Face Recognition under Camouflage and Adverse Illumination

Matthew Shreve, Vasant Manohar, Dmitry Goldgof, Sudeep Sarkar

Abstract—This paper presents a method for face identification under adverse conditions by combining regular, frontal face images with facial strain maps using score-level fusion. Strain maps are generated by calculating the central difference method of the optical flow field obtained from each subject’s face during the open mouth expression. Subjects were recorded with and without camouflage under three lighting conditions: normal lighting, low lighting, and strong shadow. Experimental results demonstrate that strain maps are a useful supplemental biometric in all three adverse conditions, especially in the camouflage condition, where a 30% increase in rank 1 recognition is observed over a baseline PCA-based algorithm.

I. INTRODUCTION

Over the last decade, a broad range of face identification systems have been proposed for application in surveillance and human authentication. While many of these systems work well under normal conditions, they perform poorly at identifying an individual with an altered appearance due to makeup, or more severely, camouflage. While some of these alterations may be inadvertent due to fashion, there is a clear indication that such a maneuver may be taken by individuals wanting to evade such systems [8]. Other adverse conditions such as uneven lighting conditions (e.g., bright sunlight or poorly illuminated nighttime settings) have also been repeatedly shown to drastically reduce the recognition rates of well-known solutions.

In this paper, we propose a bi-modal biometric method that uses score level fusion to combine standard, frontal images of a subject with information about his/her facial dynamics, or the non-rigid motion made during an expression. There are two main motivations for this approach: (i) recent psychological studies have indicated that human recognition is, at least in part, based on facial expressions [19]. For example, in [15] it was shown that a subject could be identified strictly by the projection of facial motion vectors observed during an expression on a generic model. And (ii) facial dynamics have been shown to be robust to lighting changes and other adverse conditions [5][16][20][9].

One way to observe the effects of facial tissue deformation is to calculate optical strain patterns as a force is applied, which occurs naturally during facial expressions. This has several advantages: (i) the strain pattern can easily be used in conjunction with many existing face recognition methods, such as PCA; (ii) the strain pattern captures the facial dynamics observed during an expression; (iii) the strain pattern represents the elasticity of facial skin, a distinct biomechanical property; (iv) as long as reliable motion vectors can be extracted from the video, strain patterns can be regarded as stable with respect to illumination changes and face camouflage; (v) as demonstrated in [14] strain patterns are robust to moderate amounts of head movement, which is common in surveillance application.

In this paper, we calculate optical strain maps using the central difference method over a robust estimation of optical flow [1]. We chose this approach for calculating optical strain because of its computational efficiency, thus making it a candidate for real-time application. A PCA-based recognition algorithm is used to generate scores for both regular intensity images and strain maps. Then, a score-level fusion technique is used to generate final rank recognition rates.

A. Related Work

There have been several recent literature surveys that encompass general face recognition algorithms [22], with some focusing on illumination invariant approaches [23][13]. However, face recognition under camouflage has had very little attention in the literature. Hence, we will briefly discuss relevant approaches from these surveys that relate to illumination-invariance of our approach, along with some of our earlier results relating to this study.

In general, we found that illumination-invariant methods for face recognition using standard imaging devices (e.g., a camcorder) fall under two categories: Those that preprocess or transform an image to remove the adverse effects of uneven lighting conditions, and those that collect multiple frames of information from a subject, so that collectively the effects of bad illumination are reduced. Those in the first category include techniques which may attempt to locally normalize areas in an image where there is poor illumination [18]. Also, this category includes methods that exploit illumination invariant features such as edges [6], relative image gradients and self-quotients [17], and facial symmetry [21].

Those in the second category fuse data from multiple images in order to gain the intrinsic illumination-invariant characteristics of the face. Our approach falls into this category. In the work by Chen et al.[5], regular intensity images were augmented with features generated from a set of motion vectors, resulting in increased robustness to illumination changes. Tsai et al. proposed a PCA feature-level fusion technique for combining features based on appearance and facial expression spaces [16]. Using a Digital Image Skin Correlation (DISC) method to establish motion vectors between two images, Pamudurthy et al. were successful at
identifying a small set of individuals, some wearing makeup [12]. In [7], Local Binary Patterns were used for generating spatio-temporal textures used for face recognition. In the work by Canavan et al. [4], the selection of images was based on the idea that implicit 3D information about a face can be obtained when it is rotated in front of a camera. A recognition rate increase from 63% using a single frame to 85% using 7-frame feature-level fusion was achieved for experiments where the gallery contained normal lighting conditions and the probe consisted of the strong shadow condition.

In our previous work [20][9], we identified several important qualities of strain maps. In contrast to the methods above that explicitly use motion vectors for recognition, strain maps are a derivation of motion fields. Hence, they represent both the facial dynamics observed during an expression and the elasticity of facial skin. Early experiments suggested that strain maps are robust to poor illumination conditions and camouflage, but on the profile face pose. We demonstrated similar, preliminary results for the camouflage condition on a frontal face dataset consisting of 5 subjects in [10]. Our results also indicated that using strain maps, inter-subject and intra-subject variation increased under the camouflage condition, most likely due to more accurate optical flow distortion due to noise. The second approach allows for the

II. Method

Facial strain patterns are derived from the observed facial motion found during an expression. Hence, optical strain maps can be calculated directly from the pixel intensity by integrating the strain definition into optical flow equations, or it can be calculated after optical flow has initially been estimated. The first approach requires the direct calculation of high order derivatives, which may result in increased distortion due to noise. The second approach allows for the optical flow fields to be smoothed through filtering and outlier detection before optical strain maps are calculated. Since our goal is to have robust and consistent strain maps across several adverse and potentially noisy conditions, we chose the second approach.

A. Optical Flow Estimation

Optical flow is a widely-used motion estimation technique based on the brightness conservation principle [1]. In order to obtain a reliable solution, it assumes (i) the intensity of any point in an image remains constant over a pair (or sequence) of frames, and (ii) pixels movement is consistent within a small image window. It is typically expressed by the following equation:

\[(\nabla I)^T \mathbf{p} + I_t = 0\]  

where \(I(x, y, t)\) represents the temporal image intensity function at point \(x\) and \(y\) at time \(t\), and \(\nabla I\) represents the spatial and temporal gradient. The horizontal and vertical motion vectors are represented by \(\mathbf{p} = [p = dx/dt, q = dy/dt]^T\).

We tried many different methods for generating high quality motion data, including Ogale [11] flow and SIFT [3] flow, however, we chose the framework given in [1] because it yielded the most consistent and reliable results. It is worth mentioning that perhaps an ensemble of these methods may lead to even better stability in flow estimation.

B. Optical Strain

The projected 2-D motion of any deformable object can be expressed by a vector \(\mathbf{u} = [u, v]^T\). If the motion is small enough, then the corresponding finite strain tensor is defined as:

\[\varepsilon = \frac{1}{2} [\nabla \mathbf{u} + (\nabla \mathbf{u})^T],\]

which can be expanded to the form:

\[\varepsilon = \begin{bmatrix} \varepsilon_{xx} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x} & \varepsilon_{xy} = \frac{1}{2} \left( \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \\ \varepsilon_{yx} & \varepsilon_{yy} = \frac{1}{2} \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \end{bmatrix}\]

where \((\varepsilon_{xx}, \varepsilon_{yy})\) are normal strain components and \((\varepsilon_{xy}, \varepsilon_{yx})\) are shear strain components.

Since each of these strain components are a function of displacement vectors \((u,v)\) over a continuous space, each strain component is approximated using the discrete optical flow data \((p,q)\):

\[p = \frac{\partial x}{\partial t} = \frac{\Delta x}{\Delta t}, u = p \Delta t,\]

\[q = \frac{\partial y}{\partial t} = \frac{\Delta y}{\Delta t}, v = q \Delta t\]

where \(\Delta t\) is the change in time between two image frames. Setting \(\Delta t\) to a fixed interval length, we can estimate the partial derivatives of (4) and (5):

\[\frac{\partial u}{\partial x} = \frac{\partial p}{\partial x} \Delta t, \frac{\partial u}{\partial y} = \frac{\partial p}{\partial y} \Delta t,\]

\[\frac{\partial v}{\partial x} = \frac{\partial q}{\partial x} \Delta t, \frac{\partial v}{\partial y} = \frac{\partial q}{\partial y} \Delta t,\]

Finally, the second order derivatives are calculated using the central difference method. Hence,

\[\frac{\partial u}{\partial x} = \frac{u(x + \Delta x) - u(x - \Delta x)}{2 \Delta x}, \frac{\partial u}{\partial y} = \frac{p(x + \Delta x) - p(x - \Delta x)}{2 \Delta x}\]

\[\frac{\partial v}{\partial y} = \frac{v(y + \Delta y) - v(y - \Delta y)}{2 \Delta y}, \frac{\partial v}{\partial x} = \frac{q(y + \Delta y) - q(y - \Delta y)}{2 \Delta y}\]

where \((\Delta x, \Delta y) \approx 2-3\) pixels.

Assuming that the stress applied uniformly across the entire image, then pixels with large strain values will correspond to a low elastic modulus, and vice versa. Hence, these values can then be normalized to (0-255) in order to generate strain images which reveal the underlying dynamics of the deformable object.
III. EXPERIMENTAL DESIGN

For simplicity in the following sections, regular intensity images will refer to the normal frames obtained from a video camera. The following section will describe data acquisition, strain map calculation, pre-processing, and experimental setup.

A. Video Collection

Videos were acquired from subjects as he/she opened his/her mouth under three different lighting conditions (normal, low, and strong shadow). A total of 35 subjects participated, 17 of which also applied camouflage. Videos were recorded using a JVC GY-HD100 camcorder with a pixel resolution of 1280×720. A few samples obtained under the three lighting conditions and camouflage are shown in Figures 1 and 2 respectively.

Fig. 1: Sample frames from normal, low, and strong lighting conditions.

Fig. 2: Sample frames of subjects wearing camouflage.

B. Facial Strain Maps

Strain maps were computed between a starting, neutral frame of the subject and the peak of the open mouth expression. The peak of the expression was automatically detected using algorithm described in [14]. In summary, this is done by first calculating optical flow between each pair of consecutive frames in the open mouth sequence. Next, vector linking is used to calculate the motion between the starting frame and every other frame. Once the expression boundaries are known, the peak of the expression is determined by locating the inclusive frame interval with the largest summed strain magnitude. Once final optical flow vectors are obtained, the data is smoothed using a Gaussian kernel (5 x 5) in order to reduce potential noise. Finally, the central difference method was used to generate facial strain maps for all subjects under all lighting and available skin conditions. Figure 3 shows some examples.

Fig. 3: Sample strain maps for open mouth expression, before geometric normalization and masking.

C. Geometric Normalization and Masking

Both the regular images and strain maps were geometrically normalized using a linear conformal transformation based on the eye coordinates of each subject. These points were manually identified since eye-trackers were not reliable for the strong shadow and camouflage conditions. An elliptical mask was applied to both the regular images and strain maps, in order to segment the boundary of the face. Additional, fixed masking was applied to the boundary regions of the eyes and mouth regions only for the strain maps, since these areas often contain noise and are not representative of the elasticity of the skin. All masking was done automatically based on fixed distances to the location of the eyes. Figure 4 contains final, normalized images for both modalities.

Fig. 4: Examples of geometrically normalized and masked images.

D. Experiments

The framework given in [2] was used to perform PCA-based face recognition experiments. The training and gallery sets contained images or strain maps obtained under the normal lighting condition. Table I lists all three experiments performed. Score-level fusion was then performed for all experiments containing the same lighting conditions for the gallery and probe sets, but with different modalities.

IV. RESULTS

In this section, we report the rank recognition rates for both the regular intensity modality along with our proposed method for each experiment shown in Table I. For all
TABLE I: Experimental setup for gallery and probe.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Gallery</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp-1</td>
<td>Normal light</td>
<td>Camouflage</td>
</tr>
<tr>
<td>Exp-2</td>
<td>Normal light</td>
<td>Low light</td>
</tr>
<tr>
<td>Exp-3</td>
<td>Normal light</td>
<td>Shadow light</td>
</tr>
</tbody>
</table>

experiments, min/max normalization was performed on both modalities. Rank recognition rates are reported using the cumulative match curve (CMC).

A. Camouflage (Exp-1)

In general, for appearance-based recognition, camouflaged faces pose a difficult challenge. Using only regular intensity images, the PCA-based face recognition algorithm achieves a low rank 1 recognition rate of 12%. This is mainly due to the severe changes in many discriminant features on the skin of each subject after camouflage is applied. It is interesting to note that although optical strain was not designed to be the leading (or dominant) biometric, it is observed to be a stronger modality than regular intensity images under this condition. Figure 5 shows some sample subjects with their corresponding strain maps. The scores of both two modalities were fused using the sum-rule. After fusion, we observe a significant gain of 30% in accuracy for the rank 1 recognition rate over regular intensity images (Figure 6). Moreover, we achieve on average a 20% increase in recognition rates for the first five ranks.

B. Low Lighting (Exp-2)

For regular images without fusion (Figure 7), the low lighting condition achieves relatively high (88% - 98%) recognition rates for early ranks (Figure 8). This robustness can mainly be attributed to the pre-processing performed before PCA experiments are run. In particular, histogram equalization is done on each normalized face image, thus reducing the effects of the low lighting condition. Because of this, there is only little potential gain that can be achieved by fusing strain map information. Still, a 3% increase is observed for rank 1, while a 100% recognition rate is achieved at rank 4 (opposed to rank 6 when only using regular intensity images). However, there is a small drop in performance (3%) for ranks 2 and 3. For this experiment, a weighted sum rule was used for fusion. Since regular intensity images are the more reliable modality for this condition, we found that biasing it by a factor of 4 worked well in our results.

C. Shadow Lighting (Exp-3)

The strong shadow lighting condition contains images of subjects with a single light source next to their head (Figure 9). This condition initially has a higher recognition rate using regular intensity images. However, past rank 2, using weighted fusion leads to an average 10% rate increase for all future ranks (Figure 10). It should be noted that this condition had the most detrimental effects on strain map calculation. Due to the increased smoothness of the skin texture caused from strong lighting, the optical sflow estimation became less precise, causing the corresponding strain maps to be less accurate. Similar to the low lighting condition, a weighted sum rule was used for fusion (regular intensity images were biased by a factor of 4).
THE RIGHT TWO COLUMNS ARE THE CORRESPONDING STRAIN MAP
REGULAR IMAGE INTENSITY SAMPLES, FROM THE GALLERY AND PROBE.

Fig. 9: Strong shadow condition. The left two columns are regular image intensity samples, from the gallery and probe. The right two columns are the corresponding strain map samples.

V. DISCUSSION

In this paper, we present a score-level fusion based method that increases the identification rates of a baseline appearance-based method under camouflage and adverse illumination. Our method is based on the idea that strain, a bio-mechanical property that represents the elasticity of facial soft tissues, can be estimated to provide discriminant information based on the facial dynamics and facial skin properties of an individual.

We have shown that fusing strain maps with regular intensity images leads to an overall increase in recognition rates for all adverse conditions. We observed a 30% increase in rank 1 recognition for identifying subjects who are wearing camouflage, while an overall 10% average increase in all ranks has been observed for the Shadow Light condition. Although the low lighting condition achieves relatively high recognition rates using only regular intensity images, fused results provided an increase of 3% for rank 1 and peaks at rank 4 (opposed to rank 6). Since strain maps can be easily calculated, they become a viable modality which can be easily fused with larger and more established biometrics.

REFERENCES


