

Near real-time face detection and recognition using a wireless camera network

Francesco Nicolo, Srikanth Parupati, Vinod Kulathumani, Natalia A. Schmid

Dept. of Computer Science and Electrical Engineering
West Virginia University, Morgantown, WV 26506

ABSTRACT

We present a portable wireless multi-camera network based system that quickly recognizes face of human subjects. The system uses low-power embedded cameras to acquire video frames of subjects in an uncontrolled environment and opportunistically extracts frontal face images in real time. The extracted images may have heavy motion blur, small resolution and large pose variability. A quality based selection process is first employed to discard some of the images that are not suitable for recognition. Then, the face images are geometrically normalized according to a pool of four standard resolutions, by using coordinates of detected eyes. The images are transmitted to a fusion center which has a multi-resolution templates gallery set. An optimized double-stage recognition algorithm based on Gabor filters and simplified Weber local descriptor is implemented to extract features from normalized probe face images. At the fusion center the comparison between gallery images and probe images acquired by a wireless network of seven embedded cameras is performed. A score fusion strategy is adopted to produce a single matching score. The performance of the proposed algorithm is compared to the commercial face recognition engine Faceit G8 by L1 and other well known methods based on local descriptors. The experiments show that the overall system is able to provide similar or better recognition performance of the commercial engine with a shorter computational time, especially with low resolution face images. In conclusion, the designed system is able to detect and recognize individuals in near real time.

Keywords: Camera network, collaborative face detection, face recognition, score fusion, near real-time processing.

1. INTRODUCTION

Face recognition systems have traditionally been applied for identification from previously acquired photographs and videos. But it is now becoming increasingly important to apply these systems for human identification in real-time environments. In this paper, we describe the design and implementation of a distributed face recognition system using a network of embedded cameras that is able to acquire and process face images of human subjects in near real-time. We have specifically considered a *choke-point* scenario where a network of cameras are deployed over small critical access regions such as entrances, lobbies, walkways etc. in public places for human identification. Potential applications for such a system include raising an alert when criminals are detected entering critical places such as airports and shopping malls, raising an alert when lost individuals are detected and automatic recognition and registration of individuals at international port of entry.

Distributed camera networks can potentially improve the accuracy of human identification by offering multiple views of a scene and being tolerant to occlusions. However, in order to be scalable and to be applicable in real-time identification, multi-camera networks pose a design challenge in terms of the trade-off required between local processing and centralized computation. On the one hand local processing is needed prior to transmitting all the acquired data to a fusion center so that the network and the fusion center are not overloaded with too much data. On the other hand, it is also crucial to acquire as many images as possible at the individual cameras so as to improve the chances of accurate recognition. This is because a subject could be constantly moving his or her head and there could be a small duration when the face pose with respect to a camera is favorable for identification and it is desirable to capture this frame. Therefore, it is important to keep computational overhead

E-mail: {fnicolo,sparupat}@mix.wvu.edu, {Vinod.Kulathumani,Natalia.Schmid}@mail.wvu.edu

low at the distributed cameras so that the data can be processed at a high frame rate. Our approach in this paper is to use the distributed cameras to perform face detection and transmit only the region containing faces detected in each frame to the base station.

Contributions: We consider a system comprising a set of embedded cameras deployed along the sides of a narrow alley or corridor that simulates passages in indoor public spaces such as airports. We use a Logitech 9000 series camera,¹ a Beagleboard² (with an OMAP processor), and a wireless card to assemble an individual embedded camera unit. Each embedded camera performs pre-processing operations such as background subtraction to detect an individual in a given scene. This is followed by a Haar cascade classifier-based face detection^{3,4} and a filtering operation to suppress images that might not be suitable for face recognition due to various reasons (e.g., poor resolution, bad pose, motion blur etc.). Only frontal face images likely to yield high recognition accuracy are transmitted to a fusion center for face recognition. Using this setup, we perform an experiment with 29 individuals walking through a network of 7 embedded cameras. The probe images that are received at the fusion center are then used for face recognition by matching against a gallery database.

In previous work,⁵ we have used a commercial software (L1 Identix FaceIt) to verify the achievable recognition performance using a multi-camera network and showed how multiple cameras improve the chance of acquiring a suitable face image for recognition. In this paper, we use the same dataset and introduce a new face recognition method which shows promise in terms of performance and speed when applied to low-resolution images.

Prior to performing recognition, each image is normalized to a standard size and then processed with a bank of Gabor filters. Filtered images are encoded using a simplified Weber descriptor. The number of standard resolutions, of Gabor filters and the scale parameter of the local descriptor are tuned to guarantee near real time classification of subjects. The operator can be applied to encode and match images for both single spectrum and for performing cross spectral matching (visible light vs. SWIR).

Related work: Multiple cameras have been used for face recognition in an active control mode in which a combination of a fixed camera and a PTZ camera are used for close-up tracking of humans and subsequent identification.⁶⁻⁸ In our approach instead of continuously tracking an individual at close quarters to eventually get a good view that is suitable for recognition, we rely on redundancy offered by multiple camera views to opportunistically acquire a suitable face image for identification.^{5,9}

In the paper by D. Kisku et al.,¹⁰ Gabor filters are used to encode face at multi-views (frontal and non-frontal) and the filtered data is further encoded with PCA and a generalization of LDA (Linear Discriminant Analysis). The obtained feature vectors are linearly combined with weights estimated on a training set. The classification is performed with three different classifiers, namely, K-Nearest Neighbor, SVM with linear kernel and SVM with RBF. The algorithm is designed to work off-line and it is tested on the UMIST¹¹ dataset having controlled and good quality images (200 pixels of resolution, no blur). By way of contrast, our recognition technique encodes Gabor filters responses with a local operator, preserving spatial information and providing a distinctive representation of facial features.

In the work of E. Kokiopoulou et al.,¹² the problem of distributed classification of an object is addressed based on multiple views of the object that are collected using an ad-hoc camera network. An average consensus algorithm is implemented for estimating the unknown object class by adopting an objective function. At the initial stage the objective function is locally estimated based on observation of a single node. As the algorithm progresses all observations are gradually taken into account in the estimation of the objective function. In our approach for distributed classification, instead of using observations from all the cameras, we implement a score fusion strategy that discards images of low quality which generate low matching scores.

Outline of the paper: In Section 2, we present our embedded camera network based face detection and filtering service and describe our experiments to acquire face images in real-time. In Section 3, we describe our face recognition technique. In Section 4, we evaluate the performance of our recognition system and conclude in Section 5.

2. FACE DETECTION

The system that we consider in this paper consists of a long, linear network of embedded cameras deployed along both sides of a secured passageway (e.g., an aisle or a corridor). This simulates an indoor walkway in a public

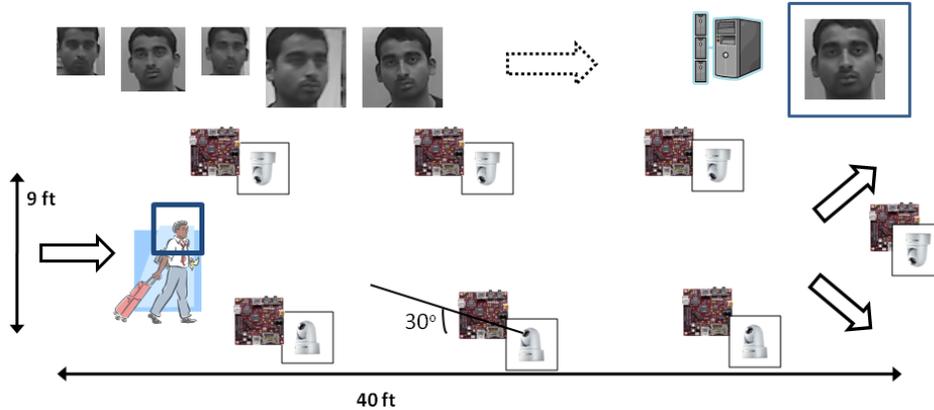


Figure 1. Schematics for network based face recognition: 7 embedded cameras are deployed along a rectangular corridor. The cameras opportunistically capture images of subjects within their field of view, apply pre-selection and filtering algorithms and transmit to a fusion center for recognition.

space that is subjected to surveillance. The embedded cameras are connected wirelessly to a base station via a single hop. Multiple cameras capture images of the subject(s) walking through the network. The acquired images may contain variations in pose, resolution and even illumination, depending on the external conditions. The requirement is for the subject to be recognized with high confidence using the set of images captured within the camera network.

2.1 Assembly of embedded camera network

For the experiments described in this paper, we use the OMAP3 processor based BeagleBoard² as our embedded camera processor. The processor supports Linux operating system and can be integrated with off-the-shelf USB enabled cameras generating medium or high resolution images. For our experiments, we use the Logitech 9000 camera. While currently we use an 802.11 based wireless card, we can easily replace this with a low power wireless network using IEEE 802.15.4 enabled transceiver. Thus, the assembled system provides us with flexibility with respect to the camera as well as the radio platforms. The assembled platform can be powered by a 5V battery or external AC power.

Next, we assemble a camera network using the above platform. Individual units are portable and are configured into a programmable network that is attached to a fusion center for recognition. Fig. 1 shows a schematic of the camera network for face recognition as described in this paper. Seven cameras are placed at a height of 6.5 feet from the ground with a 9 feet spacing along a length of 40 feet. In this work, we focus on the case where the cameras are placed at the specified height to acquire frontal or partially frontal face images. Using overhead or ceiling-placed cameras for face recognition is beyond the scope of this work. We do not use camera calibration information for face recognition. Therefore, the cameras do not have to be tightly calibrated which simplifies the logistics of deployment. The cameras are deployed facing the region under surveillance but biased toward the direction of the entrance at an angle of 30 degrees. This allows us to acquire frontal face images when the subject looks towards the direction of the exit. This is highlighted in Fig. 1. The individual camera units can be programmed at run time. The cameras form a 1-hop network, but can be extended to form a multi-hop network.

2.2 Distributed processing

We use an OpenCV based implementation of the Haar Cascade detector⁴ that is tuned to extract face segments for frontal face detection. Sometimes, the Haar Cascade detector generates false positives for detected faces by including side views as well as non-facial images. In order to reduce these false positives, we perform an additional step of eye detection on the output of the face detector. The eye detector algorithm is also a Haar Cascade based classifier that is trained to detect eyes in an image. The extracted face images are then further analyzed for size and motion blur. We use Discrete Cosine Transformation (DCT) to estimate and detect motion blur. DCT detects low and high frequency components, and the degree of motion blur can be estimated from

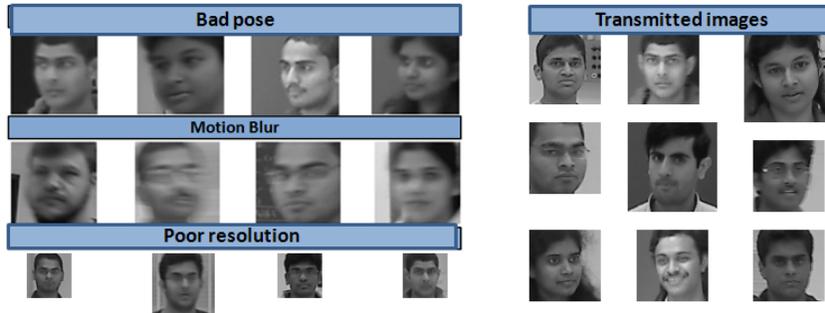


Figure 2. (Left) Images that did not qualify for transmission to the fusion center due to bad pose (top), motion blur (middle), poor resolution (bottom); (Right) Example face images that were transmitted to the fusion center.

the number of high frequency components.¹³ We determine the presence of motion blur by observing the DCT coefficients of the top 10% of the high frequency components. During the process of filtering images prior to transmission to the fusion center, we apply a predetermined threshold on the size of a face image and the blur factor to discard frames. These thresholds are empirically decided, based on an offline analysis of the impact of different quality images on the match scores. We empirically determine a size of 22×22 pixels as the minimum resolution for a detected face to be transmitted to the fusion center. If blurred and poor resolution images were to be transmitted, it will result in excess processing time at the fusion center without adding value to the fusion. Fig. 2 (left) shows example face images that were filtered (i.e., dropped) due to poor pose, blur and resolution, respectively. Fig. 2 (right) shows example face images that were transmitted to the fusion center.

At the end of this phase, a total of 223 face images are obtained for all the 29 subjects with a minimum of 5 and a maximum of 12 images per subject. These face images are used to test the performance of our recognition technique and compare with other algorithms. We save approximately 90% of the bandwidth on each transmitted frame by extracting only the face portion of the image. Furthermore, by transmitting only a small subset of frames, we significantly reduce the required network bandwidth.

3. FACE RECOGNITION

3.1 Preprocessing and normalization

The detected face images are brought to a canonical form prior to being encoded. A similarity transformation (scaling, rotation, translation) is employed to project eye locations to two fixed points in the image space. The selected points correspond to the eye coordinates of a single face image taken as a reference. The positions of the eyes are detected with Haar features and an AdaBoost classifier³ using the OpenCV libraries.⁴ To reduce the number of false alarms we enforce few geometrical constraints: detection areas for left and right eyes, maximum vertical distance and minimum horizontal distance between candidate landmarks. In a surveillance scenario, images are usually captured in unconstrained environments. The distance to the camera is unknown, and various distortions are observed in the images. In particular, the acquired face images may have variable resolution (the number of pixels representing the face in the image). Hence, before we detect the eye coordinates we downsample high resolution face images and upsample low resolution face images to deal with face images at variable scale. This strategy is adopted only for the purpose of eye detection.

At the stage of feature extraction during the enrollment, instead of defining a standard resolution for resampling the normalized images, we create feature templates at different image resolutions (a predefined set). This solution avoids excessive upsampling to a given resolution when captured images are extremely small (e.g. 28×28 pixels). The adopted strategy increases the processing time at the enrollment stage, which is usually done off-line, but has the advantage of being able to decrease the average processing time at the testing stage. Particularly, low resolution images will be processed much faster compared to high resolution images. A detailed summary is provided in Sec. 4.1 and Sec. 4.4.

3.2 Feature extraction

The proposed feature extraction method is based on a double encoding mechanism that adopts a bank of 32 Gabor filters as the first step and a Simplified Weber Local Operator (SWLD) as the second step.¹⁴ The Weber Local Descriptor was introduced by Chen et al.¹⁵ It consists of two joint descriptors: a differential excitation operator and a gradient orientation descriptor. In this work we adopt only the differential excitation operator, and we employ it to encode the magnitude response of the Gabor filters. The SWLD is a robust edge detector that preserves the intensity values along the edges. We compare the performance of SWLD with that of the following alternate approaches: (1) commercial software FaceIt G8 by L1, (2) the method of Zhang et al.¹⁶ by using the magnitude responses of a bank of Gabor filters and (3) the method of Ahonen et al.¹⁷ that employs a 8-bits LBP operator directly on raw images. We show that our approach achieves better verification performance than Gabor-LBP,¹⁶ LBP,¹⁷ and comparable or better verification performance to the Faceit G8 software. We also demonstrate that our method can be extremely fast when applied to low resolution images compared to FaceIt G8 that has a processing time almost independent of the resolution.

To extract face features, firstly, the normalized images are convolved with a bank of 32 Gabor Kernels given by:

$$G_{\alpha,\beta}(z) = \frac{\|k_{\alpha,\beta}\|}{\sigma^2} \exp\left[\frac{\|k_{\alpha,\beta}\|^2 \|z\|^2}{2\sigma^2}\right] [e^{ik_{\alpha,\beta}z} - e^{-\sigma^2/2}], \quad (1)$$

where σ^2 is the variance of the Gaussian kernel, $k_{\alpha,\beta}$ is the wave vector and $z = (x, y)$. We adopt a Gabor kernel parameterized by the following wave vector:

$$k_{\alpha,\beta} = k_{\beta} e^{i\phi_{\alpha}}, \quad (2)$$

where $k_{\beta} = (2^{-\beta/2})\pi$ with $\beta = 1, 2, 3, 4$ and $\phi_{\alpha} = \frac{\pi}{\alpha} \times 8$ with $\alpha = 1, 2, \dots, 8$. The Gaussian kernel has standard deviation $\sigma = \pi$.

Secondly, each Gabor magnitude response is further encoded with the SWLD descriptor defined as:

$$SWLD_{l,r,s}(x) = \mathcal{Q}_l \left\{ \tan^{-1} \left[\sum_{i=0}^7 \left(\frac{x_i - x}{x} \right) \right] \right\}, \quad (3)$$

where x_i are the neighbors of x at radius $r = 1$ and \mathcal{Q}_l is a uniform quantizer with l quantization levels. In the following experiments we adopt $l = 64$ levels to discretize the output of the \tan^{-1} function.

Finally, all obtained patterns are partitioned into non-overlapping regions of 8×8 pixels, and a 64 bin histogram of feature values is assembled for each block. Individual histogram sequences are concatenated resulting in a template vector. We note that the overall feature extraction process does not require any kind of training.

To compare two images A and B we adopt the symmetric I-divergence distance as a matching score:

$$d(A, B) = \sum_{k=1}^K (H_A(k) - H_B(k)) \log \frac{H_A(k)}{H_B(k)}, \quad (4)$$

where K is the length of the feature vectors $H_A(k)$ and $H_B(k)$ extracted from images A and B . The feature vectors are normalized such that matching pairs at different resolutions produce distance scores within the same range.

4. PERFORMANCE EVALUATION

4.1 Dataset and Preprocessing

During enrollment 5 images per subject are collected at close distance from a single camera to obtain a total 145 gallery images. Also recall from Section 2, that we obtain 223 probe images with an average of about 7 images per face class using our multi-camera network based face detection experiments. The captured probe images have an out-of-plane rotation up to ± 20 degree with respect to the frontal view. They also have variable

resolution and different amount of blur involved due to the limited capture volume of cameras in the network. The eye detector is applied to the gallery images (at the enrollment stage) obtaining a total of 136 detected gallery images and to probe images (at the testing stage) resulting in a total of 216 probe images. The eye detection rate can be improved by adding an eye-glasses detector. However, to keep the processing time low we adopt a single Haar cascade trading off complexity versus performance. The normalized probe images are partitioned into 7 resolutions: 28×28 , 38×38 , 47×47 , 56×56 , 75×75 , 103×103 , and 122×122 pixels. The distribution of probe images grouped by their original resolution is shown in Fig. 3. Note that most of the probe images have low resolution.

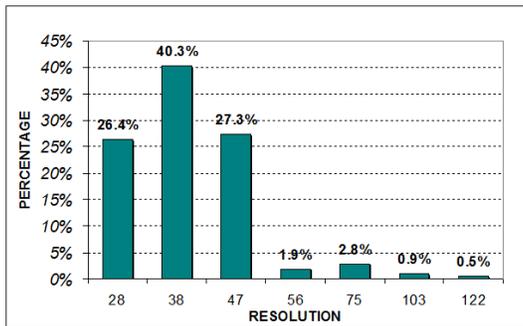


Figure 3. Probe images grouped by resolution.

As explained in Sec. 3.1, we predefine a set of resolutions to rescale the probe images prior to encoding and matching them. All high resolution probe images (6% of all probe images) are scaled down to 64 pixels; instead, images with resolution 47, 38, and 28 pixels are mapped to 48, 40, and 32 pixels by using a bicubic interpolation. In general, once the sets of standard resolutions have been defined, the incoming probe images can be rescaled to the closest one in this set.

In the following two subsections, we analyze the performance of our recognition system. We first evaluate the performance by considering all the probe images that are collected for the 29 subjects in computing the receiver operating characteristics. Next, we evaluate the performance after employing a score fusion strategy by which for each subject, the probe image that yields the best score with respect to each gallery image is selected and the receiver operating characteristics are computed based on these selected probe images.

4.2 Recognition Performance based on scores from all probe images

In this section we evaluate verification performance (Receiver Operating Characteristic curves) of the proposed encoding and matching method. The performance of the method is compared to the performance of the commercial software FaceIt G8, of the Gabor-LBP approach and of the LBP matcher. FaceIt G8 employs a similarity measure as matching score while Gabor-SWLD (64 levels), Gabor-LBP (8 bits), LBP (8 bits) employ a symmetric I-divergence distance. Other measures of distance such as Chi-Square¹⁷ or Histogram Intersection,¹⁶ can be adopted. Based on our experiments, the I-divergence distance performs reliably compared to other metrics. In the following experiments the set P of 216 probe images is matched against the gallery set G composed of 136 images. Each probe image is individually submitted for verification. The resulting Receiver Operating Characteristic (ROC) curves from this experiments are shown in Fig.4

The results show that LBP matcher provides very poor performance. The performance of the Gabor-LBP approach is lower compared to the performance of the Gabor-SWLD (our approach). The FaceIt G8 software provides best performance for the operational points in the range 0.05 to 1 of False Accept Rate (FAR). Its performance is similar to the performance of the Gabor-SWLD matcher for the operational points in the range 10^{-4} to 0.05 of FAR. The cumulative results show that all four matchers provide low verification rates ($< 50\%$) at FAR below 10^{-2} .

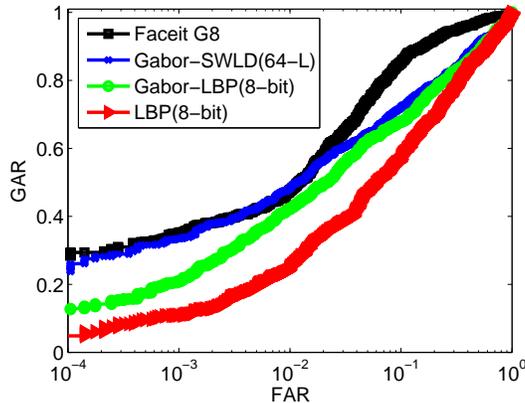


Figure 4. Verification performances for FaceIt G8, our method (Gabor-SWLD), Gabor-LBP, LBP matchers for a single camera.

4.3 Recognition performance after selecting probe images based on score-fusion

Prior to performing experiments, we group the set P of 216 probe images (see its description in Sec. 4.1) into 29 groups (the number of groups is equal to the number of walking subjects $i = 1, 2, \dots, 29$). Let $Q_i, i = 1, \dots, 29$ be the i -th subset with the images of the i -th class. At the fusion center, we compare each group Q_i of probe images against all images in the gallery G . Each group Q_i contains a different number of probe images (the minimum number is 4 and the maximum number is 10 images) having different resolutions, pose and motion blur level. Since the image collection process was repeated a number of times for each subject walking in the field of view of the network to obtain a larger number of testing images, in the following experiments, we assume that any group of images Q_i is a set of images acquired from an equivalent larger camera network (10 cameras) where not always all cameras were able (e.g. because of pose or occlusion) to detect the face of the subject walking in the network.

To compare each group Q_i of probe images against the images in the gallery G , we employ a score fusion strategy. In particular, we adopt the max or min rule¹⁸ depending on the tested matcher. At the fusion center k_i images of the current group Q_i are compared against the first image of the gallery G and the lowest matching score among the k_i is selected. This strategy is applied to Gabor-SWLD, Gabor-LBP, LBP matchers, since the output of these matchers is a distance score. When FaceIt G8 is employed, the adopted fusion rule selects the maximum of the scores, since the output of FaceIt G8 is a similarity score. This normalization process is applied to all comparisons performed against remaining gallery images. The ROC curves characterizing the verification performance of the multiview face recognition method are shown in Fig. 5.

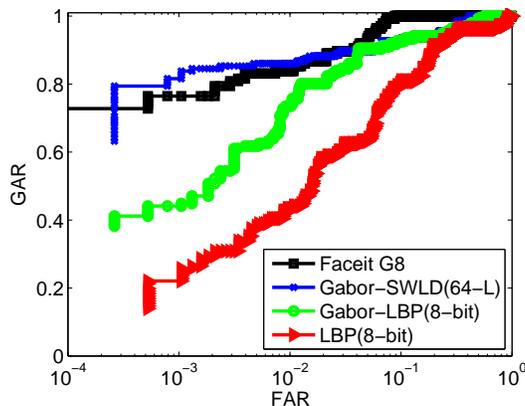


Figure 5. Verification performance for FaceIt G8, our method (Gabor-SWLD), Gabor-LBP, and LBP matchers using a camera network detecting an average of 7 probe images per walking subject.

The results demonstrate that the use of a multi-camera network for acquiring data and the use of a score fusion strategy leads to considerable performance improvement compared to the results in Fig. 4. This is especially noticeable for Gabor-SWLD approach and FaceIt G8 matcher. Note that our matcher performs best in the FAR range between 10^{-3} and 10^{-2} , while FaceIt G8 outperforms our method in the range of FAR between 10^{-1} and 1.

| MATCHER | d-prime | EER | GAR for FAR at 10^{-2} | GAR for FAR at 10^{-3} |
|------------|---------|--------|--------------------------------|--------------------------------|
| Gabor-SWLD | 3.054 | 7.35% | 86.03 % | 80.51% |
| Faceit G8 | 2.097 | 5.19% | 83.82 % | 76.46% |
| Gabor-LBP | 2.719 | 8.09% | 73.53 % | 44.12% |
| LBP | 1.942 | 15.51% | 43.38 % | 18.12% |

Table 1. d-prime, Equal Error Rate (EER), Genuine Accept Rate (GAR) evaluated at FAR 10^{-2} and FAR 10^{-3} for FaceIt G8, our method (Gabor-SWLD), Gabor-LBP, LBP matchers.

Table 1 presents a comparison of the four matching methods based on few operating points. The table shows that the Equal Error Rate (EER) of FaceIt G8 is the lowest, however, our approach provides the best separability index (d-prime) and verification rates (Genuine Accept Rate) at FAR 10^{-2} and at FAR 10^{-3} .

4.4 Processing time

In this section we compare the processing time of the proposed face recognition method against the processing time of FaceIt G8. Broadly speaking, the processing time can be divided into two parts: (1) creation of a probe template from the detected face image which we denote as operation *CT* and (2) matching the probe template against gallery templates which we denote as operation *Match*. In our approach, the operation *CT* is composed of eye detection, normalization of face image into a canonical form using eye positions, selection of appropriate resolution s from a predefined set S , rescaling the face image to resolution s and finally the extraction of features from the face image. For the FaceIt software the corresponding operations are performed by function *CreateTemplate*.

In our approach, the operation *CT* for a probe image of resolution of 64 pixels (the largest resolution supported by the set S) requires only 92.6% of the processing time employed by FaceIt G8 and the *Match* function at such resolution takes 80% of the matching time of FaceIt G8, when performed on the same machine. Furthermore, the processing time of our method reduces dramatically as the resolution decreases. Conversely, the processing time of FaceIt G8 remains constant and is independent of the resolution of probe images. In particular, at the lowest resolution of 32 pixels the operation *CT* for our approach is approximately 4 times faster compared to FaceIt G8, and the *Match* function at such resolution for our method is 5 times faster than FaceIt G8. We summarize the processing time employed by our approach and Faceit G8 in Table 2. The obtained processing times are evaluated on a PC with an Intel Pentium® 4 (2004) single-core CPU at 3.20 GHz and 1GB of RAM.

| Resolution | <i>CT</i> | | <i>Match</i> | |
|------------|---------------------|----------------------|---------------------|----------------------|
| | Faceit G8 (msec) | Gabor-SWLD (msec) | Faceit G8 (msec) | Gabor-SWLD (msec) |
| 64×64 | 730 | 676 | 100 | 80 |
| 48×48 | 730 | 246 | 100 | 46 |
| 40×40 | 730 | 195 | 100 | 29 |
| 32×32 | 730 | 177 | 100 | 19 |

Table 2. The table indicates processing time (in milliseconds) for *CT* and *Match* operations. The computational times for the Faceit G8 and Gabor-SWLD matchers are reported as functions of image size and resolution.

5. CONCLUSIONS

In this paper, we demonstrated the use of a wireless camera network to perform near real-time face recognition that yields high accuracy and robustness. We assembled a portable camera network using off-the-shelf components in which individual cameras perform background subtraction and face detection, and transmit only relevant face images to a fusion center for recognition. Our technique relies on opportunistically acquiring suitable face images as a subject walks through the multi-camera network. By only transferring the face images and further filtering them based on suitability for recognition, we are able to significantly reduce the required network bandwidth. We developed a new robust face recognition algorithm and optimized it to process distributed data in near real time. We evaluated the performance of this system for face recognition using a dataset of 29 subjects collected using the portable camera network. Our networked face recognition system is able to achieve comparable recognition performance to the commercial L1 FaceIt software while reducing the required processing time.

In this paper, we have only considered front face images for recognition. In recent work,⁹ we have described our design of a collaborative technique for extraction of multi-view face images in real-time by exploiting the geometry of the network. Integrating our face recognition software with this real-time face detection service and investigating the ability of our face recognition software to perform human identification based on multi-view face images is a subject of our current research. The design of our own face recognition algorithm offers us the flexibility to tune design parameters which in turn contribute to optimizing the performance (e.g. decreasing the processing time for low resolution images). It also gives us the flexibility to adapt the design for non-frontal face images which is important for our future work.

REFERENCES

- [1] Logitech, “Logitech quickcam pro 9000.” http://www.logitech.com/en-us/webcam_communications/webcams/devices/6333.
- [2] BeagleBoard, “Beagle Board System reference Manual Revision C 3.0 2009.” <http://www.beagleboard.org>.
- [3] Viola, P. and Jones, M., “Rapid Object Detection using a Boosted Cascade of Simple Features,” in [*IEEE Computer Vision and Pattern Recognition (CVPR)*], (2001).
- [4] Bradski, G., “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools* (2000).
- [5] Kulathumani, V., Parupati, S., Ross, A., and Jillela, R., “Collaborative face recognition using a network of embedded cameras,” in [*Distributed Video Sensor Networks*], 373–389, Springer (2011).
- [6] Krahnstoeber, N., T. Yu, S.-N. L., Patwardhan, K., and Tu, P., “Collaborative real-time control of active cameras in large scale surveillance systems,” in [*Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*], (2008).
- [7] Zhou, X., Collins, R., Kanade, T., and Metes, P., “A master-slave system to acquire biometric imagery of humans at distance,” in [*ACM International Workshop on Video Surveillance*], (2003).
- [8] Wheeler, F., Weiss, R., and Tu, P., “Face recognition at a distance system for surveillance applications,” in [*Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*], 1–8 (2010).
- [9] Parupati, S., Bakkannagari, R., Sankar, S., and Kulathumani, V., “Collaborative acquisition of multi-view face images in real-time using a wireless camera network,” in [*Fifth ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*], 1–6 (Aug. 2011).
- [10] Kisku, D., Mehrotra, H., Gupta, P., and Sing, J., “Robust multi-camera view face recognition,” in [*Computing Research Repository*], (2010).
- [11] Graham, D. and Allinson, N., [*Characterizing virtual eigensignatures for general purpose face recognition*], vol. 163 of *Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences*, 446–456, F. Fogelman-Soulie, and T. S. Huang (1998).
- [12] Kokiopoulou, E. and Frossard, P., “Distributed classification of multiple observation sets by consensus,” *IEEE Transactions on Signal Processing* **59**(1), 104–114 (2011).

- [13] Park, U., Jain, A., and Ross, A., “Face Recognition in Video: Adaptive Fusion of Multiple Matchers,” in [*Computer Vision and Pattern Recognition (CVPR)*], (2007).
- [14] Nicolo, F. and Schmid, N., “A Method for Robust Multispectral Face Recognition,” in [*International Conference on Image Analysis and Recognition*], **2**, 180–190 (2011).
- [15] Chen, J., Shan, S., He, C., Zhao, G., Pietikinen, M., Chen, X., and Gao, W., “WLD: A Robust Local Image Descriptor,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(9), 1705–1720 (2009).
- [16] Zhang, W., Shan, S., Gao, W., Chen, X., and Zhang, H., “Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition,” in [*IEEE International Conference on Computer Vision*], **1**, 786–791 (2005).
- [17] Ahonen, T., Hadid, A., and Pietikainen, M., “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(12), 2037–2041 (2006).
- [18] Kittler, J., Hatef, M., Duin, R., and Matas, J., “On Combining Classifiers,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(3), 226–239 (1998).