SPECTRAL LWIR IMAGING FOR REMOTE FACE DETECTION

Dalton Rosario

U.S. Army Research Laboratory

ABSTRACT

We show the utility of hyperspectral (HS) Longwave IR (LWIR) imaging for remote sensing human face detection. A proof of principle experimentation considers a limited but challenging dataset of calibrated LWIR HS data cubes for four skin tone diverse human subjects, standing outdoors at three distinct ranges. An algorithm is developed to capitalize on two spectral features suitable for small sample size targets, using all of the available bands. The algorithm maps the two spectral features—each consisting of sample sizes significantly smaller than the number of frequency bands, as it simultaneously generates two large sets (reference and testing) of independent contrasts in a lower dimensional subspace. The large sample size in the new subspace allows for the development of a strong hypothesis test that functions as a canonical target detector. Results using real HS imagery are encouraging for a specific face detection scenario, where the range is assumed to be known a priori—200, 300, 400ft.

Index Terms— human face detection, hyperspectral, longwave infrared.

1. INTRODUCTION

Traditionally, face recognition systems use primarily spatial discriminants that are based on geometric facial features [1]. Many of these systems have performed well on databases acquired under controlled conditions. However, these approaches often exhibit significant performance degradation in the presence of changes in range or face orientation. Several of the limitations of current face recognition systems can be overcome by using spectral information. Recently, hyperspectral (HS) imaging has been applied to the face recognition problem [2]. On these efforts, justifiably, researchers often concentrate on exploiting absorption features in the visible region, owing to the presence of melanin and hemoglobin in human skin [3]. But, unfortunately and in contrast to HS LWIR imagery, imagery from passive visible HS systems can only be exploited when data are collected during daylight.

This paper checks the utility of using hyperspectral (HS) Longwave IR (LWIR) imaging for remote face detection. In particular, we present a non-kinematic based approach to autonomous target detection using a new set of Independent and Indirectly Generated Attributes (IIGA)

from HS imagery. IIGA specifically addresses the detection of rare signals, i.e., rare appearance of targets relative to the abundance of non-targets, or targets consisting of a small sample size. Examples of rare targets are specific motor vehicles and human faces.

IIGA shows an innovation in augmenting the target's sample size in a meaningful way and without loss of distinctness between targets and non-targets in the frames. IIGA aims at transforming while isolating an initially available target sample (of size n representing a few B-dimensional reference spectra) into a lower dimensional space m (the number of scored features), where the sample size N in this new space is large (N much greater than n and m). It does so by contrasting the target sample to N randomly selected blocks of data from a subsequent frame, where each block also has size n , in order to yield the reference IIGA set. The same blocks of data are then contrasted against a given testing sample, yielding the testing IIGA set. Since both attribute sets have a large sample size N , a sound multivariate statistical test can now be applied to perform target detection or rejection. The process is repeated until the entire frame is tested for the presence of target(s). The approach is flexible for the introduction of new contrasting metrics, and offers day and night capability. No assumption is given to the underlying distribution of data, and results using real LWIR HS imagery are encouraging.

The remainder of this paper is organized as follows: Section 2 discusses the sensor and data, Section 3 presents the IIGA method, and Section 4 shows some results.

2. SENSOR & DATA COLLECTION

Our data collection utilized a hyperspectral camera (HyperCam) from Telops, Inc. The spectral measurements are performed using a Fourier-Transform Spectrometer. It uses a 320 x 256 LWIR PV-MCT focal plane array detector. In its LWIR version, the instrument has good sensitivity over the 8-12 μm band.

As a proof of principle, we collected and calibrated HS data cubes using the LWIR HyperCam for four human subjects, representing males of distinct skin tones. Data were acquired at approximately 12:30 p.m., while the subjects were standing outdoors at three different ranges: 200 ft, 300 ft, and 400 ft. The data collection location is Picatinny Arsenal, NJ, where the day in September 2010 was mostly sunny with ground temperature near 80° Fahrenheit, 35%

humidity; the sky featured thin and scattered clouds. Figure 1 shows the LWIR band average image using 320 bands for two ranges—200 ft and 300ft. The HS data cubes have the same spatial resolution of 160×128 and a spectral resolution of 1.54 cm^{-1} wavenumbers (420 bands).

3. IIGA METHOD

The conceptual framework of IIGA is depicted in Fig. 2, where two HS samples $\mathbf{X}_{n \times B}^{(reference)} = \mathbf{X}_{n \times B}^{(frame\ i)}$ and $\mathbf{X}_{n \times B}^{(test)} = \mathbf{X}_{n \times B}^{(frame\ i+k)(r)}$ taken from two frames (*frame i* and *frame i+k*), are available for comparison, where $n \ll B$ (n representing the number of pixels in the sample and B the number of bands), i and k being integers greater than zero, and the image spatial index $r=1, \dots, R$. The goal is to automatically determine whether $\mathbf{X}_{n \times B}^{(test)}$ and $\mathbf{X}_{n \times B}^{(reference)}$ belong to the same population of objects.

Figure 1. Algorithmic framework.

Typically, since $n \ll B$, $\mathbf{X}_{n \times B}^{(reference)}$ and $\mathbf{X}_{n \times B}^{(test)}$ would be compared—directly—using a standard metric (e.g., Spectral Angle Mapper—SAM), but, instead, we propose to perform this comparison, indirectly, by first contrasting the reference sample $\mathbf{X}_{n \times B}^{(reference)}$ against a series of N randomly selected HS samples of the same size $\mathbf{Z}_{n \times B}^{(1)}, \dots, \mathbf{Z}_{n \times B}^{(N)} = \mathbf{Z}_{n \times B}^{(frame\ i+k)(1)}, \dots, \mathbf{Z}_{n \times B}^{(frame\ i+k)(N)}$ (see Fig. 1), where N is a free parameter. This process yields a new attribute set $\mathbf{Y}_{N \times m}^{(reference)}$, where N is much greater than m ($N \gg m$) and m equals the number of scored features that will be discussed later. This process continues by contrasting the same series of random selections $\mathbf{Z}_{n \times B}^{(1)}, \dots, \mathbf{Z}_{n \times B}^{(N)}$ against the test sample $\mathbf{X}_{n \times B}^{(test)}$ in order to yield the corresponding new attribute $\mathbf{Y}_{N \times m}^{(test)}$ for testing. Fig. 1 illustrates that the underlying distributions of the new attribute sets, shown for

convenience in Fig. 1 as univariate, can now be exploited using large sample statistical theory for a chosen large N . In particular, the means of the underlying multivariate distributions of $\mathbf{Y}_{N \times m}^{(reference)}$ and $\mathbf{Y}_{N \times m}^{(test)}$ can be contrasted using a version of the Hotelling's T^2 test [4].

The motivation for pursuing the indirect comparison approach shown in Fig. 1 is that we discovered that object distinctness can be preserved or accentuated by generating contrasts between a target's sample in the original data space and a series of other unrelated samples also in the same data space, where distinctness is represented by features of the new attribute sets' underlying distributions, as depicted in Fig. 2 (see the univariate based blue and red curves, shown in figure for illustration purpose only).

Interestingly, the individual contrast between $\mathbf{X}_{n \times B}^{(reference)}$ and the b^{th} sample $\mathbf{Z}_{n \times B}^{(b)}$ may not be unique, relative to the contrast between $\mathbf{Z}_{n \times B}^{(b)}$ and some other sample (e.g., $\mathbf{X}_{n \times B}^{(test)}$), but the object uniqueness, or distinctness, is preserved or accentuated by collecting these contrasts between $\mathbf{X}_{n \times B}^{(reference)}$ and each sample in the series $\{\mathbf{Z}_{n \times B}^{(b)}\}_{b=1}^N$. In this context, the contrast may be obtained by first computing mean averages (mean and median) of the available HS samples and then applying a linear or non-linear metric between the resulting mean averages. The new attribute set is finally generated for a particular object (known or unknown) by first capitalizing on the object's sample original data space distinctness, using all of the bands (contrasts between spectral averages), and then on the fact that a sufficiently large number of contrasts will draw a better picture on the distinctness of the particular object in the new attribute space. Recall that the number of contrasts (N) is a free parameter.

For the number m of features, we found the following features to be surprisingly effective in the context of difficult target detection: the spectral *mean* average and a coarse estimation of the spectral *skew* tendency, which can be readily attained by the difference between the spectral *mean* and *median* averages; illustratively, using univariate variables $\beta = \hat{\mu} - \tilde{\mu}$, where $\hat{\mu}$ is the mean average and $\tilde{\mu}$ is the median average. Notice that large values of $|\beta|$, where $|\cdot|$ is the absolute value operator, indicate stronger skew tendencies, either positive or negative; smaller values indicate weaker skew tendencies. So, in this paper, although $m=2$, three orthogonal features are inherently being considered, i.e., $\hat{\mu}$, $|\beta|$, and the sign of β .

The new attribute sets $(\mathbf{Y}_{N \times m}^{(reference)}, \mathbf{Y}_{N \times m}^{(test)})$ then consist of a large number N of bivariate scored features for $m=2$.

The IIGA conceptual framework discussed in this section can be formalized in the context of target detection as follows: Let us assume that an initial cue $\mathbf{X}_{n \times B}^{(f)}$, a HS

sample, is available as a reference set from image frame f , where $n \ll B$ and another HS sample $\mathbf{X}_{n \times B}^{(f+k)(r)}$ is also available as a testing set taken from another image, frame $f+k$, where r indexes a particular spatial location in frame $f+k$. These samples are rearranged as

$$\begin{aligned} \mathbf{X}_{n \times B}^{(f)} &= [\mathbf{x}_{01}^{(f)}, \dots, \mathbf{x}_{0n}^{(f)}]^t \\ \mathbf{X}_{n \times B}^{(f+k)(r)} &= [\mathbf{x}_{11}^{(f+k)(r)}, \dots, \mathbf{x}_{1n}^{(f+k)(r)}]^t \end{aligned} \quad (1)$$

where $\mathbf{x}_{ij}^{(f)} \in \mathbb{R}^B$, $\mathbf{x}_{ij}^{(f+k)(r)} \in \mathbb{R}^B$, B is the number of frequency bands, $i = (0,1)$, $j = (1, \dots, n)$, and the operator $(\bullet)^t$ means transposed.

We would like to transform $\mathbf{X}_{n \times B}^{(f)}$ and $\mathbf{X}_{n \times B}^{(f+k)(r)}$ to a feature space where their representations, namely $\mathbf{Y}_{N \times m}^{(f)}$ and $\mathbf{Y}_{N \times m}^{(f+k)(r)}$, respectively, have a large sample size N relative to m number of components within any vector in the new space, i.e., $(N \gg m)$. We also would like the vectors within $\mathbf{Y}_{N \times m}^{(f)}$ to be statistically independent, likewise for $\mathbf{Y}_{N \times m}^{(f+k)(r)}$.

The independence requirement can be handled by autonomously and randomly selecting N blocks of data $\mathbf{Z}_{n \times B}^{(f+k)(b)} (b=1, \dots, N)$ from frame $f+k$, such that the mean and skew estimates are computed using each available HS sample; followed by the application of a contrasting metric, in this case SAM, as a function of m . Mathematically, this approach using $\mathbf{X}_{n \times B}^{(f)}$ and $\mathbf{X}_{n \times B}^{(f+k)(r)}$, separately, against $\mathbf{Z}_{n \times B}^{(f+k)(b)} (b=1, \dots, N)$ yields $\mathbf{Y}_{N \times m}^{(f)}$ where

$$\mathbf{y}^{(f)(b)} = \begin{bmatrix} y_{Mean}^{(f)(b)} \\ y_{Skew}^{(f)(b)} \end{bmatrix}^t, \quad (2)$$

such that the f^{th} component is the corresponding SAM

$$\begin{aligned} y_{Mean}^{(b)} &= \arccos \left(\frac{\hat{\boldsymbol{\mu}}_0^T \hat{\boldsymbol{\mu}}_1^{(b)}}{\|\hat{\boldsymbol{\mu}}_0\| \|\hat{\boldsymbol{\mu}}_1^{(b)}\|} \right) \\ y_{Skew}^{(b)} &= \arccos \left(\frac{\boldsymbol{\beta}_0^T \boldsymbol{\beta}_1^{(b)}}{\|\boldsymbol{\beta}_0\| \|\boldsymbol{\beta}_1^{(b)}\|} \right) \end{aligned} \quad (3)$$

where $0 \leq y_{Mean}^{(b)} \leq \pi/2$, $0 \leq y_{Skew}^{(b)} \leq \pi/2$; $\boldsymbol{\beta}_0 = \hat{\boldsymbol{\mu}}_0 - \tilde{\boldsymbol{\mu}}_0$, $\hat{\boldsymbol{\mu}}_0 \in \mathbb{R}^B$ and $\tilde{\boldsymbol{\mu}}_0 \in \mathbb{R}^B$ are the skew, mean and median averages, respectively, using $\mathbf{X}_{n \times B}^{(f)}$; and, similarly, $\boldsymbol{\beta}_1 = \hat{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{\mu}}_1$, $\hat{\boldsymbol{\mu}}_1 \in \mathbb{R}^B$ and $\tilde{\boldsymbol{\mu}}_1 \in \mathbb{R}^B$ are the skew, mean and median averages, respectively, using the b^{th} random block of data $\mathbf{Z}_{n \times B}^{(f+k)(b)}$; and $\|\bullet\|$ is the square root of the squared component sum.

Likewise, $\mathbf{Y}_{N \times m}^{(f+k)(r)}$, where its components are

$$\mathbf{y}^{(f+k)(r)(b)} = \begin{bmatrix} y_{Mean}^{(f+k)(r)(b)} \\ y_{Skew}^{(f+k)(r)(b)} \end{bmatrix}^t \quad (4)$$

using the b^{th} random block of data $\mathbf{Z}_{n \times B}^{(f+k)(b)}$ to calculate the estimates.

Equipped with a large sample size statistical problem, a strong hypothesis test can now be utilized to function as a binary classifier, as shown next.

Using $\mathbf{Y}_{N \times m}^{(f)}$ and $\mathbf{Y}_{N \times m}^{(f+k)(r)}$, let

$$\Delta_{N \times m}^{(f, f+k)(r)} = \mathbf{Y}_{N \times m}^{(f)} - \mathbf{Y}_{N \times m}^{(f+k)(r)} \quad (5)$$

be the difference between $\mathbf{Y}_{N \times m}^{(f)}$ and $\mathbf{Y}_{N \times m}^{(f+k)(r)}$ over the SAM scored features, where

$$\Delta_{1 \times m}^{(f, f+k)(r)(b)} = \mathbf{y}_{1 \times m}^{(f)(b)} - \mathbf{y}_{1 \times m}^{(f+k)(r)(b)} \quad (b=1, \dots, N). \quad (6)$$

Using (5) and (6) as input to compute corresponding mean average and covariance estimates yields $\bar{\Delta}_{1 \times m}^{(f, f+k)(r)}$ and $\Sigma_{m \times m}^{(f, f+k)(r)}$, respectively. Notice that $\bar{\Delta}_{1 \times m}^{(f, f+k)(r)}$ represents a normalized sum of independent random variables, where, according to [4], one can utilize a α -level test of $H_0 : \boldsymbol{\delta}_{1 \times m} = \mathbf{0}_{1 \times m}$ versus $H_1 : \boldsymbol{\delta}_{1 \times m} \neq \mathbf{0}_{1 \times m}$, rejecting H_0 if the observed

$$\begin{aligned} T^{2(f, f+k)(r)} &= N \left(\bar{\Delta}_{1 \times m}^{(f, f+k)(r)} - \boldsymbol{\delta}_{1 \times m} \right) \left(\Sigma_{m \times m}^{(f, f+k)(r)} \right)^{-1} \left(\bar{\Delta}_{1 \times m}^{(f, f+k)(r)} - \boldsymbol{\delta}_{1 \times m} \right)^T \\ &> \frac{(N-1)m}{N-m} F_{m, N-m}(\alpha), \end{aligned} \quad (7)$$

where $F_{m, N-m}(\alpha)$ is the upper $(100\alpha)\%$ percentile of an F -distribution with m and $N-m$ degree of freedom.

In essence, equation (7) tests whether the average difference $\bar{\Delta}_{1 \times m}^{(f, f+k)(r)}$ is statically close to a zero vector $\mathbf{0}_{1 \times m}$ of the same size. The test is repeated across the testing imagery by using $\mathbf{Y}_{N \times m}^{(f)}$ and $\mathbf{Y}_{N \times m}^{(f+k)(r)}$ and changing spatial index $r=1, \dots, R$. This produces a 2-dim output surface. Examples of output surfaces are shown in Section 4. The overarching process can again be repeated using $\mathbf{X}_{n \times B}^{(f)}$ as reference to test a different image frame.

4. EXPERIMENTAL RESULTS

We conducted an experiment to determine, given the limitations of the data, whether the algorithm described in Section 3 can be used to detect the frontal face of a specific targeted human subject standing at 200 ft, 300ft, and 400 ft. A key element in this experiment is that the features are computed using samples of the target while it stood at a closer range (200 ft) and tested using the same features while the target stood at different ranges, although the range in each image is assumed known. We repeated the test by posing each human subject as the target.

Experimental results using the pseudo color depicted in Fig. 2 are shown in Fig 3, using the frontal faces of Subjects A (far left), B (second from left), C (second from right), and D (far right) individually as targets (initial cues), where cues ($\mathbf{Y}_{N \times m}^{(f)}$) were only taken from frame 1 but tested in frame 1 (range 200ft), frame 2 (range 300ft), and frame 3 (range 400ft), $n = 187$, $B = 320$, N was fixed to 600 for all frames, and the testing window size was also fixed to 5×5 for all frames. Independently of the frame being tested, at any given testing location, contrasting features in $\mathbf{Y}_{N \times m}^{(f+k)(r)}$ were estimated between the cue and one of the randomly selected blocks of data and between the testing location and the same blocks of data. This process yields a scalar T^2 , as in (7). The interpretation of T^2 relative to the hypotheses H_0 and H_1 and the pseudo color used to display the output surfaces is shown in Fig. 2 for a probability of error.

Figure 2. Targets are detected if T^2 falls below the threshold $\frac{(N-1)m}{N-m} F_{m, N-m}(\alpha)$. The pseudo color used to display output surfaces depicts white for the presense of targets and yellow through black for the absense of targets, where yellow depicts non-target results closer to the threshold while black shows the results farthest away.

Results shown in Fig. 3 are organized as follows, from top to bottom and left to right: the first horizontal row (row 1) shows the LWIR average band images for the given ranges; from row 2 through row 5 the output surfaces using the proposed algorithm are shown. The output surfaces shown in row 2 depicts the results for detecting Subject A, while rows 3, 4, and 5 depict results for detecting Subjects B, C, and D, respectively. Although those results are qualitative, we were pleasantly surprised and pleased with

depicted performances. Fig. 3 shows that, for the example data cubes, the proposed algorithm was able to isolate the frontal faces of the targeted human subjects, given that targeted cues were taken from a closer range but also tested at farther ranges. The limited dataset is sufficiently challenging, given the goal of the experiment and skin tone diversity of the human subjects.

Figure 3. Algorithm output surfaces using individual target cues from frame 1 only. In Frame 1 (top left), targets are at 200ft; Frame 2 (center) the reange is 300ft; and Frame 3, the range is 400ft.

There are many unknowns for attempting to explain the reason why this approach worked so well under the experiment bounds. We conjecture that the mean and skew estimates, which used all of the 320 bands for the computations, jointly with the large sample size N representing the target's individual contrasts against independent blocks of data played a positive role. Perhaps the concept of testing two samples indirectly by generating differences between each sample and a multitude of other samples and comparing, instead, the resulting outputs, does have its merit in separating the target from non-targets in the new attribute space.

5. REFERENCES

- [1] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Am. A*, vol. 14, pp. 1,724–1,733, 1997.
- [2] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1552–1560, Dec. 2003.
- [3] M. van Gemert, S. Jacques, H. Sternborg, and W. Star, "Skin optics," *IEEE Trans. Biomed. Eng.*, vol. 36, pp. 1146–1154, December 1989.
- [4] E. L. Lehmann. *Testing Statistical Hypotheses*, New York: transferred to Chapman & Hall, 2nd Edition, 1993.