individuals, we learn facial as well as social context information. While facial information is utilized by a face recognition algorithm to perform traditional face recognition, social context is utilized during testing to improve the results. The procedure for learning facial features from gallery data depends on the selected face recognition algorithm. In order to learn social context information, a social graph is created in which each individual is represented as a node, the relationships between individuals are denoted by edges, and the strength of these relationships is measured by the edge weights. When a video/image is presented to the system for recognition, the face recognition results obtained by matching facial features are retrieved using the specified face recognition algorithm. These results contain ranks and match scores for each possible match found in the gallery. Social context scores are then computed for each match which are combined with existing results to obtain the final recognition result. In the following subsections, we provide the details of the steps involved in the proposed algorithm.

2.1. Social Graph Creation

A traditional face recognition algorithm detects face images from the given photograph, segments the face portion(s) from the image, and extracts discriminative features from it. These extracted features are stored and utilized to perform matching and recognition during runtime. If group videos/photographs from social media are provided to such an algorithm, it loses out on the valuable information about the co-occurrence of different individuals in the

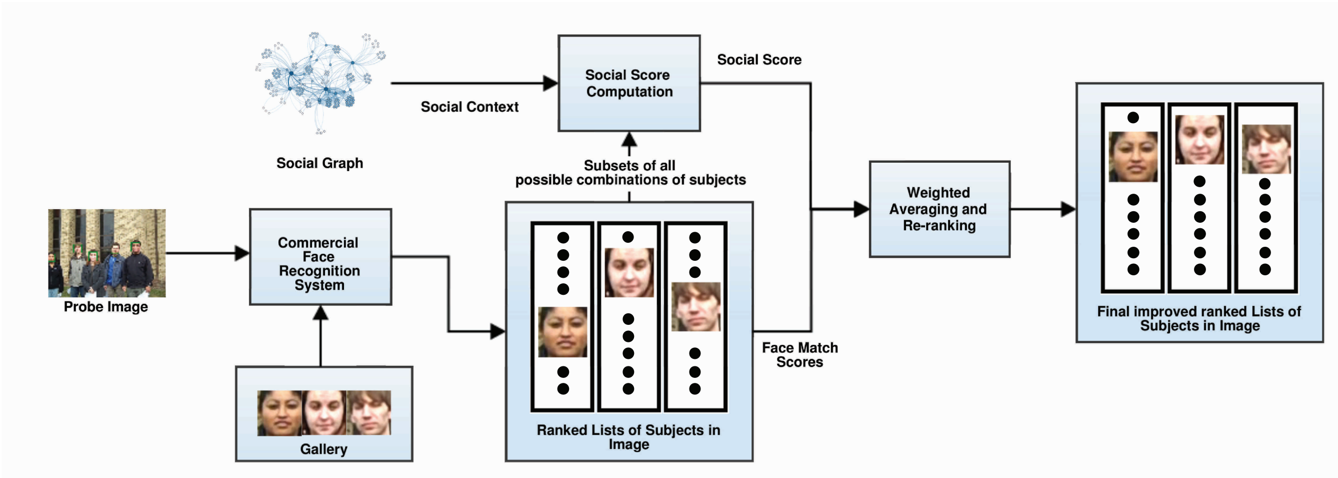


Figure 2: An overview of the proposed social context augmented face recognition algorithm.

same photograph/video. The proposed algorithm attempts to utilize this information as well. During training, the proposed algorithm creates a social graph from labelled group videos/photos which contains social context information. The procedure for generating a social graph from a collection of images is detailed in algorithm 1.

First, the vertices and edges of the social graph ($[\mathbf{V_{gallery}}, \mathbf{E_{gallery}}]$) are initialized to an empty set and then the algorithm iterates through each image in the gallery. All the subject identifiers ($\mathbf{S}$) present in each image are added to the social graph as vertices if they are not a member of the graph already. Then each of the subject identifiers that share the same image are paired with every other subject identifier in the image and edges are created using all such possible pairs of subjects. If an edge is already present in the social graph its weight is incremented by one, otherwise a new edge is added to the graph with an initial weight of one. After the procedure completes iterating through all the gallery images, the social graph contains all the unique subject identifiers in the form of vertices. Each edge weight records the number of images in which the connected subjects co-occur in the gallery. Edge weights are normalized before being combined with face recognition match scores in order to ensure that both lie in the same range of $[0, 1]$. Two different kinds of normalization techniques are explored and a comparison of these is presented in the results section. The social graphs extracted from the Gallagher and SN-Flip datasets are presented in Figure 3.

2.2. Augmenting Face Recognition with Social Context: Proposed Algorithm

The social graph maps the network of social connections that exist among the users enrolled in the gallery. The premise of the proposed algorithm is that if the weight of an edge is high then it implies that the individuals connected via that edge have co-occurred in a large number of gallery images and it is highly likely that they might appear together again in a probe image during runtime. Subsequently, if a certain combination of subjects have a low edge weight or no edge connecting them, it is less likely for them to appear in the same image in future. The proposed algorithm utilizes such cues and combines them with confidence scores obtained using face recognition algorithms.

Input: A set of gallery images: $\mathbf{G}$ with corresponding image-wise subject identifiers $\mathbf{S}$
Output: A social graph of the gallery: $[\mathbf{V_{gallery}}, \mathbf{E_{gallery}}]$
Algorithm:
$\mathbf{V_{gallery}} = \phi$
$\mathbf{E_{gallery}} = \phi$
for each G_i in $\mathbf{G}$:
  for each S_{i_j} in $\mathbf{S_i}$:
    if $S_{i_j} \notin \mathbf{V_{gallery}}$:
      $\mathbf{V_{gallery}} = \mathbf{V_{gallery}} \cup S_{i_j}$
    for each $S_{i_k} \forall k \neq j$ in $\mathbf{S_i}$:
      if edge $(S_{i_j}, S_{i_k}) \notin \mathbf{E_{gallery}}$:
        add edge (S_{i_j}, S_{i_k}) to $\mathbf{E_{gallery}}$
      else:
        $E_{gallery}(S_{i_j}, S_{i_k}) += 1$

Algorithm 1: Proposed social graph creation algorithm

The proposed approach is explained in algorithm 2. First, the face recognition match score lists $\mathbf{F}$ are obtained using a Commercial Off the Shelf face recognition system. $\mathbf{F}$ contains one match score list for each of the detected subjects in a probe image. The match score list for the i^{th} de-

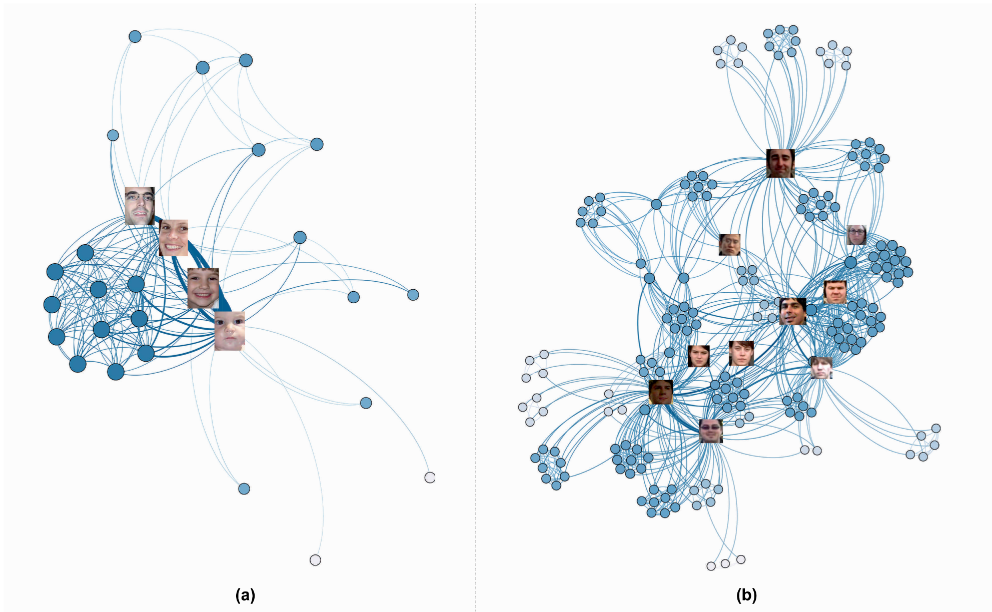


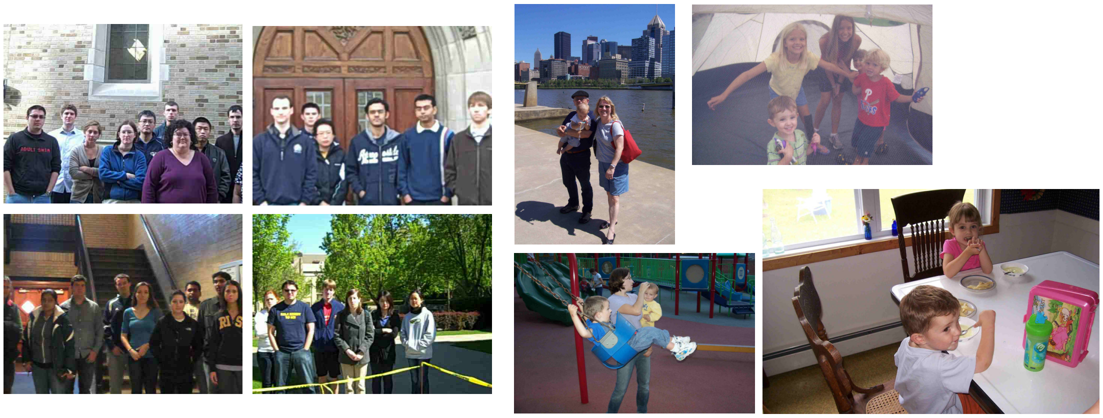
Figure 3: Social graph created from the (a) Gallagher dataset and (b) SN-Flip dataset in Yifan Hu layout. Darker/Larger nodes indicate a higher degree of the node.

Input: Face recognition match score lists $\mathbf{F}$ and the social graph $[\mathbf{V_{gallery}}, \mathbf{E_{gallery}}]$
Output: Final match scores $\mathbf{M}$
Algorithm:
$\mathbf{R} = F_1 \times ... \times F_n = \{(p_1, ..., p_n) | p_i \in F_i \forall i = 1, ..., n\}$
for each relation $R_i \in \mathbf{R}$:
  if $\{p_j, p_k\} \in \mathbf{E_{gallery}} \forall p_j, p_k \in R_i, j \neq k$:
    $C_i = \alpha \times \sum E_{gallery}(p_j, p_k) \forall p_j, p_k \in R_i, j \neq k$
  else:
    $C_i = 0$
for each match score list $F_i \in \mathbf{F}$:
  for each person $p_j \in F_i$:
    $r_j = relation\ with\ maximum\ C\ value\ containing\ P_j$
    $X_j = F_i[p_j] \times C_{r_j}$
    for each person $p_k \in R_{r_j}$:
      $X_j = X_j \times F_l[p_k] \quad p_k \in F_l$
    $M_j = w_{sc} \times X_j + (1 - w_{sc}) \times F_i[p_j]$

Algorithm 2: Proposed algorithm for combining social context with conventional face recognition

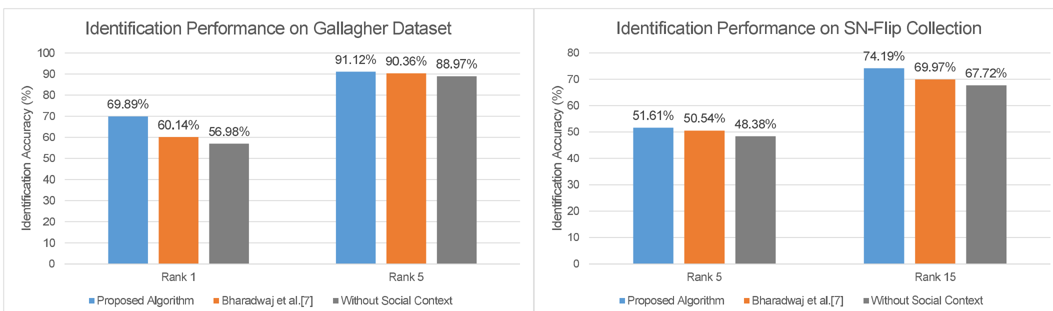
tected person in a probe image is denoted by F_i. The match score list F_i contains a list of possible candidate identities for the i^{th} person in the probe. Each entry in the list is a person from the gallery along with the corresponding match score. The j^{th} person in F_i is denoted by p_j and its match score is denoted by $F_i[p_j]$. This match score $F_i[p_j]$ signifies the confidence of the face recognition system that the true identity of the i^{th} individual in the probe is p_j.

In the proposed algorithm, we augment this match score with a social context score, X_j. To compute this score, a set of all possible social relationships between the candidate subjects in all the match score lists is created by computing the Cartesian product $\mathbf{R}$ of all the subjects in $\mathbf{F}$ and ignoring any entry with duplicates. A social relationship can have as little as two subjects and as many as the number of detected subjects in the probe. Next, each relation is evaluated for social connectivity, i.e., all the members of each relationship should be directly connected to each other in the social graph created from the training data as explained in Section 2.1. Two people are considered to be directly connected if there is an edge that connects them. A relationship score C_i is computed for the i^{th} relation which is 0 if the aforementioned social connectivity criterion does not hold. Otherwise it is equal to the product of all the edge weights of unique pairs in the relation. α is a parameter of the proposed algorithm which acts as a proportionality constant. In order to compute the social context score X_j for the j^{th} subject in the i^{th} list, all of the relationships that contain p_j are evaluated and the relationship which has the maximum possible social context score r_j is selected. This score is then multiplied by the face recognition score of each subject in the relation r_j to obtain the final social context score for the j^{th} subject in the i^{th} list. The final step is to then combine face match score with the social context score. A weighted sum approach is utilized to combine these scores, w_{sc} is the social context weight which is another parameter of the algorithm. The final match scores $\mathbf{M}$ are utilized to re-order



(a) (b)

Figure 4: Sample images from the (a) SN-Flip dataset [2] and (b) Gallagher dataset [7].



(a) (b)

Figure 5: Demonstrating the effect of using social context to augment face recognition: (a) Gallagher dataset [7] and (b) SN-Flip dataset [2].

the subjects in each rank list and perform recognition.

3. Experimental Results

The proposed algorithm is evaluated on two publicly available datasets: The Gallagher Collection Person Dataset [7] and the SN-Flip Crowd Video Database [2]. Sample images from these datasets are presented in Figure 4. These datasets are selected because of their inherent social nature. The commercial-off-the-shelf system provided by FaceVACS [1] is utilized for face recognition. In the following subsections, we describe the database, experimental protocol, and results of the proposed algorithm on two databases. We also present a comparative analysis of the proposed algorithm with another context aware face recognition algorithm proposed by Bharadwaj *et al.* [9].

3.1. Experiments on The Gallagher Collection Person Dataset

The Gallagher dataset consists of 589 personal images of 32 subjects, representing the photographs taken in a family setting. Since these photos are representative of typical photos of an event or a gathering, most of the subjects appear in two or more photos with an average of 18.4 images per subject and 1.56 subjects per image. This high co-occurrence of subjects allows for the creation of a social context graph with more accurate edge weights. The dataset is split in 50:50 gallery to probe ratio such that for every subject appearing in two or more images, at least half of that subject's images are used as the gallery and the remaining are used as the probe. This is accomplished by finding overlapping images for two subjects and including the common image in gallery. After this split, 72 out of 468 probe faces and 49 out of 455 gallery faces were discarded as the COTS was unable to detect them. Since these faces are deemed unfit for processing by the COTS itself, i.e., these faces incur the Failure to Enroll (FTE) error, these faces are not included in any recognition experiment. Therefore, these faces do not contribute to the performance improvement observed using the proposed algorithm.

The performance of the proposed algorithm is compared with traditional face recognition as well as Bharadwaj *et al.* [9] and results are shown in Figure 5(a). We observe a 12.91% improvement at rank 1 and 2.17% improvement at rank 5 on the Gallagher dataset as opposed to traditional face recognition. These results are encouraging and suggest that social context can indeed improve the recognition performance in a substantial manner for scenarios where it can be applied. In comparison with Bharadwaj *et al.* [9], the proposed algorithm achieves higher performance gain, i.e. the proposed algorithm achieves about 12% improvement at rank 1 accuracy whereas existing algorithm achieves an improvement of around 3% by using social context.

Normalization of edge weights in the social graph is important since the range of face recognition match scores is $[0, 1]$ and the range of edge weights is $[0, N]$, where N is the total number of images in the gallery. In order to do so, in our experiments, we have performed min-max normalization. We have also conducted various experiments to analyze the effect of parameters involved in the proposed algorithm on the identification performance. We have observed that the optimal value for α is 0.5 and w_{sc} is 0.9 which implies that social context score alone does not offer the best performance and the optimal approach is to combine it with the original face recognition score.

3.2. Experiments on SN-FLIP Dataset

The SN-Flip dataset contains 28 crowd videos of 190 subjects, recorded using a consumer grade Cisco Flip camera. These videos are representative of the quality of typical videos shared on social media. The dataset is split in 50:50 ratio (50% Gallery, 50% Probe) in a pseudo-random manner similar to the approach followed with the Gallagher Collection Dataset. Since the dataset consists of videos, face frames are used as gallery and training images. After face recognition scores for all probe face frames are obtained, the grouped face recognition score for each face track is computed by taking the average of face match scores obtained for frames in that track. The graph is generated using the co-occurrence of two subjects in a video the information for which was obtained from the ground truth provided in the dataset.

Identification performance of the proposed algorithm on the SN-Flip dataset is presented in Figure 5(b). COTS yields rank-5 and rank-15 accuracies of 48.38% and 67.72% respectively. Incorporating social context improves the accuracies by 3.23% and 6.47% respectively. On the same protocol, Bharadwaj *et al.* [9] yields improvements of 2.16% and 2.25% in rank-5 and rank-15 accuracies. Compared to the Gallagher Personal Collection database, the percentage improvement is lower on the SN-Flip database because the number of videos in the SN-Flip dataset is smaller and the social graph is only extracted from limited labeled information, leading to not a very strongly connected graph.

The results on the two databases show that the proposed algorithm requires good amount of training data to generate the social context graph. If the social context data is available in abundance, the proposed algorithm can provide significant improvements to the results of face recognition.

4. Conclusions and Future Work

In this paper, we propose a novel algorithm to create a social graph from a gallery of group photographs and then utilize the graph to obtain social context information. The social context is then combined with face recognition score to perform efficient face identification. Based on the results demonstrated on two publicly available datasets, we can infere that the proposed algorithm is successfully able to fuse social context with face recognition to improve performance. As part of future work, we plan to create a large database of social photographs and evaluate the algorithm in terms of scalability and efficiency when the extracted social graph is large and complex.

References

[1] `http://www.cognitec.com`.

[2] J. Barr, L. Cament, K. Bowyer, and P. Flynn. Active clustering with ensembles for social structure extraction. In *WACV*, pages 969–976, 2014.

[3] M. Bloess, H.-N. Kim, M. Rawashdeh, and A. El Saddik. Knowing who you are and who you know: Harnessing social networks to identify people via mobile devices. In *Advances in Multimedia Modeling*, pages 130–140. 2013.

[4] Y.-Y. Chen, W. H. Hsu, and H.-Y. M. Liao. Discovering informative social subgraphs and predicting pairwise relationships from group photos. In *ACMMM*, pages 669–678. ACM, 2012.

[5] T. Dhamecha, R. Singh, M. Vatsa, and A. Kumar. Recognizing disguised faces: Human and machine evaluation. *PLoS ONE*, 9(7):e99212, 2014.

[6] G. Goswami, R. Bhardwaj, R. Singh, and M. Vatsa. MDL-Face: Memorability Augmented Deep Learning for Video Face Recognition. In *IJCB*, 2014.

[7] A. Gallagher and T. Chen. Clothing cosegmentation for recognizing people. In *CVPR*, 2008.

[8] S. Z. Li. *Handbook of face recognition*. Springer, 2011.

[9] S. Bharadwaj, M. Vatsa and R. Singh. Aiding face recognition with social context association rule based re-ranking. *IJCB*, 2014.

[10] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In *CVPR*, volume 2, pages 994–1000, 2005.

[11] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. IEEE*, 94(11):1948–1962, 2006.