PERSON RE-IDENTIFICATION IN MULTI-CAMERA SURVEILLANCE SYSTEMS

Thesis

Submitted to

The School of Engineering of the UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for

The Degree of

Master of Science in Electrical Engineering

By

Kevin C. Krucki

UNIVERSITY OF DAYTON

Dayton, Ohio

December, 2015
PERSON RE-IDENTIFICATION IN MULTI-CAMERA
SURVEILLANCE SYSTEMS

Name: Krucki, Kevin C.

APPROVED BY:

Vijayan K. Asari, Ph.D.
Advisor Committee Chairman
Professor, Department of Electrical and Computer Engineering

Yakov Diskin, Ph.D.
Committee Member
Adjunct Faculty, Department of Electrical and Computer Engineering

Eric J. Balster, Ph.D
Committee Member
Associate Professor, Department of Electrical and Computer Engineering

John G. Weber, Ph.D.
Associate Dean
School of Engineering

Eddy M. Rojas, Ph.D., M.A., P.E.
Dean
School of Engineering
© Copyright by

Kevin C. Krucki

All rights reserved

2015
ABSTRACT

PERSON RE-IDENTIFICATION IN MULTI-CAMERA SURVEILLANCE SYSTEMS

Name: Krucki, Kevin C.
University of Dayton

Advisor: Dr. Vijayan K. Asari

In a system of cameras, it can be beneficial to track and identify people as they move through the scene. To solve this problem (called human re-identification) appearance matching through feature extraction must be applied to detected humans. We propose an algorithm that combines color features with soft biometric features; namely clothing identification that distinguishes pants from shorts, and long sleeve shirts from short sleeve shirts and backpack and sling bag detection. First, person extraction is performed using a mixture of Gaussians background subtraction and simple blob analysis. Next, a subject's color features are calculated using color histograms in the RGB and HSV color spaces. Clothing identification is performed by detecting skin on the top and bottom half of an extracted person. Backpack and sling bag detection is accomplished with advanced blob analysis. A label is calculated for each of these soft biometrics depending upon the percentage of frames a detection occurs in. Once a person is located, a signature of that real world subject is obtained by combining the labels and color features. An average signature of each individual that appears in a given camera (training camera) is then calculated by averaging all the signatures from a specific camera of that subject. The signatures gathered from other cameras (testing cameras)
are then compared to the average signatures from the training camera using chi-squared distance measurements between the color histograms and between their cumulative distribution functions. These distances are ranked and then reranked according to the labels of the signatures. A correctly identified person has a ranking of one. These final ranks are shown in Cumulative Matching Characteristic (CMC) curves. The most significant of the many challenges involved are the variation of illumination conditions, pose, and viewpoint across the cameras. The algorithm is tested on the SAIVT-SoftBio dataset and promising results for human re-identification on multi-camera systems are observed. Research work is progressing for human re-identification based on computer rendered human models.
For my parents Tim and Kathy. You taught me everything I know.
ACKNOWLEDGMENTS

I would like to sincerely express my appreciation to my advisory committee, Dr. Vijayan Asari, Dr. Eric Balster, and Dr. Yakov Diskin. Special thanks go out to Dr. Asari for advising throughout my time in graduate school and for giving me the opportunity to work in the Vision Lab. It has truly been the most rewarding experience of my life and I will never forget it. All of my colleagues in the Vision Lab are greatly appreciated as well for their suggestions and encouragement throughout this process. I would also like to express my gratitude to the ECE academic staff at the University of Dayton. They gave me the tools to succeed in completing this work. Lastly, thanks to my parents, family, and friends for their support throughout this process.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Challenges and Motivation</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Research Objectives and Proposed Solution</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Thesis Outline</td>
<td>5</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Hard Biometrics</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Appearance Models</td>
<td>9</td>
</tr>
<tr>
<td>2.2.1 Feature-Based Approaches</td>
<td>9</td>
</tr>
<tr>
<td>2.2.2 Learning Approaches</td>
<td>12</td>
</tr>
<tr>
<td>III. HUMAN RE-IDENTIFICATION USING SOFT BIOMETRIC RERANKING</td>
<td>14</td>
</tr>
<tr>
<td>3.1 Person Extraction</td>
<td>14</td>
</tr>
<tr>
<td>3.2 Color Feature Model</td>
<td>16</td>
</tr>
<tr>
<td>3.3 Backpack and Sling Bag Detection</td>
<td>19</td>
</tr>
<tr>
<td>3.3.1 Sling Bag Detection</td>
<td>19</td>
</tr>
<tr>
<td>3.3.2 Backpack Detection</td>
<td>22</td>
</tr>
<tr>
<td>3.4 Clothes Identification</td>
<td>26</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1.1 A basic human re-identification scenario. A person should have the same label when walking through several camera views. .......................................................... 2

1.2 The proposed method for solving the human re-identification problem. ............. 6

3.1 An example of background subtraction and person extraction ............................ 16

3.2 An example of a persons body being split at the waist and neck. The head is discarded because of its view variance. Color histograms are taken of the upper and lower body. .......................................................... 17

3.3 3D Representation of HSV color space. Hue goes around a circle from 0-360 degrees, while value and saturation control the chromacity of the hue from 0-255. 18

3.4 Three subjects transformed from RGB color space to HSV color space. ............. 18

3.5 The integral image and an example operation. On the left is a sample of image values. The image on the right shows the calculated integral image. Underneath each image is the way area is calculated. The integral image uses two less adds to get the same number. .......................................................... 20

3.6 Slingbag appearance from different viewpoints with different human pose. ....... 20

3.7 A Bradley conversion and minimum bounding box calculation ....................... 21

3.8 Backpacks from the left, right, front and back look extremely different. ............ 22

3.9 In black is the extracted backpack. The red dashed line shows the convex hull. The blue circles are convex points and the green circles are convex defects. The angle measured from convex defects to convex points, \( \theta \), is used to detect a backpack. 23
3.10 Angles corresponding to those that signal a backpack detection. The left image shows a top defect, the middle image shows a left defect, and the right image shows a right defect.

3.11 Confusion matrix of bag detection from [1]. Slingbags are labeled backpacks and vice versa a little more than 20% of the time. Therefore, the descriptor is combined into one category of “bag” and “no bag”.

3.12 The clothes classification process.

3.13 An overview of the feature extraction and labeling process for one person in a one camera.

4.1 Examples of STTF being applied to face detection (top row), haze/fog removal(second row) and color restoration, (third row).[2]

4.2 Nonlinear intensity transformations for image enhancement through STTF. The tunable parameter adjusts the transformation based on the local features of each pixel.[2]

4.3 An example of STTF enhanced background subtraction. The top row is the original background subtraction without STTF. The bottom row is the STTF enhanced background subtraction. The left image is the original frame, the middle image is the extracted person, and the right image is the background subtraction of the extracted person. The STTF enhanced background subtraction is markedly better than the original background subtraction.

5.1 The layout of cameras in the SAIVT dataset.

5.2 Examples of images in each of the cameras in the SAIVT dataset.

5.3 CMC curves using simple identifiers. All identifiers perform at about the same level.

5.4 CMC curves using soft biometric reranks and STTF rerank. The rerank and STTF improvement is significant.
LIST OF TABLES

3.1 Confusion matrix for bag detection on the SAIVT-SoftBio dataset. . . . . . . . . . . 25

3.2 Confusion matrix for shirt identification on the SAIVT-SoftBio dataset. . . . . . . 26

3.3 Confusion matrix for lower body clothing identification on the SAIVT-SoftBio dataset. 27

4.1 Confusion matrix for bag detection on the SAIVT-SoftBio dataset with STTF pre-

4.2 Confusion matrix for shirt identification on the SAIVT-SoftBio dataset with STTF

4.3 Confusion matrix for lower body clothing identification on the SAIVT-SoftBio dataset

4.4 True positive improvement of bag detection and clothing identification using STTF. 33

5.1 Rank statistics for several different identifiers. . . . . . . . . . . . . . . . . . . . 36

5.2 Rank statistics for several different identifiers. . . . . . . . . . . . . . . . . . . . 39
CHAPTER I

INTRODUCTION

Human re-identification in video surveillance has become an increasingly important research topic in recent years. Due to the reducing cost of cameras and advancement of data storage technologies, more companies, schools, and government facilities have installed camera networks than ever before. Public transportation centers, including bussing companies and more notably airports are at the forefront of this effort. The proliferation of such camera systems has allowed governments to gather intelligence, prevent crime, and analyze previously recorded video to a much larger extent than ever before. Traditionally, video monitoring has relied on individual security officers to watch an increasingly large number of video feeds and make observations. Unfortunately, as more video feeds are added, humans find it more difficult to make accurate assessments of the information provided to them. Therefore, video analytics that can process several camera feeds at the same time and make the same astute observations that a human can need to be developed. In particular we will focus on human re-identification. The goal of human re-identification research is to be able to recognize specific people in a system of cameras with non-overlapping views. More precisely, we would like to develop an algorithm that can see a person in one camera and extract a set of features that the algorithm can learn. When the same person walks through a different camera, we want the algorithm to be able to compare them to the learned features and identify them. A basic human
re-identification scenario can be in Fig 1.1, where several subjects pass through different camera feeds but retain the same label.

Figure 1.1: A basic human re-identification scenario. A person should have the same label when walking through several camera views.

There are several different real world scenarios in which human re-identification can be useful. The general scenario is that cameras are placed and then data is recorded and evaluated in real-time or after an event (such as a crime). An algorithm will sift through the saved video, determine where humans are located, and output which persons are the best matches to one another.

1.1 Challenges and Motivation

While the human re-identification problem could be seen as a retrieval task, where every person wearing a certain type of clothing (e.g. white shirt and khaki pants) is recalled from the cameras, we are more concerned with distinguishing within these categories to find an exact person. There are many different challenges to solving this problem. Most real-world cameras have insufficient resolution or are too far away from a detected person to perform high-level face recognition upon.
In many cases, a person will walk through a camera without showing their face, rendering face detection and recognition impossible. Therefore, a person's entire appearance must be used to perform person re-identification. There are several challenging issues to consider when implementing this type of algorithm.

The most challenging and difficult problems to solve in relation to human re-identification are camera view variation and person appearance variation. Camera view variation occurs when human re-identification is attempted on cameras that have different fields of view. This could be due to, amongst other reasons, the position, tilt, zoom, or model of the camera. For example, a camera positioned on the floor pointing parallel to the ground shows a person differently than a camera raised above the floor pointing down. This will be true even if these cameras are pointed at the same area. Lighting changes due to the environment may also cause significant camera view variation. This problem is worsened when a camera system has cameras placed both indoors and outdoors.

A person's appearance can change drastically depending on a multitude of factors. The clothes a person is wearing, their body shape, where they are detected in a scene, what they are carrying and which direction they pass through a camera's field of view can all be causes of person appearance variation. Every person looks markedly different from the front, back, and side. If a person is carrying an item their left and right side may differ as well. This problem is compounded when a person is wearing a shirt with a logo on the front but not on the back. Human re-identification can be even more difficult when a person is wearing a backpack, because the color on the front of the shirt will completely differ from that on the back in a 2D image. In addition, the lighting differences espoused in the previous paragraph can create distinct differences in color of a scene. This means that the same article of clothing could be viewed as a different color in different cameras.

Another significant problem for human re-identification algorithms is person detection and extraction. Firstly, detecting a person in a cluttered scene can be extremely difficult. More occlusions
in a scene leads to a harder human detection. This is especially true when those occlusions are other people. Pose variation of humans can also affect human detection, especially if people are not in an upright position. Secondly, extracting a person from a scene's background poses an interesting challenge. Typically, motion segmentation is performed but this does not account for a person standing still in a scene for a long period of time. In addition, segmentation can be negatively affected by lighting changes in the camera's field of view.

1.2 Research Objectives and Proposed Solution

Our specific research objectives are to:

1. Use soft biometrics to improve basic color based human-re-identification. The soft biometrics used are an advanced backpack and sling bag detector and skin detection based clothes identification to create labels in three categories: whether or not a person is carrying a sling bag or backpack, whether they are wearing pants or shorts, and whether they are wearing a short sleeve shirt or a long sleeve shirt. This information will be used in conjunction with color histograms to create a feature model. The color features between humans in different cameras will be compared using different distance metrics. These distances are sorted and ranked in descending order and then re-ranked based on the soft biometric labels to create final rankings. Person re-identification results will then be shown based on these re-rankings.

2. Use a self-tunable transformation function (STTF) to improve algorithm performance through image enhancement as a pre-processing stage. This image enhancement will help to overcome lighting variations between cameras and improve person extraction through the use of motion detection. STTF will also improve the soft biometric labeling process.

To accomplish these objectives several other steps need to take place. First, motion detection of a scene is completed using a mixture of Gaussians method and a foreground frame is calculated.
Large blobs in the foreground scene are extracted as humans. This allows us to extract humans from the background of the scene. Next, color histograms and their cumulative distribution functions in the RGB and HSV space are gathered and calculated. Then, backpack, sling bag, and skin detection for clothing identification are performed and labels for each classification are assigned to the extracted person. This occurs for each frame that a person appears in for a particular camera, resulting in two color histograms and three labels for each frame. The histograms from these frames are averaged to create a single average color model and three overall labels for each individual person are determined using majority voting. The average histograms and overall labels form a feature model. Feature models are calculated for all individuals in all cameras in the system. For testing, a particular training camera is chosen and the feature models in of people that appear in this camera compose the probe set. All other cameras in the system are testing cameras and people in these cameras compose the gallery set. A feature model from the testing camera is compared to all feature models in other cameras. Color histograms are compared using chi-squared distance measurements. The distances will be sorted and then ranked. The smallest distance will be ranked one, the next smallest two, and so on. The ranks will then be updated and re-ranked based on the amount of labels matched between the feature models. An overview of this process can be seen in Fig. 1.2 and will be investigated further in Chapters 3-5.

1.3 Thesis Outline

The ensuing material in this thesis is outlined as follows: Chapter 2 is a literature review of human re-identification topics. Chapter 3 describes the proposed method in full, including human extraction, color feature histogram calculation, backpack and sling bag detection, and clothing identification. Chapter 4 discusses using a self-tuning transfer function (STTF) as a pre-processing step of the algorithm. Chapter 5 discusses the experiments performed and results from the experiments,
focusing on the impact made by soft biometric re-ranking and STTF. Chapter 6 provides conclusions and discusses future research in the field of human re-identification.

Figure 1.2: The proposed method for solving the human re-identification problem.
CHAPTER II

LITERATURE REVIEW

Human re-identification as a research topic has expanded significantly in recent years. This has led to several innovative and interesting algorithms being developed to solve the human re-identification problem. Most researchers have attacked the problem in three stages: person extraction, person feature representation, and person matching. The first problem, person detection, already has extensive research dedicated to it, and therefore human re-identification research has assumed that this problem has been solved. Therefore, most have looked to improve feature representation, person matching, or both. These methods can be broken down into two categories: hard biometric methods and global appearance methods.

Hard biometrics, including gait, iris, and face recognition, are some of the most effective human re-identification features because they are extremely unique. This is opposed to soft biometrics or attributes (e.g. hair color, eye color, and body size) that describe a person in more general terms. Unfortunately, most camera systems are placed in cluttered environments that have trouble extracting detailed enough information to perform these tasks. Therefore a different approach must be taken: appearance modeling.

Appearance based methods attempt to provide a robust model that categorizes a person across different camera views. Appearance methods search for features and soft biometrics that transfer
across views (e.g. color, texture, and height). Before delving into general appearance models we first discuss hard biometric methods.

2.1 Hard Biometrics

Hard biometrics are characteristics that define a person specifically and can be either physical or behavioral. Physical biometrics include DNA, facial features, iris features, retina scan, and fingerprints. Behavioral biometrics include handwriting, gait, and voice. A camera is limited in that it can only observe the features that are visual: iris recognition, retinal scans, gait and face recognition. Unfortunately, most cameras are too far away to perform retinal scans or iris recognition. In addition, gait recognition is an extremely difficult task unless a ground-level camera from the side of a person is readily available. In most cases this is not realistic, and therefore we will discuss only face recognition.

Face recognition has been a major research topic for a number of decades. Turk and Pentland [3] were the first to attain successful results using the Eigenface method. This method uses principal component analysis (PCA) to reduce the dimensionality of face images in each class. Then, each face can be represented as several vectors which match the eigenvectors from its class. This process was improved using Fisher’s linear discriminant analysis (LDA) [4] to minimize the variation within a class while maximizing the variation between classes. LDA also improved face recognition under different lighting conditions [5]. This method was named Fisherfaces. These algorithms work extremely well as long as there are no pose variations in the face.

More recently, local features have been used as accurate facial feature representations. These features include Gabor Wavelets [6], Discrete Cosinus Transform [7] and Local Binary Patterns [8]. Most successfully, Yaigman et.al.[9] designed DeepFace, a machine learning algorithm that utilizes neural networks to identify humans. The algorithm performs irrespective of lighting and viewpoint
variations and is as good at face recognition as a human is. Unfortunately, these programs all need
detailed facial features to train their algorithms. This is simply not a realistic scenario for most real-
life camera systems. People frequently walk through a camera scene without showing their face at
all and even if they do there is no guarantee that the resolution of the camera will be high enough to
identify them. Therefore, we need to utilize appearance models for human re-identification.

2.2 Appearance Models

Appearance based models face several challenges, namely variety of camera viewpoints, illumina-
tion variance, and human pose variation. Therefore, an appearance based model must be invariant
to these problems but still be distinctive enough to distinguish between classes of people. Models
for humans are represented by one or many features that are extracted from an image. These fea-
tures will compose a signature for the person. The most basic forms of appearance based models
take low-level features in histogram format. More advanced algorithms looked at the relationships
between feature spaces to find correlations. Overall, appearance based algorithms fall into two cat-
egories: feature-based approaches and learning approaches. The following sections will cover these
topics separately.

2.2.1 Feature-Based Approaches

Feature-based approaches do not require any human supervision or annotation before use. There-
fore, feature-based approaches must be robust to view and pose change, occlusions, lighting vari-
ation, and cross camera differences. There are three main features associated with human re-
identification: color, shape, and texture. Color information is found in a variety of spaces, including
but not limited to RGB (red, green, blue), HSV (hue, saturation, value), and yCbCr(luminance,
blue difference chroma, red-difference chroma). Local-binary patterns (LBP) [10]and Gabor [11]
filters are examples of both local and global features that describe texture and shape. While color
[12, 13, 14, 15] is the most important feature, it is extremely sensitive to differences in lighting of a scene and camera parameters and is not a distinct enough feature to rely on alone. Therefore many have looked to combine color features with light invariant texture and shape-based features [16, 17, 18, 19]. Matching for feature-based algorithms is typically completed using nearest neighbor distance algorithms (Manalobis, L1, L2, cosine distance, chi-squared, Kolmogorov, or Bhattacharyya). More advanced matching algorithms are discussed in section 2.2.2.

Gray and Tao [19] take an ensemble of localized features (ELF) composed of color (RGB, HSV, YCbCr) and texture (Schmidt and Gabor filters) features. In their work they found the most discriminative features were hue, saturation, red channel, blue channel, and Schmidt and Gabor filters equally. Gallagher et.al. [20] uses clothing segmentation in addition to face recognition. Using a clothing mask learned from mutual information extracted from images with the same appearance, clothing regions are represented using 5-dimensional vectors composed of RGB color values and horizontal and vertical edge detectors. Park et.al. [21] take color histograms over the shirt and pants region that are manually extracted (1/5 of the image is the head cutoff and 3/5 is the waist). Only shirt and pants regions are considered due to the pose variant nature of the human head. Histograms are computed in the HSV space with a heavy weight on the hue channel.

Cai et.al. [22] utilizes a patch-based approach using edge detection. Square regions (patches) are extracted around edges and represented by its most dominant color and its frequency within the patch. The similarity of patches together is matched along with geometric information. Wang et.al. [23] segment an image into regions and characterize their colors using a co-occurrence matrix. A histogram of oriented gradients (HOG) [24] in the Log-RGB color space is calculated due to the illumination invariant nature of the Log-RGB color space. A modified shape context algorithm is utilized to identify human body parts. Lastly, a occurrence/co-occurrence function that describes probability distributions and their correlations over the image region is used for matching.
Madden et al. [25] look to minimize the effect of lighting variation between cameras. A histogram equalization process called “controlled equalization” is applied to a detected person and k-means clustering extracts the major colors. These colors are combined with the equalized histogram and similarity is calculated using a Kolmogorov distance.

Bak et al. [26] use the deformable parts model and STMicroelectronics background subtraction for person extraction. A normalized image is divided into dense grid structures with 32x32 regions and a 16 pixel overlap. Each region has an RGB, LBP, and HOG histogram, and a covariance averaged on a Riemannian manifold [27]. These features are then compared using the Brownian descriptor.

Xiaojing et al. [28] use 6 equally sized horizontal strips, 8 color channels (RGB, HS from HSV, CbCr from yCbCr), 21 texture features (13 Schmidt filters, 8 Gabor filters) and soft biometrics to rerank after matching. They train support vector machines (SVM) for backpacks, jeans, carrying an item, short hair, and male and rerank individuals based on these categories. Unfortunately, the SVM performance is so poor that reranking improvement is small.

More advanced features, such as SURF (Speeded Up Robust Features) [29] and SIFT (Scale Invariant Feature Transform) [30] have also been used in human re-identification. Hamdoun et al. [31] compiles SURF points across time for an individual and matches them using a sum of absolute differences and KD-trees [32]. Zhao et al. [33] use SIFT features in addition to chromium information to address pose-variation. These methods are not experimented with as frequently in human re-identification as in other computer vision fields because Gheissari et al. [34] proved that these descriptors perform worse than color and texture based features.
2.2.2 Learning Approaches

Learning approaches can be broken down into two categories: metric learning methods and discriminative methods. In human re-identification, a person manually provides label information for the system to learn. Metric learning methods involve offline training where data is given in positive (same person in two different cameras) and negative (different person in different cameras) pairs. On the other hand, discriminative methods are trained online in real-time.

Metric learning methods that improve feature representation do so through direct appearance modeling [35, 36] or indirect appearance modeling through feature mapping [15, 37]. There are also methods that improve matching using distance metric learning [38, 14]. More specifically, Prosser et.al. [39] built brightness transfer functions to learn lighting variations in different camera views. Li et.al. [40] proposed a Filter-Pairing Neural Network (FPNN) that is trained using deep learning. FPNN can learn feature representation, geometric and lighting variance between cameras, and identity matching at the same time. Gray, first introduced in section 2.2.1, uses Adaboost to learn the similarity function between two appearances. Dikmen et.al. [41] propose a cost function similar to large margin nearest neighbor (LMNN) [42] called LMNN-R. This algorithm introduces the rejection condition, where the system returns no match if nearest neighbors are at a great distance from the model.

Discriminative methods look to amplify distinctive characteristics of a specific individual by looking at the appearances of others. Schwartz et.al. [43] uses the Partial Least Squares statistical tool. A feature set of color, texture, and edge descriptors is extracted and trained to give higher weights to features that are discriminative. Nakajima et.al. [44] trains multi-class SVMs to learn color, shape, and local shape features. Each appearance is separated from others based on the tree structure of pairwise SVMs. Teixera et.al. [45] uses a bag of SIFT features to improve matching accuracy. SIFT vectors are calculated and put into visual words using the vocabulary tree to define
a descriptor space. Lastly, Hirzer et al. [14] uses principal component analysis (PCA) to reduce the dimensionality of a dense grid of features. Due to the need for manually annotated data, learning methods fare poorly in real-world scenarios compared to feature-based models.
CHAPTER III

HUMAN RE-IDENTIFICATION USING
SOFT BIOMETRIC RERANKING

As stated before, the proposed method takes several steps to complete. The algorithm goes through four main steps before testing: person extraction, color feature calculation, backpack and sling bag detection, and clothes identification. This chapter describes how a feature model for each person is calculated.

3.1 Person Extraction

In order to extract humans from a scene, background subtraction is performed using a mixture of Gaussians model [46]. This model is online and adaptive. It considers the values of a pixel over time to be a “pixel process”. The “pixel process” is a series of pixel values (in our case vectors because the images are RGB) and at any time $t$ what is known about a pixel $\{x_0, y_0\}$ is its history

$$\{X_1, ..., X_t\} = \{Im(x_0, y_0, i) : 1 \leq i \leq t\} \quad (3.1)$$

where Im is the image sequence and $\{X_1, ..., X_t\}$ is the “pixel process”. If a scene contained a static background and no lighting variation the pixel value would be relatively constant. If independent Gaussian noise is experienced at the pixel then its density could be described by one Gaussian distribution centered at the mean pixel value. However, lighting conditions could change or a static object could be added to the scene and its corresponding pixels could be considered foreground until
it was there longer than the previous background. To overcome these issues, a mixture of Gaussians, rather than a single Gaussian distribution, is used. The recent history of each pixel, \( \{X_1, \ldots, X_t\} \), is modeled by a mixture of \( K \) Gaussian distributions. The probability of observing the current pixel value is

\[
P(X_t) = \sum_{i=1}^{K} w_{i,t} \ast \eta(X_t, \mu_{i,t}, \Sigma_{i,t})
\]  

(3.2)

where \( K \) is the number of distributions, \( w_{i,t} \) is an estimate of the weight of the \( i^{th} \) Gaussian in the mixture at time \( t \), \( \mu_{i,t} \) is the mean value of the \( i^{th} \) Gaussian in the mixture at time \( t \), \( \Sigma_{i,t} \) is the covariance matrix of the \( i^{th} \) Gaussian in the mixture at time \( t \), and \( \eta \) is a Gaussian probability density function shown below.

\[
\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t-\mu)^T \Sigma^{-1} (X_t-\mu)}
\]  

(3.3)

Once these parameters are initialized an initial foreground detection is made. The first \( B \) Gaussian distribution which exceeds an arbitrary threshold \( T \) is kept for a background distribution.

\[
B = \text{argmin}(\sum_{k=1}^{b} w_k > T)
\]  

(3.4)

A mixture model for every pixel in the image is computationally costly. Therefore an on-line K-means approximation is implemented. Every new pixel value, \( X_t \), is checked against the existing \( K \) Gaussian distributions, until a match is found. A match is defined as a pixel value within 2.5 standard deviations of a distribution. The weights are then updated accordingly depending on whether or not there was a match. If there is no match then the pixel is kept as foreground.

Once all foreground pixels are detected, 2D connected component analysis is completed to collect all significant blobs. It is assumed all blobs that are of a significant size are humans. A
bounding box of the person is then extracted and saved. An example of background subtraction and person bounding box extraction using this method can be seen in Fig. 3.1

![Figure 3.1: An example of background subtraction and person extraction](image)

### 3.2 Color Feature Model

The color representation of a person can be affected by many attributes including hair color, skin color, type of clothes, and shoe color, amongst others. These attributes can vary greatly with respect to view. For instance, a person's head from the front, back and side has different colors depending on how much hair they have. To combat this variance of view, we would like to find the most consistent color features in an image, irrespective of pose or camera view. The attributes with the strongest invariance to view variation are color of lower body clothing and upper body clothing. Therefore, we take color histograms of the upper and lower body without considering the head. The process of splitting the body into lower half, torso, and head regions is based on one laid out in Denman et al. [47]. Search areas sized to be the width of the image by 10% of the height of the image are setup at likely positions of the neck (80% of height of the image) and waist (50% of the height of the image). Then each column's vertical gradient is calculated. The waist and neck lines are set where the maximum gradient location is in each column. An example of this process is laid out in Fig. 3.2. Color histograms are then calculated from these two parts of the body. A color histogram is a representation of the distribution of colors in an image. Several different color spaces
Figure 3.2: An example of a person's body being split at the waist and neck. The head is discarded because of its view variance. Color histograms are taken of the upper and lower body.

were considered for the color feature. After testing, the HSV (hue, saturation, value) and RGB (red, green, blue) color spaces were observed to yield the best test results. Equations (3.5-3.7) show the conversion from the RGB color space to the HSV space.

\[
V = \max(R, G, B) \tag{3.5}
\]

\[
S = \begin{cases} 
\frac{V - \min(R, G, B)}{V} & \text{if } V \neq 0 \\
0 & \text{otherwise.} 
\end{cases} \tag{3.6}
\]

\[
H = \begin{cases} 
60(G - B) & \text{if } V = R \\
\frac{120 + 60(B - R)}{V - \min(R, G, B)} & \text{if } V = G \\
\frac{240 + 60(R - G)}{V - \min(R, G, B)} & \text{if } V = B 
\end{cases} \tag{3.7}
\]

Saturation and value are values from 0-255 while hue is in degrees from 0-360. A 3-dimensional representation of the HSV color space can be seen in Fig. 3.3. Three dimensional HSV and RGB histograms are computed using 15 bins per channel on a detected human. Therefore, each histogram is $15^3$ or 3375 elements long. These histograms are normalized in accordance with how many pixels are detected from the background subtraction. This step is critical because it allows the system to
Figure 3.3: 3D Representation of HSV color space. Hue goes around a circle from 0-360 degrees, while value and saturation control the chromacity of the hue from 0-255.

maintain invariance to the size of the detected human. The final results are the normalized color histograms $H_{RGB}$ and $H_{HSV}$. For distance testing purposes the cumulative distribution functions of these histograms ($CDF_{RGB}$ and $CDF_{HSV}$) is also computed. This is done so that pixels on the edge of bins are penalized less in the distance measurement. An example of three images transformed to the HSV color space for three different people is shown in Fig. 3.4. The differences illustrate why the HSV colorspace is an effective descriptor to differentiate between subjects.

Figure 3.4: Three subjects transformed from RGB color space to HSV color space.
3.3 Backpack and Sling Bag Detection

For backpack and sling bag detection we use a process similar to that proposed in [1]. Based upon observations, bags are typically darker than clothes. Therefore, a morphological operation must be performed on a detected person to extract the darker pixels. To start, we consider the upper body as calculated in Section 3.2. This region of interest is normalized to 75 x 100 pixels.

3.3.1 Sling Bag Detection

To begin, a Bradley local thresholding is applied to the extracted ROI [48]. A Bradley threshold utilizes an integral image computed as follows,

\[ I(x, y) = f(x, y) + I(x - 1, y) + I(x, y - 1) - I(x - 1, y - 1). \]  \hspace{1cm} (3.8)

where \( I(x, y) \) is the integral image and \( f(x, y) \) is the original image.

In simpler terms, an integral image is calculated from the original image on a pixel by pixel basis. Each pixel in the integral image is a sum of all the pixels above and to the left of a particular pixel in the original image. This process is illustrated in Fig. 3.5. Once the integral image is calculated, the sum of any rectangle of the original image can be computed using the equation

\[ \sum_{x=x_{1}}^{x_{2}} \sum_{y=y_{1}}^{y_{2}} f(x, y) = I(x_{2}, y_{2}) - I(x_{2}, y_{1} - 1) + I(x_{1} - 1, y_{2}) + I(x_{1} - 1, y_{1} - 1). \]  \hspace{1cm} (3.9)

where \( x_{1}, y_{1} \) and \( x_{2}, y_{2} \) are the \( x \) and \( y \) coordinates of the top left and bottom right corners of the desired area. Fig. 3.5 shows an example operation, where the four points listed in Eq. 3.9 correspond to the bold points in the integral image. The Bradley threshold then computes the average of an \( s \times s \) window of pixels centered around each pixel. If the value of the current pixel is an arbitrary \( T \) percent less than this average then it is set to 255. Otherwise it is set to 0. The pixels that are set to 0 are the darker pixels that we use to determine whether or not a sling bag or backpack is present.

As seen in Fig. 3.6 sling bags can appear in many different ways on a person. However, the main
Figure 3.5: The integral image and an example operation. On the left is a sample of image values. The image on the right shows the calculated integral image. Underneath each image is the way area is calculated. The integral image uses two less adds to get the same number.

Figure 3.6: Slingbag appearance from different viewpoints with different human pose.

constant is the dark black strap over a lighter background. The strap is a long rectangle, meaning its borders have parallel lines Therefore, two different detectors designed to detect the existence of nearly parallel lines are designed. The first detector uses geometrical properties to measure the
parallelism of the strap. Given the perimeter of a blob extracted using a Bradley thresholded image, $L_c$, we compute the minimum area bounding box of the approximated contour. An example of the Bradley transformation and minimum bounding box calculation can be seen in Fig 3.7. The parallelism of the blob is then calculated as follows:

$$P = \frac{2 \times \max(b_w, b_h)}{L_c}$$

(3.10)

where $b_w$ and $b_h$ are the width and height of the minimum area. As $P$ increases, so does the parallelism of the blob. If the blob is long and rectangular then $P$ should be close to 1. Therefore, a threshold of $P \geq 0.8$ is chosen to detect the presence of a sling bag.

The second detector utilizes the edges of the potential bag region extracted using a canny edge detector. This canny edge detector is applied to the original image and no Bradley thresholding is taking place. A probabilistic Hough transform is then computed on these edges to find nearly parallel lines with width between them less than 15 pixels, a minimum length of 20 pixels, and an angle between $35^\circ$ and $90^\circ$. Whereas the first detector is unable to detect a case where a dark background blends with the strap, this detector is able to detect the partial near-parallel lines even
though the blob is not rectangular. If a sling bag is detected in the first step then the second step is skipped.

### 3.3.2 Backpack Detection

As shown in Fig. 3.8 backpacks look extremely different from several viewpoints. The same Bradley threshold that was performed for the sling bag detection takes place on the original image and the largest dark blob is considered. A contour analysis is used to extract the external boundary of the blob. The convex hull is then calculated and convex points and defects are located. Fig 3.9 shows an example of this process with the convex hull in red, convex points in blue, and convex defects in green.

![Figure 3.8: Backpacks from the left, right, front and back look extremely different.](image)
Figure 3.9: In black is the extracted backpack. The red dashed line shows the convex hull. The blue circles are convex points and the green circles are convex defects. The angle measured from convex defects to convex points, $\theta$, is used to detect a backpack.

After discarding convex defects with a depth from convex points of less than 3 pixels, the angles from each convex defect to its closest convex points are calculated and three types of defects are defined; top, left and right (see Fig. 3.10 for all defect definitions). A top defect which composes

Figure 3.10: Angles corresponding to those that signal a backpack detection. The left image shows a top defect, the middle image shows a left defect, and the right image shows a right defect.
the ‘V’ shaped break around the shoulders is detected if the following conditions are met (angles correspond to the left image in Fig. 3.10):

\[
0^\circ \leq \theta_{T1} \leq 85^\circ \\
95^\circ \leq \theta_{T2} \leq 85^\circ \\
30^\circ \leq |\theta_{T1} - \theta_{T2}| \leq 135^\circ
\]

A left defect that defines the left edge of the backpack is detected if (angles correspond to middle image in Fig. 3.10):

\[
65^\circ \leq \theta_{L1} \leq 170^\circ \\
195^\circ \leq \theta_{L2} \leq 270^\circ \\
30^\circ \leq |\theta_{L1} - \theta_{L2}| \leq 165^\circ
\]

A right defect that defines the right edge of the backpack is detected if (angles correspond to left image in Fig. 3.10):

\[
0^\circ \leq \theta_{R1} \leq 105^\circ \\
270^\circ \leq \theta_{R2} \leq 345^\circ \\
30^\circ \leq |\theta_{R1} - \theta_{R2}| \leq 165^\circ
\]

This process is performed on each extracted person. For a single person, each frame then has either a “no bag”, “sling bag”, or “backpack” label. In Chua et.al. there were several instances where slingbags were labeled backpacks and vice versa (see Fig. 3.11). Therefore we combine these two labels to create one binary category: “bag” and “no bag”. If the percentage of frames that contain a label of “bag” is greater than 50% this label is termed a “bag” for that person. Otherwise the label is “no bag”.

24
Figure 3.11: Confusion matrix of bag detection from [1]. Slingbags are labeled backpacks and vice versa a little more than 20% of the time. Therefore, the descriptor is combined into one category of “bag” and “no bag”.

Results of the sling bag and backpack detection from testing on the SAIVT dataset (see section 5.2 for more details) in the form of a confusion matrix can be seen in Table 3.1. In Table 3.1, results are shown as correctly identified if a majority of frames have a correct detection. An ideal case would be if the top left cell and bottom right cell were equal to 100%. This would mean that all people with bags (“actual”) were “identified” as “bag” and all people without bags (“actual”) were identified as “no bag”. Although this is not quite the case, detection results are strong, with a bag being detected over 88% of the time and no bag being detected over 85% of the time.

Table 3.1: Confusion matrix for bag detection on the SAIVT-SoftBio dataset.

<table>
<thead>
<tr>
<th>Identified</th>
<th>No Bag</th>
<th>Bag</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Bag</td>
<td>85.64%</td>
<td>14.36%</td>
</tr>
<tr>
<td>Bag</td>
<td>11.23%</td>
<td>88.77%</td>
</tr>
</tbody>
</table>
Table 3.2: Confusion matrix for shirt identification on the SAIVT-SoftBio dataset.

<table>
<thead>
<tr>
<th></th>
<th>Identified Short Sleeve</th>
<th>Identified Long Sleeve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td>83.11%</td>
<td>16.89%</td>
</tr>
<tr>
<td>Sleeve</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long</td>
<td>13.06%</td>
<td>86.94%</td>
</tr>
</tbody>
</table>

3.4 Clothes Identification

For clothes identification we do not directly train a classifier for identifying short-sleeve shirts, long-sleeve shirts, shorts, or pants. Instead we perform skin detection in the same manner as [49]. Skin pixels reside in a small range of the H and S channels of the HSV color space. More specifically, skin pixels are generally found in the H band between $0^\circ$ and $50^\circ$ and in the S band between 55 and 175 (see Section 3.2 for details on the HSV color space). This operation is performed on the background subtracted and extracted human. Unfortunately there are some pixels that fall into this category that may not be skin, found mostly in the background. Therefore, a k-means clustering is performed on the original extracted person without background subtraction, meaning all pixels in the bounding box are considered. All k-means blobs that have more than 80% of pixels classified as skin are then considered arms or legs. Therefore, depending upon which part of the body is being analyzed, either shorts or a short-sleeve shirt, is present. This process takes place on the lower 50% of the upper body and lower body images as calculated in Section 3.2 in all frames of a person in one camera. If the percentage of frames that have skin detection in the upper body is greater than 50% then the label is “short sleeve shirt” and otherwise it is “long sleeve shirt”. If the percentage of frames that have skin detection in the upper body is greater than 50% then the label is “shorts” and otherwise it is “pants”. An overview of this process can be seen in Fig. 3.12.

Results of skin detection and clothes identification from testing on the SAIVT-SoftBio dataset (see section 5.2 for more details) in the form of a confusion matrix can be seen in Tables 3.2 (shirt) and 3.3. Results are shown as correctly identified if a majority of frames have a correct detection.
Table 3.3: Confusion matrix for lower body clothing identification on the SAIVT-SoftBio dataset.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shorts</td>
</tr>
<tr>
<td>Shorts</td>
<td>91.45%</td>
</tr>
<tr>
<td>Pants</td>
<td>17.19%</td>
</tr>
</tbody>
</table>

An overview of the feature extraction and labeling process can be seen in Fig. 3.13.
Figure 3.12: The clothes classification process.
Figure 3.13: An overview of the feature extraction and labeling process for one person in a one camera.
CHAPTER IV

EFFECTIVE IMAGE ENHANCEMENT USING STTF ON SOFT BIOMETRICS AND PERSON RE-IDENTIFICATION

As a pre-processing technique, the self-tunable transformation function (STTF) [2] algorithm has already been proven to improve face detection, color restoration, and haze/fog removal (Fig. 4.1). However, it has never been applied to human re-identification, background subtraction, backpack and sling bag detection, or skin detection.

Figure 4.1: Examples of STTF being applied to face detection (top row), haze/fog removal(second row) and color restoration, (third row).[2]
STTF is a pixel-based nonlinear transformation technique that uses a sine function with a tunable parameter that changes based on the pixels characteristics. Like the human eye, the algorithm will tune sine functions locally by considering the statistical information of its neighborhood region. The idea is to create one function that can treat both dark and light regions in an image. This function is represented as an ArcSin function:

$$I_{new}(k) = \frac{2}{\pi} \arcsin(x^k)$$

and its characteristics can be seen in Fig. 4.2.

Figure 4.2: Nonlinear intensity transformations for image enhancement through STTF. The tunable parameter adjusts the transformation based on the local features of each pixel.[2]

4.1 Impact on Background Subtraction, Bag Detection and Clothing Identification

Unfortunately, the impact that STTF has on background subtraction cannot be quantified due to a lack of ground truth data. However, it can be visualized in several examples seen in Fig. 4.3. Improvement of this type can be seen throughout the dataset. It is especially prominent in areas that go from dark to bright very quickly.
Figure 4.3: An example of STTF enhanced background subtraction. The top row is the original background subtraction without STTF. The bottom row is the STTF enhanced background subtraction. The left image is the original frame, the middle image is the extracted person, and the right image is the background subtraction of the extracted person. The STTF enhanced background subtraction is markedly better than the original background subtraction.

There is a way to quantify the effect STTF has on backpack and bag detection and clothes detection. Tables 4.1, 4.2, and 4.3 show the updated confusion matrices.

Table 4.1: Confusion matrix for bag detection on the SAIVT-SoftBio dataset with STTF pre-processing.

<table>
<thead>
<tr>
<th>Actual</th>
<th>No Bag</th>
<th>Bag</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Bag</td>
<td>96.42%</td>
<td>3.58%</td>
</tr>
<tr>
<td>Bag</td>
<td>5.25%</td>
<td>94.75%</td>
</tr>
</tbody>
</table>

A table showing true positive improvement attained using STTF can be seen in Table 4.4. There is improvement across the board for all classifications.
Table 4.2: Confusion matrix for shirt identification on the SAIVT-SoftBio dataset with STTF pre-processing.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Calculated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short Sleeve</td>
</tr>
<tr>
<td>Short Sleeve</td>
<td>89.87%</td>
</tr>
<tr>
<td>Long Sleeve</td>
<td>8.78%</td>
</tr>
</tbody>
</table>

Table 4.3: Confusion matrix for lower body clothing identification on the SAIVT-SoftBio dataset with STTF pre-processing.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Calculated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shorts</td>
</tr>
<tr>
<td>Shorts</td>
<td>93.63%</td>
</tr>
<tr>
<td>Pants</td>
<td>10.66%</td>
</tr>
</tbody>
</table>

Table 4.4: True positive improvement of bag detection and clothing identification using STTF.

<table>
<thead>
<tr>
<th></th>
<th>No STTF</th>
<th>STTF</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Bag</td>
<td>85.64%</td>
<td>96.42%</td>
<td>+10.82%</td>
</tr>
<tr>
<td>Bag</td>
<td>88.77%</td>
<td>94.75%</td>
<td>+5.98%</td>
</tr>
<tr>
<td>Short Sleeve</td>
<td>83.11%</td>
<td>89.87%</td>
<td>+6.76%</td>
</tr>
<tr>
<td>Long Sleeve</td>
<td>86.94%</td>
<td>91.22%</td>
<td>+4.28%</td>
</tr>
<tr>
<td>Shorts</td>
<td>91.45%</td>
<td>93.63%</td>
<td>+2.18%</td>
</tr>
<tr>
<td>Pants</td>
<td>82.91%</td>
<td>88.34%</td>
<td>+5.43%</td>
</tr>
</tbody>
</table>
CHAPTER V

EXPERIMENTS AND RESULTS

The testing and matching process, dataset overview, and evaluation of the algorithm is presented in this chapter. First, the initial testing and matching with only color models is discussed. Next, the re-ranking process is defined. Lastly, the dataset being tested on is described and results are shown. Results shown illustrate the effectiveness of soft biometric re-ranking and STTF on human re-identification.

5.1 Testing and Matching Process

The testing process takes the features gathered from the process shown in Fig 3.13 and compares them across different cameras. To begin, the color features, composed of four different identifiers: an RGB histogram $H_{RGB}$, an HSV histogram $H_{HSV}$, and both of their cumulative distribution functions $CDF_{RGB}$ and $CDF_{HSV}$ from a testing camera are compared using the chi-distance squared function. This distance function is described in Eq. 5.1, where $D$ is the distance, $H_{test}$ is the test identifier and $H_{train}$ is the train identifier. The identifiers are compared across all elements $j$ to total elements $P$.

$$D = \sum_{j=0}^{P} \frac{(H_{test}(j) - H_{train}(j))^2}{H_{test}(j) + H_{train}(j)}$$  \hspace{1cm} (5.1)

These distances are calculated for each identifier in the feature model and then sorted and ranked. The smallest distance will receive the rank 1, the next smallest 2, and so on.
These ranks are then re-ranked based on the number of matching labels gathered from bag detection and clothing identification. This will take place for all labels, meaning that ranks who have all three labels matched with the original will move to the top, those with two matched labels move directly behind them, and so on. Ranks with the same amount of matches retain the same order from the color measurements.

5.2 Dataset

The dataset tested on is the SAIVT-SoftBio dataset [50]. This dataset contains 152 people filmed in 8 cameras. Each person is not in each camera. The dataset is extremely challenging due to lighting variations between cameras, occlusions in the form of furniture and people, and the wide variety of items being carried or worn by people in the dataset (e.g. backpacks, hats, helmets, and bags). The layout of these cameras can be seen in Fig. 5.1 and corresponding example views can be seen in Fig. 5.2.

![Figure 5.1: The layout of cameras in the SAIVT dataset.](image)

5.3 Results

Results are shown using Cumulative Matching Characteristic (CMC) curves. CMC curves show the probability (y-axis) of a person being ranked within the top \(x\) (x-axis) matches. Tables will also
be presented to compare to state of the art methods at specific ranks. We want to show that re-ranking and STTF pre-processing improve human reidentification. CMC curves calculated without STTF pre-processing are shown in Fig. 5.3. Each color identifier individually, combined, and without re-ranking has its own CMC curve. Although it is difficult to tell since the results are extremely similar, the best result comes from using $CDF_{RGB}$, as seen in Table 5.1. Table 5.1 shows the CMC curves of the different identifiers at different ranks. This shows that each of the different identifiers performs at about the same rate. The ranks correspond to the closest matches. Therefore, when using $CDF_{RGB}$ a person will be correctly identified if asking for only the closest match 1.07% of the time. A person will be correctly identified in the 5 closest matches 15.83% of the time and in the 25 closest matches 43.38% of the time.

Since the $CDF_{RGB}$ identifier led to the best results and because identifiers perform at about the same level, we will only analyze this identifier for the remaining stages. Fig. 5.4 shows CMC
Figure 5.3: CMC curves using simple identifiers. All identifiers perform at about the same level.
Figure 5.4: CMC curves using soft biometric reranks and STTF rerank. The rerank and STTF improvement is significant.

curves calculated using only $CDF_{RGB}$, $CDF_{RGB}$ with re-ranking (called Rerank), and re-ranking with STTF (STTF-Rerank). The best result is gathered from using STTF and re-ranking, although most of the improvement is due to the basic reranking. We can see from these results that STTF and re-ranking have a massively positive influence on the algorithm. The reranking by itself raises performance substantially. The addition of STTF also improves performance though in a less substantial way. The improvements at ranks 1, 5, and 20 can be seen in Table 5.2.

Table 5.2 also shows a comparison to state of the art algorithms from Bak et.al [26]. Although the first rank for our method is less than others, the rankings for 10 and 25 are far greater than all
other methods. This shows that our method compares favorably to others. An exact person match is less successful with our algorithm. However, when considering the best 10 or 25 matches our algorithm is the most successful.
CHAPTER VI

CONCLUSIONS AND FUTURE WORK

The goal of this thesis research was to show that the use of soft biometrics and image enhancement improved algorithms for person re-identification. First, a mixture of gaussians background subtraction was used to extract persons from the scene. Then color features in the RGB and HSV color space are extracted. Soft biometric labels are then gathered through backpack and slingbag detection and clothes identification through skin detection. Distance measurements were gathered using chi-squared distance and cumulative distribution functions. Distances are ranked and then reranked using the soft biometric. This algorithm is repeated, but a self-tunable transfer function (STTF) is applied beforehand to improve all stages.

The results prove that both soft biometric reranking and image enhancement improve human re-identification algorithms. STTF improves background subtraction, color feature extraction, backpack and slingbag detection, and skin detection. This human re-identification method was tested on the SAIVT-SoftBio dataset and compared to state of the art methods. Results shown in the form of Cumulative Matching Characteristic (CMC) curves prove the algorithm is effective.

Future work for this topic can go in two different directions. First, more soft biometrics (e.g. shirt texture (plaid, stripes, etc.), hair color, and others) could be added to the algorithm. This will in turn make re-identification a simpler task, as demonstrated by positive results from the results section of this thesis. The second area of research to look into is computer render aided human
re-identification. This research initiative utilizes computer generated images to create a potential subject at any view, pose, or lighting from a camera. Therefore, rather than comparing a person in different camera views, you can take the first view and create all possible views as a model. The renderings created could use camera parameters to optimize comparison between camera view and rendering. Preliminary results on this research is promising.
BIBLIOGRAPHY


