Sensor-assisted Face Recognition System on Smart Glass via Multi-view Sparse Representation Classification

Weitao Xu[§¶], Yiran Shen*[†], Neil Bergmann[§], Wen Hu[‡‖]
[¶]CSIRO Data61, Australia
[‖]National ICT Australia
[§]School of Information Technology and Electrical Engineering, University of Queensland, Brisbane, Australia
Email: {xuweitao005}@gmail.com {n.bergmann}@itee.uq.edu.au
[†]College of Computer Science and Technology, Harbin Engineering University, Harbin, China
Email: {shenyiran}@hrbeu.edu.cn
[‡]School of Computer Science and Engineering, University of New South Wales, Sydney, Australia
Email: {wenh}@cse.unsw.edu.au

***Abstract*—Face recognition is one of the most popular research problems on various platforms. New research issues arise when it comes to resource constrained devices, such as smart glasses, due to the overwhelming computation and energy requirements of the accurate face recognition methods. In this paper, we propose a robust and efficient sensor-assisted face recognition system on smart glasses by exploring the power of multimodal sensors including the camera and Inertial Measurement Unit (IMU) sensors. The system is based on a novel face recognition algorithm, namely Multi-view Sparse Representation Classification (MVSRC), by exploiting the prolific information among multi-view face images. To improve the efficiency of MVSRC on smart glasses, we propose a novel sampling optimization strategy using the less expensive inertial sensors. Our evaluations on public and private datasets show that the proposed method is up to 10% more accurate than the state-of-the-art multi-view face recognition methods while its computation cost is in the same order as an efficient benchmark method (e.g., Eigenfaces). Finally, extensive real-world experiments show that our proposed system improves recognition accuracy by up to 15% while achieving the same level of system overhead compared to the existing face recognition system (OpenCV algorithms) on smart glasses.**

***Index Terms*—Face Recognition, Smartglass, Sparse Representation, Sampling Optimization, IMU Sensors**

I. Introduction

Smart glasses, e.g., Google Glass and Vuzix Smart Glasses, have attracted significant attention both from researchers and industrial communities since their introduction in 2013. One of the most promising new applications enabled by this technology is to assist people in recognizing identities. Smart glasses have advantages over other smart devices as they are equipped with first-person camera which can be naturally used as a 'third eye' to deliver a significantly better user experience for face recognition.

In this paper, we aim to develop a robust and efficient face recognition system on smart glasses. Face recognition has been well researched in the computer vision community, yet there still remain many challenges on mobile devices. As discussed in [1], most of the advanced face recognition methods fail on portable smart devices because of the tension between high computation requirements and resource constraints. For instance, the battery life of smart glasses is limited by its size. It is reported that the fully charged battery on the Vuzix Smart Glass can last for one hour; however our practical experience shows that the battery would be completely drained within half an hour with display on, camera open and high CPU loading. Moreover, on smart devices, most of the applications involving face recognition are still using the inaccurate but efficient methods in the Open Source Computer Vision (OpenCV) library, e.g, Eigenfaces [2] proposed in 1991. Recently, Shen et al. [1] proposed a new face recognition system: opti-SRC, which is specifically designed for smart phones based on the sparse representation classification (SRC) algorithm [3]. However, as opti-SRC only applies on a single face image, it ignores the rich information enabled by the sensors (accelerometer, gyroscope, magnetometer, etc.) and video camera when used on smart glasses. These additional information may improve the performance of the recognition system and user experience significantly. There have been some recent face recognition systems implemented on smart glasses, e.g., Gabriel [4]. Gabriel shifts the computation burden to a cloudlet (local server) or cloud from the smart glasses while smart glasses are only used for images capture and results display. Gabriel provides assistance services to the users such as face recognition and object recognition. However, the usability of the cloud-based recognition systems relies on the wireless connectivity. The cost of wireless transmission depends greatly on the quality of the wireless connection. Furthermore, according to the results reported in [5], wireless transmission of a bit requires over 1,000 times more energy than a single 32-bit computation. Therefore, the *in-situ* approaches are preferable considering the relatively high cost of wireless transmission and the inconvenience of relying on wireless connections.

To overcome the challenges and facilitate the useful infor-

*Corresponding Author

mation provided by smart glasses, we propose and implement a novel sensor-assisted face recognition system which runs locally on smart glasses by exploiting the information from both the camera and sensors on smart glasses to improve the recognition accuracy and reduce the energy consumption. The system recognizes the identities based on face image sequences collected from different view angles and utilizes the IMU sensors to improve its efficiency. To the best of our knowledge, our work is the first to consider *in-situ* face recognition on smart glasses by fusing IMU sensors. The proposed system presents a humble step forward for *in-situ* face recognition on smart glasses. The contributions of this paper are threefold:

- We propose a novel face recognition algorithm called Multi-view Sparse Representation based Classification (MVSRC). It exploits the high agreement among the sparse representations of the face images from different view angles and applies a novel weighted SRC model to improve the Signal to Noise Ratio (SNR). Our evaluation on several datasets show that MVSRC outperforms several state-of-the-arts multi-view face recognition algorithms.
- We propose a Support Vector Regression (SVR)-based estimation model to relate the recognition accuracy to the angle information obtained by IMU sensors. Then we design a sampling optimization approach: Maximum Accuracy Sampling Optimization (MASO) based on the estimation model to improve the efficiency of MVSRC while preserving its high recognition accuracy. We refer to MVSRC after sampling optimization as fast-MVSRC.
- We implement a face recognition system based on the proposed methods on smart glasses and demonstrate that it significantly outperforms the existing *in-situ* face recognition algorithms on smart glasses. We also discuss the offloading approach and experimentally show that the cost of our system is in the same order of offloading to a nearby server (cloudlet).

The rest of this paper is organized as follows. In Section II, we introduce technical background on SRC [3] and opti-SRC [1]. Section III discusses the related work. We describe the system architecture in details in Section IV. In Section V, we evaluate the performance of the proposed system on several datasets. We then implement the system on smart glasses and conduct real-world experiments to evaluate the system in Section VI. Finally, we discuss the feasibility of the system in Section VII and conclude the paper in Section VIII.

II. Technical Background

In this section, we introduce the rand-SRC face recognition algorithm in [3] and opti-SRC in [1].

In [3], face recognition is cast as a sparse representation problem and is solved by a Sparse Representation Classifier (SRC). SRC is applied to solve the traditional linear equation: $y = Ax$, where $y \in \mathbb{R}^p$ is the test image vector which comes from concatenating the pixel values by rows or columns; $A \in \mathbb{R}^{p \times (N \cdot K)}$ is the dictionary consisting of K classes and each class contains N p-dimensional image vectors. With the knowledge of A and y, ℓ_1 optimization can be applied to solve the linear equation with the *sparse assumption*:

$$\hat{x} = \arg\min_x \|x\|_1 \quad \text{subject to } \|y - Ax\|_2 < \epsilon \qquad (1)$$

where ϵ is used to account for noise and the sparse assumption holds when the test image vector can be represented by one of the classes in A. Due to the large dimensionality of the image vectors, solving Eq. (1) can be computationally intensive. Inspired by the recent information theory technique of Compressive Sensing (CS) [6], [7], [8], a random projection matrix $R \in \mathbb{R}^{m \times p}$ ($m \ll p$) can be applied to improve the efficiency of the ℓ_1 optimization. In particular, the projection matrices are randomly generated from Bernoulli or Gaussian distributions because of their information preserving properties:

$$\hat{x} = \arg\min_x \|x\|_1 \quad \text{subject to } \|Ry - RAx\|_2 < \epsilon \qquad (2)$$

After obtaining the *sparse* representation vector $\hat{x} \in \mathbb{R}^{N \cdot K}$, the class results can be determined by checking the *residuals* based on the Euclidean distance. The definition of the residual for class i is: $r_i(y) = \|y - A\delta_i(\hat{x})\|_2$, where $\delta_i(\hat{x}) \in \mathbb{R}^{N \cdot K}$ contains the coefficients related to class i only (the coefficients related to other classes are set to be zeros). Then the final result of the classification will be: $\hat{i} = \arg\min_{i=1,\ldots,K} r_i(y)$, i.e., the right class produces the minimal residual.

To further improve the performance of SRC, [1] proposes a heuristic algorithm to find the optimal projection matrix instead of the random one. The classification accuracy is improved by up to 12% with the optimized projection matrix.

III. Related work

Face recognition has been well researched in computer vision community. It invokes new research challenges when used on smart devices. With the availability of OpenCV, many apps such as friends tagging have appeared on the app markets. There are three face recognition algorithms in OpenCV: EigenFaces [2], FisherFace [9] and LBPFace [10]. Although these methods can be used in real-time on smartphones, the recognition accuracies are unsatisfactory [1]. SRC [3] outperforms these three methods; however, it cannot provide consistent high recognition accuracy and is computationally intensive. To overcome its limitations, [1] proposed opti-SRC by optimizing the projection matrix to provide consistently better accuracy while solving the computation efficiency issue. Much efforts have also been made on multi-view based face recognition to further improve recognition accuracy [11], [12], [13]. Compared to existing face recognition methods in computer vision community, our work is significantly distinguished by exploiting information from multi-view face images and IMU sensors. The advent of smart glasses makes face recognition easier to perform and more interactive with user by the first-person camera. In [4], a cloud-based system Gabriel was developed on Google Glass to provide cognitive assistance services, such as face recognition, object recognition and optical character recognition (OCR). However, Gabriel could

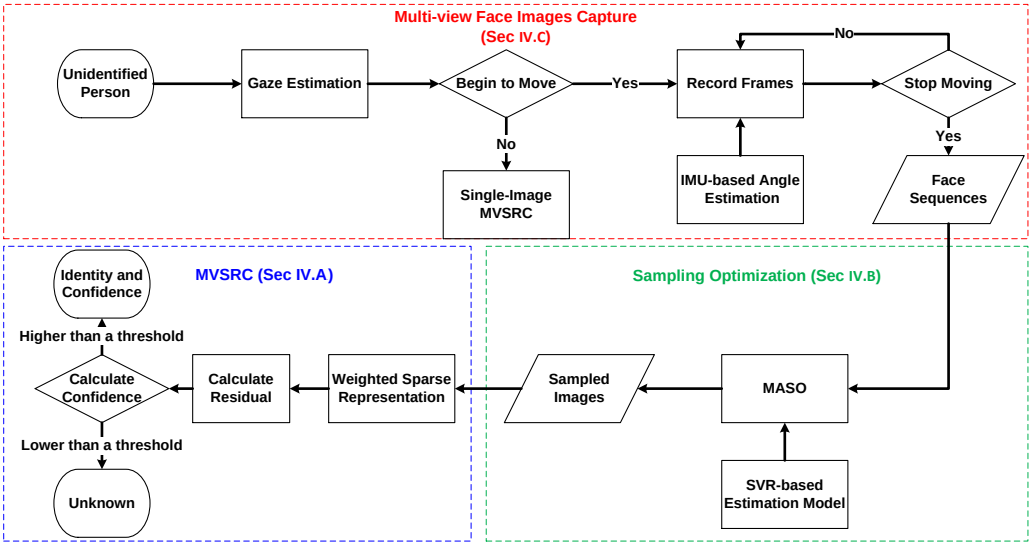


Fig. 1: The pipeline of face recognition system

not work without connecting to servers. In comparison, our system works locally on smart glasses.

There have been several papers which propose sensor-assisted biometric authentication system. [14] developed a multimodal system against spoofing attacks by fusing the information from the camera and fingerprint sensor. [15] proposed a face authentication system to prevent 2D media attack and virtual camera attack by utilizing the motion sensors. In [16], the authors used motion sensors to compensate the tilt of the smartphone for face detection. In difference, the motion sensors are used to assisted in face recognition and improve computation efficiency in the proposed system.

Several papers have exploited the sparsity of multiple measurements to improve the system performance. [17] used CS to compress GPS signals and exploits the information of various propagation paths to improve the SNR of GPS signals. In [18], the authors improved activity classification accuracy by fusing several channel state information (CSI) vectors. The multimodal signal processing on resource constrained platforms is challenging because of the limited computation capability and energy supply. Some efforts have been made to address the problem by applying CS. Bo et al. [19] applied a SRC-based acoustic classification system on Pandaboard and proposed a column reduction procedure to reduce the computational complexity. In [20], a new background subtraction algorithm based on the compressed projections was proposed to enable the real-time tracking with low-power camera nodes.

IV. System Architecture

In this section, we will introduce the proposed system by walking through an example scenario and then describe the architecture in details.

Example Scenario. One day in a party, Tom wants to know the name of the man standing near him. Tom moves a few steps around the man and the smart glass pops up the name of Bob on the display. Then Tom says *hi* to Bob and they have a nice conversation.

System Overview. As shown in Figure 1, the face recognition system starts with acquiring face images when the user starts to move (i.e., walk a few steps around the subject). The angle information of the face images are estimated by the IMU sensors embedded on the smart glasses. Then the images and the associated angle labels are recorded and stored for further processing. After user stops recording face images, the optimal subset of the frames is chosen via the sampling optimization algorithm. Finally, the MVSRC is applied on the samples (i.e. fast-MVSRC) to obtain the classification result and the smart glass prompts the name on the display.

A. MVSRC

Multi-view Sparse Representation based Classification (we call it MVSRC for short) is built on the single image approaches [3], [1]. The key assumption behind MVSRC is that face images obtained from different view angles tend to have a high agreement on the sparse representations because each of the face images from the same person should be linearly represented by the same class in the dictionary. Suppose we have acquired a set of M feasible face images from the camera. Following the single image approach described in Section II, we can obtain a set of estimated coefficients vectors $\hat{X} = \{\hat{x}_1, \hat{x}_2, ..., \hat{x}_M\}$ by solving the ℓ_1 optimization problem for each of the face images. Theoretically, a precise sparse representation will only contain the non-zero entries at the locations related to the specific class. However, noise exists in the empirical estimations. Therefore, the estimated coefficients vector of the test image m can be expressed as:

$$\hat{x}_m = x + \epsilon_m \tag{3}$$

where x is the theoretical sparse representation of the face image vector and ϵ_m is used to account for noise. The image vector could be misclassified due to low Signal to Noise Ratio (SNR). To enhance the SNR of the classification system, we propose a new sparse representation model by exploiting the

information from the multi-view face images. The new sparse representation model can be expressed as:

$$\hat{x}_{sum} = \sum_{m=1}^{M} \alpha_m \hat{x}_m \tag{4}$$

where α_m is the weight assigned to $\hat{x}_m$ based on the *Sparsity Concentration Index* (SCI) defined in [3]:

$$SCI(\hat{x}_m) = \frac{K \cdot max_j \|\delta_j(\hat{x}_m)\|_1 / \|\hat{x}_m\|_1 - 1}{K-1} \in [0,1] \tag{5}$$

the SCI measures how concentrated the coefficients are in the dictionary. $SCI(\hat{x}_m) = 1$, if the test image can be strictly linearly represented using images from only one class; and $SCI(\hat{x}_m) = 0$, if the coefficients are spread evenly over all classes. The weight of $\hat{x}_m$ is obtained by normalizing the SCIs among the multi-view face images:

$$\alpha_m = SCI(\hat{x}_m) / \sum_{n=1}^{M} SCI(\hat{x}_n) \tag{6}$$

In the new face recognition model, the SNR is improved in two aspects: 1) The estimated coefficients vector can be divided into the theoretical part (signal part) and noise part. The theoretical parts among the sparse representations of the multi-view face images from the same identity have a high agreement. However, due to the random nature of the noise, the agreement among the noise signals is low. It is straightforward to prove that the SNR of the face recognition system tends to be improved by summing up the coefficients vectors obtained from conducting sparse representation on different face images. 2) Normalized weights assigned to each of the coefficients vectors are derived from their SCIs. SCI is designed to approximate the sparsity of the coefficient vectors. A higher SCI represents a more accurate approximation achieved by solving ℓ_1 optimization. Therefore, the coefficients vector with higher SNR will be assigned a relatively larger weight. Meanwhile a coefficients vector with a high noise level will be depressed by being assigned a smaller weight.

With the knowledge of $\hat{x}_{sum}$, the compressed residual of each class is computed as:

$$r_i(y_{sum}) = \|R_{opt} y_{sum} - R_{opt} A \delta_i(\hat{x}_{sum})\|_2 \tag{7}$$

where $y_{sum} = \sum_{m=1}^{M} \alpha_m y_m$ is the weighted summation of all the feasible face image vectors obtained by the glasses. Following the same approach in [3], [1], the final classification result is obtained by finding the minimal residual.

To recognize individuals that are not in the system, we adapt the same principle in [1] by using confidence level defined as:

$$confidence = \left(\frac{1}{K} \sum_{i=1}^{K} r_i - \min_{i=1,\ldots,K} r_i \right) / \frac{1}{K} \sum_{i=1}^{K} r_i \tag{8}$$

The confidence level is in the range of $[0,1]$ and should be close to 1 if a subject is known by the recognition system; otherwise it will be close to 0. An appropriate threshold (0.2 in our system) can be chosen by data-driven approach to make the recognition system robust to intruders.

B. Optimized Sampling Strategy

Considering the computation and energy consumption issues of the smart glasses, applying MVSRC straightforwardly on all of the M face images is not a desirable choice because it requires M ℓ_1 optimizations. Evaluation in [1] shows that only a single ℓ_1 optimization takes almost 2/3 of the total computation time. Moreover, a large amount of redundant information exists among the adjacent frames as the face images with similar view angles contain large overlaps. This makes a downsampling strategy possible to improve the efficiency of MVSRC while preserving its accuracy.

To find the best sampling strategy, we propose to optimize the downsampling on the face images set with a predefined energy budget. The energy budget E_b is preset and we aim to find the optimal subset Ω_s of the face images set I to achieve the highest recognition accuracy A_c within the preset budget as follows:

$$\Omega_s = \arg\max_{\Omega} A_c \quad \text{s.t. } E_{total} \leq E_b, \Omega \subseteq I \tag{9}$$

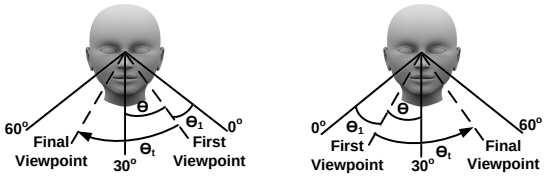
where Ω is one of the arbitrary subsets of I and E_{total} is the total energy consumption for face recognition.

To solve the optimization problem, we start with analyzing the parameters affecting the face recognition accuracy. According to the processes of face recognition with smart glasses, we define a potential parameters list X and aim to relate this list to recognition accuracy by machine learning. The parameters included in X in our system must satisfy two conditions: 1) it can be quantified and 2) it can be estimated by sensors on smart glasses. Using these two conditions, we build the list $X = (\theta_1, \theta_t, \theta_{s1}, \theta_{si}, n_s)$ consisting of the following parameter variables:

- θ_1: the view angle of the first recorded face image which is estimated by image processing method.
- θ_t: the total rotation angle displacement between the leftmost (rightmost) and rightmost (leftmost) face images in the yaw direction and is estimated by the IMU sensors.
- θ_{s1}: the view angle of the first face image in the chosen subset and is estimated by combining the result of θ_1 and analysis on IMU sensor readings.
- θ_{si}: the view angle interval among the face images in the chosen subset and is estimated by IMU sensors.
- n_s: the number of face images in the chosen subset.

The illustrative explanation of the parameter variables is shown in Figure 2 (θ is obtained by gaze estimation in Section IV-C1). As the evaluation in Section V-B1, the feasible range of the view angle is between 30° to the left and 30° to the right of the frontal face respectively. The results are also consistent to the symmetric property of the human face. Therefore the original angle (0°) can be either the 30° view angle to the right (as shown in Figure 2(a)) or the 30° view angle to the left (as shown in Figure 2(b)). We choose the origin at the same side as the view angle of the first recorded face image.

As the exact recognition accuracy cannot be computed without the the knowledge of the groundtruth in real world



(a) θ_1 at the left profile face (b) θ_1 at the right profile face

Fig. 2: Angle coordinate system settings: θ is the relative view angle of the first face image to the frontal face, θ_1 is the view angle of the first face image to the origin ($\theta_1 = 30° - \theta$), θ_t is the rotation angle displacement between the first and last face images.

application, to estimate recognition accuracy, we model the correlation between the parameter variables and the recognition accuracy based on a novel Support Vector Regression (SVR)-based approach [21] to find the optimal observation of the parameters. The estimation model is learned offline and then used for *in-situ* accuracy estimation. We use our private dataset (see Section V for details of private dataset) which consists of 10 subjects to learn the estimation model. Each of the subject contributes 9 image sequences and each of the image sequences contains 61 face images from different view angles. In the following parts, we will describe how to build the estimation model.

We define the set of all the possible observations of X' as $\{\chi_1, \chi_2, \chi_3, ..., \chi_N\}$ and the corresponding accuracies as $\{z_1, z_2, z_3, ..., z_N\}$. Each of the observations is related to a certain subset of the face images which is determined by the values of the parameters in X'. With the information of the observations and the corresponding accuracies, we aim to find the function $f(\cdot)$ which best approximates the relation inherited between the input features X' and it can be used later on to infer the accuracy z for a new input feature X'. Specifically, the goal of regression is to find the function $f(\cdot)$ which relates the parameters list X' to the recognition accuracy z with the deviation of at most ϵ:

$$Dev(z, f(X')) \leq \epsilon \tag{10}$$

where $Dev(\cdot, \cdot)$ represents the deviation computation.We apply SVR [21] by using all the possible observations in private dataset to find the function $f(X')$ and we use the Radial Base Function (RBF) Kernel which is defined as:

$$k(x_i, x) = e^{-\gamma \|x_i - x\|^2} \tag{11}$$

where γ is a kernel parameter (0.01 in our experiment). For more details of SVR, readers are encouraged to refer [21] for the step-by-step instructions.

With the knowledge of the estimation function, we propose a computationally efficient approach to solve the optimization problem Eq. (9), i.e., Maximum Accuracy Sampling Optimization (MASO). In the real application, θ_1 and θ_t are user-specific and determined before the sampling optimization stage. The optimization approach is actually searching for the optimal observation of $(\theta_{s1}, \theta_{si}, n_s)$ under the predefined conditions (energy budget). The estimation function is used to efficiently approximate the recognition accuracy with the knowledge of the angle information (Line 7 in algorithm 1).

Algorithm 1 Maximum Accuracy Sampling Optimization

1: Input: Estimation model f, total energy consumption E_{total}, energy budget E_b, angle of the first view θ_1, total rotation angle displacement θ_t, angle of the first sampled image θ_{s1}, interval between sampled images θ_{si}, number of sampled images n_s, maximum number of sampled images $N_{max} = [(min(\theta_1 + \theta_t, 60) - \theta_{s1})/\theta_{si}]$.
2: Initialization: allocate one empty list: Y, $m = 0$.
3: **for** $n_s = 1 : N_{max}$ **do**
4: **for** $\theta_{si} = 0 : \theta_t$ **do**
5: **for** $\theta_{s1} = \theta_1 : min(\theta_1 + \theta_t, 60)$ **do**
6: **if** $(E_{total} \leq E_b)$ **then**
7: $Y_m = f(\theta_1, \theta_t, \theta_{s1}, \theta_{si}, n_s)$
8: m++
9: **end if**
10: **end for**
11: **end for**
12: **end for**
13: $(\theta_{s1}, \theta_{si}, n_s) = \underset{\theta_{s1}, \theta_{si}, n_s}{\operatorname{argmax}} Y$
14: Output: $(\theta_{s1}, \theta_{si}, n_s)$

In Algorithm 1, the total energy consumption of the system can be expressed as:

$$E_{total} = T * (P_{base} + P_{dis} + P_{imu} + P_{cam}) + E_{cpu} \tag{12}$$

where T is the total operating time for classification; P_{base} denotes the baseline power consumption of the smart glass; P_{dis}, P_{imu} and P_{cam} are the power consumed by the display, IMU sensors, and camera respectively; E_{cpu} is the total energy consumption of CPU for classification which accounts for face detection, gaze estimation, ℓ_1 optimization, residual calculations and sampling optimization. E_{cpu} can be further split into repeatable part and once-off part. The once-off part consists of the energy consumption of gaze estimation, residual calculation and sampling optimization which are operated once only during classification, while the repeatable part includes face detection and ℓ_1 optimization which are operated on every sampled face images. Assuming m face images are sampled for classification, the total energy consumption can be further expressed as:

$$E_{total} = T * (P_{base} + P_{dis} + P_{imu} + P_{cam}) + m * E_u + E_1 \tag{13}$$

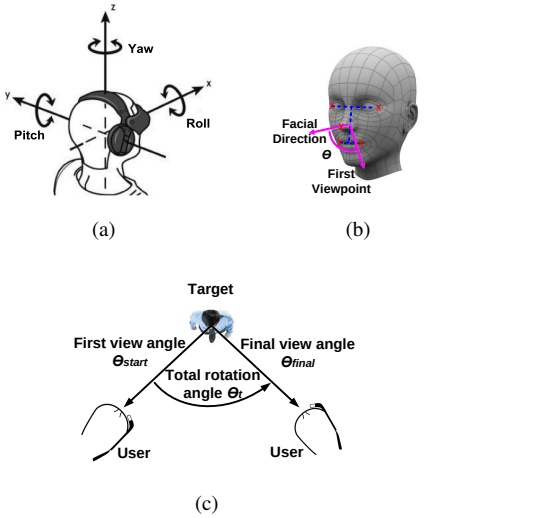
where E_u is unit energy consumption of the repeatable part on each image, E_1 is the energy consumption of once-off part.

The optimization strategy and corresponding parameter (i.e. E_b in MASO) is preset by user before recognition, and the optimization process is called online after multi-view face images are obtained. As the results in Section V-D1, the sampling optimization component only takes less than 2.7% of the total

computation time which suggests that our optimization method is computationally efficient. In order to differentiate from MVSRC, we call MVSRC after sampling optimization as fast-MVSRC. Note that the proposed MASO can also be generally applied to other multi-view face recognition methods.

C. Sensors-assisted System

As described above, the sampling optimization process requires the angle information. The angle information is obtained by gaze estimation of the first face image and angle displacement estimation with IMU sensor readings.



(a) (b) (c)

Fig. 3: (a) Head model of the user [22] (b) Gaze of the subject (c) Bird view of the recognition process.

1) Gaze Estimation: Gaze estimation is used to find the initial angle information θ_1 of the first image by the image processing technique proposed in [23]. The method uses the locations of the following five facial features: left and right eye far corners, left and right mouth corners and nose tip which are marked as red crosses in Figure 3(b). The angle θ between the view point of the first face image recorded and the frontal view is calculated by analyzing the relative positions of the five facial points. Then θ_1 in the view angle coordinate system can be obtained with the knowledge of θ ($\theta_1 = 30^\circ - \theta$ in our system). After obtaining the initial angle information, the view angles of the face images recorded later can be calculated by accumulating the angle displacements along with θ_1 as reference.

2) Angle Displacement Estimation: From Figure 3(a) and Figure 3(c), we notice that the rotation angle is actually the angle change along yaw direction of the smart glass when the user moves around the subject. In practice, substantial pitch and roll rotations rarely occur. Moreover, the slight pitch rotation caused by the height difference between the subject and user is within the tolerance of the face recognition algorithms. One can estimate rotation angle by simply integrating gyroscope readings. However, the measurements from IMU sensors suffer from bias, noise and systematic errors (e.g., misalignment between the sensor axes and non-unit scale parameters) which lead to inaccurate orientation estimations [24]. To address this issue, we implement a sensor fusion algorithm to compensate for the weakness of each sensor by utilizing other sensors' information. Here we use quaternion-based Extended Kalman Filter (EKF) proposed in [25] to estimate the orientation of the smart glass. The EKF incorporates an in-line calibration procedure for modeling time-varying biases which may affect sensors like accelerometers and magnetometers, and a mechanism for adapting their measurement noise covariance matrix in the presence of motion and magnetic disturbances. Assume the output of EKF is quaternion $q = [w, x, y, z]^T$, we could compute the three Euler angles of head model in Figure 3(a) using the following equations:

$$\begin{bmatrix} \varphi \\ \psi \\ \theta \end{bmatrix} = \begin{bmatrix} atan2\left(2\left(wz + xy\right), 1 - 2\left(x^2 + z^2\right)\right) \\ asin\left(2\left(wx - yz\right)\right) \\ atan2\left(2\left(wy + xz\right), 1 - 2\left(x^2 + y^2\right)\right) \end{bmatrix} \quad (14)$$

where φ stands for roll, ψ represents pitch and θ represents yaw rotations respectively.

To improve user experience, the IMU sensor readings are used to automatically determine the start and end of recognition process. From our observation, the gyroscope data along the yaw direction (perpendicular to the motion) exhibits large variations when the user moves during the recognition process. We first apply a low pass filter to filter out the small vibrations, then our system will start to record face images at θ_{start} when the gyroscope sensor reading along the yaw direction is larger than a threshold (0.15 rad/s in our system) and end the recording at θ_{final} when it is lower than the threshold in the sense that the user stops moving. The rotation angle can be simply obtained by $\theta_t = |\theta_{final} - \theta_{start}|$ as shown in Figure 3(c).

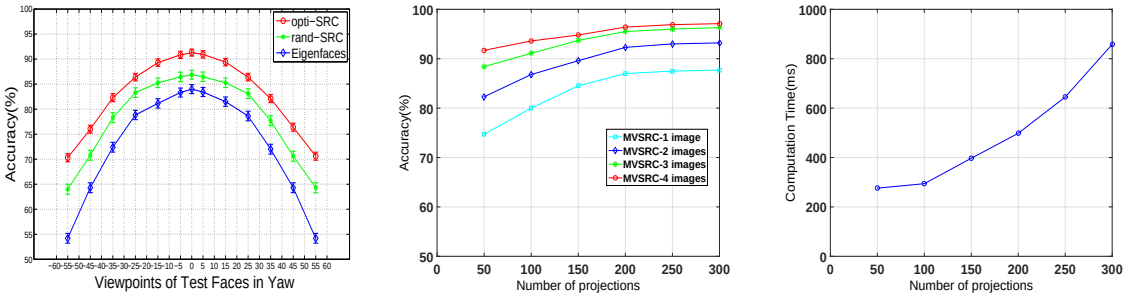
Another challenge is that the timestamps of sensors and video frames are usually not well synchronized [26]. Therefore, we apply the online calibration and synchronization method proposed in [26] to obtain the delay t_d between IMU sensors and camera, then t_d is in return used to synchronize the timestamps of sensor readings and images. For a full description of the EKF-based orientation estimation and synchronization, the reader is referred to [25] and [26] respectively.

After the user stops moving, θ_1, θ_t and face images associated with corresponding angle displacement are used in MASO as described in Algorithm 1.

V. Evaluation

A. Goals, Metrics and Methodology

In this section, we evaluate the performance of the proposed system via simulation. The goals of the simulation are fourfold: 1) to determine the choice of the key parameters including feasible range of views in the yaw direction and the number of projections used in MVSRC; 2) whether MVSRC outperforms the state-of-the-art face recognition methods in accuracy; 3) whether MASO improves the efficiency of



(a) (b) (c)

Fig. 5: Experimental results of parameters choice

MVSRC while retaining high accuracy; and 4) to evaluate the angle estimation accuracy of IMU-based method and its impact on face recognition accuracy.

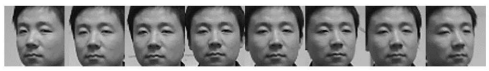


Fig. 4: Examples of face images from private dataset

The evaluations are based on two datasets: Honda/UCSD video dataset (Honda/USCD) [27] and the private dataset we have collected with the smart glasses[1]. Honda/USCD video dataset is widely used for the evaluation of multi-view face recognition methods. It consists of 59 image sequences from 20 subjects recorded in different environments and each subject contributes at least two sequences. The number of frames of the sequences vary from 12 to 645. The angle information is not available in Honda/UCSD dataset, therefore we built our private dataset by obtaining both multi-view face images and their associated view angles. Our private dataset consists of 10 subjects (2 females and 8 males) aged from 24 to 43 with different skin tones. The face images are taken under 9 different categories by combining the different expressions (neutral, happy and sad) and locations (corridor, office and outdoor). The user wearing the smart glass records the video clips of the candidate to be recognized (suppose the candidate is just facing to the user) by moving around the subject from left to right (in yaw direction) with wide range. The flow of orientation information is obtained and synchronized with the video clips. Face regions are detected by a Viola-Jones face detector [28] and cropped to 48×48 gray-scale images. We then apply the method introduced in Section IV-C to find the frame containing the face in frontal view angle. Finally we sub-sample the video clips by every 1° according to the associated angle displacement information until we reach 60° to both left and right direction. Therefore, for each video clip, we obtain a symmetric sequence of 121 face images with view angles from -60° (left) to $+60^\circ$ (right). A sequence of sample images is shown in Figure 4.

We determine the parameters (feasible angle range and number of projections) by gradually changing the parameters and the choices are made according to the evaluation results on the real datasets. In the evaluation of this section, the training set is derived from random selection as in [1]. We show the results of the average value and 95% confidence level of the performance metrics (accuracy, energy consumption and computation time) over 30 independent trails. The computation time and energy consumption are measured by running the system on Vuzix M100 Smart Glass.

[1]Ethical approval for carrying out this experiment has been granted by the corresponding organization (Approval Number 2014000589)

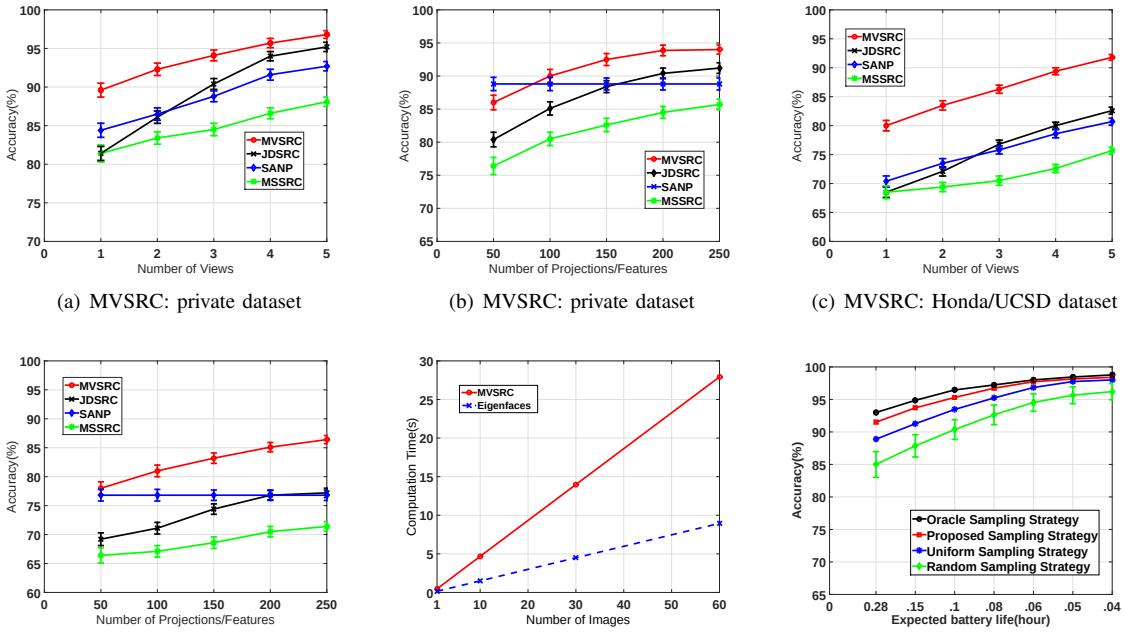
B. Parameters Choice

In this section, we will determine the choice of the parameters including the feasible angle range and the number of projections applied in the system by evaluations on the real datasets.

1) Impact of View Angles: The view angles have substantial impact on the recognition accuracy. In this experiment, we evaluate the influence of different view angles on the recognition accuracy on private dataset.

As described previously, we obtain a symmetric sequence of 121 face images with view angles from -60° (left) to $+60^\circ$ (right) for each video clip. We group the the face images by the view angles uniformly into 12 bins by every 10° and the frontal faces are picked up to form the 13th bin. Each bin represents an evaluation point. We calculate the recognition accuracy of each bin by using three single-image based face recognition methods: opti-SRC, rand-SRC and Eignfaces. We display the evaluation points at the medium degree of each bin (x-axis) in Figure 5(a). From the results, we observe that the recognition accuracy decreases when the view angles of the face images deviate from the frontal view angle and the recognition accuracy has dropped significantly when the view angle is over 30° apart from the frontal view. Therefore, we determine the feasible range of the view angles in our system as $[-30^\circ, 30^\circ]$ (the origin of the view angle is the frontal face). In addition, the work in [29], [30] also studies the effects of different poses on face recognition and their findings support our results. With this observation, we remove the images in our private dataset whose view angles are not in the feasible range. Therefore, the number of image in each sequence becomes to 61.

2) Impact of Number of Projections: It is known that the recognition accuracy can be improved by increasing the



(a) MVSRC: private dataset (b) MVSRC: private dataset (c) MVSRC: Honda/UCSD dataset

(d) MVSRC: Honda/UCSD dataset (e) MVSRC: Computation time (f) MASO: Private Dataset

Fig. 6: Evaluation results

number of projections or features. However, it also increases the computation cost significantly. To investigate the recognition accuracy on the number of projections, we evaluate the performance of MVSRC with different settings by varying the number of projections from 50 to 300. As MVSRC uses multiple face images to perform recognition, we calculate the accuracy of MVSRC with different number of *views* (the number of face images from different view angles for each classification) $n_{view} = 1, 2, 3, 4$. We group the face images of each image sequence in the test set into small subsets of n_{view} images and report the recognition accuracy of MVSRC on the small subsets of different sizes in Figure 5(b). We also evaluate the computation time of MVSRC with different number of projections. As the computation time of MVSRC is proportional to n_{view}, without loss of generality, we present the computation time of MVSRC when $n_{view} = 1$. From the results shown in Figure 5(b) and Figure 5(c), we find the growth of the recognition accuracy diminishes when the number of projections is above 200 while computation time keeps increasing substantially. Therefore, we choose the number of projections as 200.

C. Dataset Evaluation of MVSRC

In this section, we compare MVSRC with several competing face recognition methods in the literature. Note that we do not consider angle displacement information in this section.

1) Comparison with State-of-the-Art: We compare MVSRC with three state-of-the-art multi-view face recognition methods, namely, JDSRC [11], SANP [12] and MSSRC [13]. We compute the recognition accuracy of different methods under different number of views (k) as well as different number of projections/features (d) on private dataset and Honda/UCSD dataset respectively. For each dataset, we randomly choose 30 images from each subject to form the training set and the rest are used as test set. We first evaluate the accuracy with different number of views from 1 to 5 by setting $d = 200$. We then evalulate the accuracy of different methods against the number of projections/features from 50 to 200 with $k = 3$. Figure 6(a) and Figure 6(b) plot the results of private dataset. The results on Honda/UCSD dataset are shown in Figure 6(c) and Figure 6(d). Note that SANP is not a feature-based method, the accuracy of SANP in Figure 6(b) and Figure 6(d) is shown by a straight line. From the results, we can see that MVSRC consistently achieves the best recognition accuracy and is up to 7% and 10% more accurate compared to the second best recognition method on the two datasets respectively. MVSRC, JDSRC and MSSRC are based on original SRC; however, we noticed that MVSRC performs better than JDSRC and SANP when k=1. This is due to the fact that single-image MVSRC becomes opti-SRC and opti-SRC performs better than SRC. It is worth mentioning that MVSRC is approximately $5\% - 10\%$ more accuracy than direct major voting in our evaluation. Due to limited space, we do not plot the results of direct major voting in this paper.

2) Computation Time Evaluation: Eigenfaces is known to be efficient and is the most popular method used on resource-constrained devices. We use our private dataset to evaluate the computation time for MVSRC and Eigenfaces (with majority voting) on smart glasses with various sizes of image sequence from 1 to 60. The cost of the two methods is represented by the computation time used for one classification operation. The results in Figure 6(e) demonstrate that MVSRC requires significantly more computation time than Eigenfaces and the

gap increases with the growth of the number of images used for recognition. However, we will show in the following section that the computation time of MVSRC can be reduced significantly while preserving high accuracy with the proposed sampling optimization method.

D. Dataset Evaluation of MASO

To address the computation issue of MVSRC, we propose fast-MVSRC by combining MVSRC with MASO described in Section IV-B. In this section, we start with some preliminary experiments to investigate the computation and energy cost performance of our system. Then we compare MASO with other common sampling strategies. Finally, we compare the recognition accuracy of fast-MVSRC with Eigenfaces under various computation cost on smart glasses.

1) Preliminary Experiments: As the optimization method proposed in Section IV-B requires the energy consumption information, we conduct preliminary experiments on the Vuzix Smart Glass to obtain the energy consumption and computation time information. It is worth mentioning that the proposed system is not platform specific and is compatible with Google Glass as well.

In the preliminary experiments, we first evaluate the impact of image resolution and frame per second (FPS) on system cost. Table I shows that the cost of face detection improves significantly with the increase of image resolutions. Image downsampling reduces the system cost, however, it also leads to low recognition accuracy. Note that the recognition accuracy shown in Table I are the mean results of single-image MVSRC on private dataset without sampling optimization. The original image resolution of Vuzix Smart Glass is 432×240. As shown in Table I, we found that the recognition accuracy drops significantly when the image is downsampled to 108x60 (4 times downsampling in both dimensions). Thus the raw image is downsampled to 216×120 (1/2 downsample) in the prototype. Table II illustrates the display power, camera power and mean angle estimation error under different FPS. To balance the system cost and the accuracy of angle estimation, we set FPS to 24 in the prototype system.

TABLE I: System cost of face detection operation under different image resolutions

Resolution	Time(ms)	Energy(mJ)	Accuracy(%)
432*240	175	107	88.1
216*120	85	56	86.8
108*60	63	34	78.4

TABLE II: Display power and camera power under different FPS

FPS	P_{dis}(mW)	P_{cam}(mW)	Estimation error
27	265	177	1.9°
24	220	122	1.9°
20	204	112	2.9°
15	187	105	3.8°
10	162	92	5.7°

TABLE III: Resource consumption on Vuzix Smart Glass

Face Detection(ms)	85			
Gaze Estimation(ms)	87			
ℓ_1(ms)	350			
Residual(ms)	33			
MASO (ms)	15			
E_u(mJ)	238			
E_1(mJ)	62			
Baseline Power(mW)	35			
Display Power(mW)	220			
Camera Power(mW)	122			
Sampling Rate	NORMAL	UI	GAME	FASTEST
Frequency(hz)	5	15	50	205
IMU Power(mW)	11	29	61	295

After image resolution and FPS are determined, we evaluate the resource consumption of each component in the system. Table III shows the related specifications and resource consumptions on Vuzix Smart Glass. The computation time is obtained from the console of the Eclipse development environment and the energy consumption of each component is estimated by PowerTutor App (it was also used in [1]). The sampling rate of the IMU sensors can be set via Android API and is in four levels from low to high: NORMAL, UI, GAME and FASTEST. Considering both the energy consumption and the accuracy of the synchronization, we choose the sampling rate of the IMU sensors as GAME.

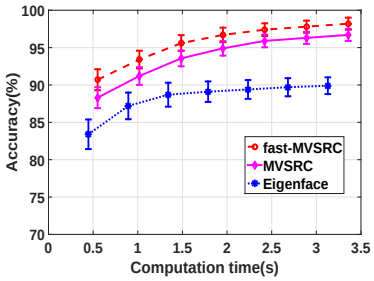
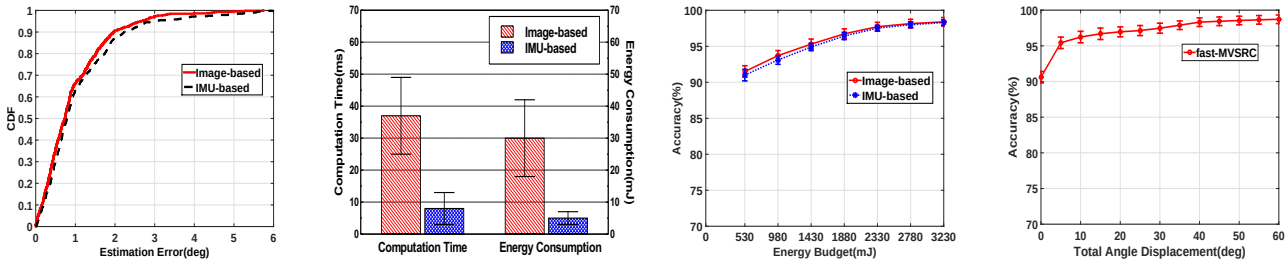


Fig. 7: Accuracy-Efficiency Curve

2) Comparison with Other Sampling Strategies: In this section, we compare the recognition accuracy of fast-MVSRC with different sampling strategies under different energy consumption budgets. The sampling strategies include the proposed algorithm MASO, random sampling strategy, uniform sampling strategy and oracle sampling strategy. For the random sampling strategy, we randomly choose a subset of the image sequences. The energy consumption of MVSRC with this subset should satisfy the budget. For the uniform sampling strategy, we divide the image sequence into uniform groups and select the face image in the middle of the group as the representative. We vary the energy budget from 530mJ to 3230mJ for each classification. We consider an offline oracle optimal strategy that provides an upper bound on recognition accuracy for a given energy budget. In terms of oracle sampling strategy, we calculate the recognition accuracy of MVSRC with all possible subsets to find the most accurate one. However it is not applicable for real-world applications as the recognition system cannot compute the recognition



(a) (b) (c) (d)

Fig. 8: Evaluation results of IMU-based gaze estimation: (a) CDF of estimation error. (b) Computation time and energy consumption on one image. (c) Impact of estimation error on recognition accuracy. (d) Accuracy of different total angle displacement.

accuracy for each of the possible subsets without the the knowledge of the groundtruth (the identity of each face image obtained manually). For illustration purpose, we convert the energy budget values (mJ) into expected battery life (hours) and plot the results in Figure 6(f). The expected battery life means the duration of running the system in Vuzix M100 Smart Glass.We can see that fast-MVSRC is comparable to the oracle sampling strategy, and achieves higher accuracy than the random and uniform approaches with the same energy budget.

3) Fast-MVSRC v.s. Eigenfaces: To demonstrate the effectiveness of fast-MVSRC, we compare the *Efficiency-Accuracy* performance of fast-MVSRC, MVSRC and Eigenfaces. We define the *Efficiency-Accuracy* performance as the recognition accuracy with respect to the computation time. We calculate the accuracy of fast-MVSRC and Eigenfaces with majority voting data fusion for multiple images under different computation time. The computation time is varied by using different number of face images for classification. From the results shown in Figure 7, we can see the recognition accuracy of fast-MVSRC is up to 9% better than Eigenfaces. The results in Figure 6(e) and Figure 7 show that fast-MVSRC improves the efficiency of MVSRC significantly while preserving high recognition accuracy.

E. Evaluation of IMU-based Gaze Estimation

In this section, we evaluate the performance of IMU-based gaze estimation method and the impact of estimation error on the face recognition accuracy. We also evaluate the impact of total angle displacement on face recognition accuracy.

1) Comparison with Image-based Gaze Estimation: In this part, we compare the estimation accuracy and resource consumption of IMU-based gaze estimation used in our system and image-based gaze estimation proposed in [23]. The results in [23] show that the image-based method can achieve a mean estimation error of 2.5° and a maximum estimation error of 6° in 1000 samples of noisy face images. We randomly select 30 face images from each subject to form the comparison image set and use the angle information obtained in Section V-B1 as corresponding estimated value of IMU-based method. For image-based gaze estimation, we perform facial features detection first and use the method in [23] to estimate the gaze. Facial features are detected by a state-of-the-art facial landmark detector *flandmark* [31]. The ground truth are obtained by annotating facial features manually and then performing the method in [23]. From Figure 8(a) and Figure 8, we can see that our method reduces computation time by 65% and energy consumption by 78% respectively, while it achieves comparable accuracy compared to image-based gaze estimation method.

2) Impact of Angle Estimation Error: As shown in Figure 8(a), the estimation errors for most of the face images (over 95%) are within 3°. Therefore, it is important to know the impact of the estimation errors on the recognition accuracy. As described in Section V-A, each subject in the private dataset has 9 image sequences collected in different categorizes. We randomly select 5 image sequences from each subject to form a training dataset and use the rest sequences as testing data. We apply MASO on each testing sequence and obtain the angles of the sampled images. Then we select corresponding images in the test sequence according to their angle information. The angles of the testing images are obtained from two methods, i.e., IMU-based method and image-based method. We vary the energy budget from 530mJ to 3230mJ and calculate corresponding recognition accuracy of IMU-based method and image-based method respectively. From Figure 8(c), we can see that IMU-based method achieves comparable accuracy to image-based method. Therefore, we conclude that the minor errors introduced by the IMU based angle estimation will not have noticeable influence on the recognition accuracy.

3) Impact of Total Angle Displacement: As the total angle displacement between the starting view angle and the ending view angle varies in the practical use, we evaluate the impact of the total angle displacement on the recognition accuracy. We gradually increase the total angle displacement from 0° to 60° by every 5° and the recognition accuracy is calculated. Figure 8(d) shows that the total angle displacement for high recognition accuracy (95%) can be as small as 5° which enables a very short image collection phase (200ms as our user study). In addition, Figure 8(d) also shows that there is a significant improvement on recognition accuracy from 0° (static) to 5°.

VI. Real-world Experiments

A. System Implementation

The prototype of our proposed face recognition system is implemented on Vuzix M100 smart glass. The CPU is an OMAP4460 at 1.2GHz and the operating system is Android 4.0.4. It is equipped with a 5-megapixel camera and the images captured in our system are 216×120@24fps. We use hardware face detection of OMAP for efficient operation and facial features (i.e. nose tip, eye outer corners and mouth corners) are detected by *flandmark* [31]. The efficient implement of ℓ_1 optimization algorithm ℓ_1-*Homotopy* [32] is used as [1], the reader can refer to [1] for complexity analysis of ℓ_1-*Homotopy*.

B. Experimental Description

We recruited 15 volunteers: 5 users and 10 subjects in the training set. The 10 subjects are the same as our private dataset which was collected under different environments. We select 30 face images from each subject to form the training set. Therefore, the training dictionary is a matrix of size 2304×300 (face image is resized to $48\times48=2304$). The experiments are conducted at two different locations: the office and outdoor. The lighting conditions are quite different for indoor (200-400 lux) and outdoor (over 1,000 lux) environments. Different lighting conditions are applied for indoor experiment (front-lighting, back-lighting and uniform lighting) and outdoor experiment (front-lighting and backlighting). Thus, the experiments are divided into 5 categories. For each category, users conducted two recognition attempts for each of the subjects. Therefore, we obtain 500 independent recognition results. The energy budget in the system is set as 550mJ (the actual energy consumption of our system was around 520mJ).

We also implement the OpenCV face recognition algorithms (OpenCV-2.4.9) on Vuzix M100 smart glass as a benchmark. OpenCV provides three face recognition methods, namely EigenFaces [2], FisherFace [9] and LBPFace [10] in its library. In the experiments, we found that these three methods achieve comparable performance in terms of recognition accuracy and computational cost. Due to limited space, we present the results of Eigenface only because the state-of-the-art cognitive assistance system Gabriel [4] uses Eigenfaces as face recognition methods. However, Gabriel uses offloading approach which is a complementary work to the proposed system.

C. Experimental Results

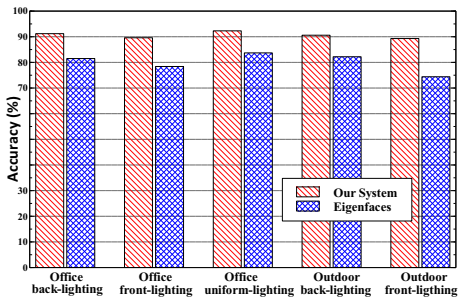


Fig. 9: Recognition accuracy

The recognition accuracy of our system and Eigenfaces in different experimental categories are shown in Figure 9. The proposed system is very stable under different lighting conditions. It outperforms Eigenfaces significantly in every experimental category and is up to 15% more accurate than Eigenfaces under outdoor front-lighting condition. We also evaluate and compare the system overhead of our system and Eigenfaces. From the results in Table IV, we can see that the cost of the proposed system is in the same order of that of Eigenfaces.

TABLE IV: System overhead

Statistic	The Proposed System	Eigenfaces
Computation Time	516-582ms	247-331ms
Energy Consumption	506-535mJ	316-410mJ
Expected Battery Life	≈0.28hr	≈0.37hr
Memory Usage	55-64MB	38-44 MB

VII. Discussion

Feasibility The implementation of the system takes advantage of the following assumption: the subject to be recognized remains still and the user needs to move subtly to the left (right) of the target to capture multi-view images. Such assumption may cause inconvenience in practical scenarios. However, as the evaluation results in Section V-E3, a total angle displacement of 5° is sufficient to obtain a reliable recognition result (over 95%) and it only takes approximately 200ms for image collection. We believe it only requires small efforts of the user and subjects for normal cases. If the user remains static, single-image MVSRC will be adopted. However, if the user is willing to make extra small efforts with sacrificing user experience on one hand, the significant higher recognition accuracy will significantly improve user experience on the other hand. The proposed system provides such options to users. In practice, face recognition may be applied in a more complex scenario, such as the user or subject is sitting. We defer face recognition in these scenarios as our future work.

Offload V.S. In-situ Offloading computationally intensive operations from mobile devices to powerful infrastructure is a common strategy to reduce computation burden on resource constrained devices. In terms of offloading approach, the smart glass is used to capture images and perform MASO, then the sampled images are transmitted to server via wireless network. Results obtained by running MVSRC on server are sent back to the smart glass. We evaluate the response time and energy consumption of smart glass by transmitting raw sampled images under two different offloading approaches: cloudlet and remote cloud. Hardware specifications of different offloading strategies are shown in Table V. The cloudlet is implemented inside Virtual Machine (VM) managed by Vmware Workstation on Windows 7 host. We use Amazon EC2 VM instances located in local magic land as remote cloud (location is anonymized for blind review process). The wireless network is based on a campus WiFi.

Figure 10 presents the response time and energy consumption of of the smart glasses using different approaches. We can

TABLE V: Server hardware specifications

Offload Strategy	CPU	RAM
Cloudlet	Intel Core 2.7Ghz 2cores	8GB
Cloud	Intel Xeon 2.5Ghz 1VCPU	1GB

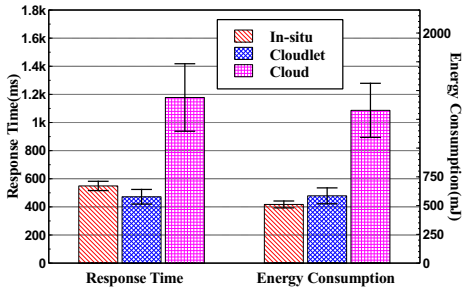


Fig. 10: Comparison of different strategies

see a significant drop in both latency and energy consumption when switching from cloud to cloudlet. The performance of offloading to remote cloud depends greatly on the network conditions. We also find that the cost of our proposed system (*in-situ*) is comparable to the offloading approach with cloudlet and is significantly lower than that with remote cloud. It is worth mentioning that the energy consumption of the proposed system largely depends on the user-specified parameter settings of optimization strategy, i.e., E_b in MASO. Meanwhile, we also note that more advanced recognition methods such as 3D techniques can be achieved in powerful server. However, offloading approaches require extra infrastructure and system cost. Furthermore, the proposed system has advantages over offloading strategies when network is not available or in poor quality.

VIII. Conclusion

In this paper, we explore the capability of smart glasses and propose a novel face recognition system which utilizes the power of multimodal sensors. The proposed system improves recognition accuracy by combining multi-view face images and exploits prolific information from IMU sensors to reduce energy consumption. Extensive dataset based evaluations and real-world experiments demonstrate that our system is both accurate and efficient compared to the state-of-the-arts.

References

[1] Y. Shen, W. Hu, M. Yang, B. Wei, S. Lucey, and C. T. Chou, "Face recognition on smartphones via optimised sparse representation classification," in *IPSN '14*. IEEE Press, 2014, pp. 237–248.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, pp. 71–86, 1991.

[3] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *PAMI*, pp. 210–227, 2009.

[4] K. Ha, Z. Chen, W. Hu, W. Richter, P. Pillai, and M. Satyanarayanan, "Towards wearable cognitive assistance," in *Mobisys '14*. ACM, 2014, pp. 68–81.

[5] K. C. Barr and K. Asanović, "Energy-aware lossless data compression," *TOCS*, pp. 250–291, 2006.

[6] R. Baraniuk, M. Davenport, R. DeVore, and W. M, "A Simple Proof of the Restricted Isometry Property for Random Matrices," *Constr Approx*, pp. 253–263, 2008.

[7] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, pp. 489–509, 2006.

[8] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, pp. 1289–1306, 2006.

[9] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *PAMI*, pp. 711–720, 1997.

[10] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *PAMI*, pp. 2037–2041, 2006.

[11] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Joint dynamic sparse representation for multi-view face recognition," *Pattern Recognition*, vol. 45, no. 4, pp. 1290–1298, 2012.

[12] Y. Hu, A. S. Mian, and R. Owens, "Sparse approximated nearest points for image set classification," in *CVPR*. IEEE, 2011, pp. 121–128.

[13] A. W. Enrique G. Ortiz and M. Shah, "Face recognition in movie trailers via mean sequence sparse representation-based classification," in *CVPR*. IEEE, 2013, pp. 3531–3538.

[14] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli, "Security evaluation of biometric authentication systems under real spoofing attacks," *IET biometrics*, pp. 11–24, 2012.

[15] S. Chen, A. Pande, and P. Mohapatra, "Sensor-assisted facial recognition: An enhanced bio-metric authentication system for smartphones," in *Mobisys '14*. ACM, 2014, pp. 109–122.

[16] X. Yang, C.-W. You, H. Lu, M. Lin, N. D. Lane, and A. T. Campbell, "Visage: A face interpretation engine for smartphone applications," in *Mobile Computing, Applications, and Services*. Springer, 2013, pp. 149–168.

[17] P. K. Misra, W. Hu, Y. Jin, J. Liu, A. Souza de Paula, N. Wirstrom, and T. Voigt, "Energy efficient gps acquisition with sparse-gps," in *IPSN '14*. IEEE Press, 2014, pp. 155–166.

[18] B. Wei, W. Hu, M. Yang, and C. T. Chou, "Radio-based device-free activity recognition with radio frequency interference," in *Sensys '2015*. ACM, 2015, pp. 154–165.

[19] B. Wei, M. Yang, Y. Shen, R. Rana, C. T. Chou, and W. Hu, "Real-time classification via sparse representation in acoustic sensor networks," in *Sensys '2013*. ACM, 2013, p. 21.

[20] Y. Shen, W. Hu, J. Liu, M. Yang, B. Wei, and C. T. Chou, "Efficient background subtraction for real-time tracking in embedded camera networks," in *Sensys '2012*. ACM, 2012, pp. 295–308.

[21] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, pp. 199–222, 2004.

[22] F. Alonso-Martín, M. Malfaz, J. Sequeira, J. F. Gorostiza, and M. A. Salichs, "A multimodal emotion detection system during human–robot interaction," *Sensors*, vol. 13, no. 11, pp. 15 549–15 581, 2013.

[23] A. Gee and R. Cipolla, "Determining the gaze of faces in images," *Image and Vision Computing*, pp. 639–647, 1994.

[24] M. Li, H. Yu, X. Zheng, and A. I. Mourikis, "High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation," in *ICRA '14*. IEEE, 2014, pp. 409–416.

[25] A. M. Sabatini, "Quaternion-based extended kalman filter for determining orientation by inertial and magnetic sensing," *IEEE Transactions on Biomedical Engineering*, pp. 1346–1356, 2006.

[26] C. Jia and B. L. Evans, "Online calibration and synchronization of cellphone camera and gyroscope," in *GlobalSIP*. IEEE, 2013, pp. 731–734.

[27] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *CVPR*. IEEE, 2003, pp. I–313.

[28] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, pp. 137–154, 2004.

[29] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, pp. 2876–2896, 2009.

[30] I. Van der Linde and T. Watson, "A combinatorial study of pose effects in unfamiliar face recognition," *Vision research*, pp. 522–533, 2010.

[31] M. Uřičář, V. Franc, and V. Hlaváč, "Detector of facial landmarks learned by the structured output svm," *VISAPP*, pp. 547–556, 2012.

[32] D. Donoho and Y. Tsaig, "Fast Solution of ℓ_1 -Norm Minimization Problems When the Solution May Be Sparse," *Information Theory, IEEE Transactions on*, vol. 54, no. 11, pp. 4789–4812, 2008.