Part-based spatio-temporal model for multi-person re-identification

A. Bedagkar-Gala, Shishir K. Shah *

Quantitative Imaging Laboratory, University of Houston, Department of Computer Science, Houston, TX 77204-3010, USA

ARTICLE INFO

Article history:
Available online 12 September 2011

Keywords:
Person re-identification
Sparse representation
Representative color descriptor
Open-set matching
Closed-set matching

ABSTRACT

In this paper we propose an adaptive part-based spatio-temporal model that characterizes person's appearance using color and facial features. Face image selection based on low level cues is used to select usable face images to build a face model. Color features that capture the distribution of colors as well as the representative colors are used to build the color model. The model is built over a sequence of frames of an individual and hence captures the characteristic appearance as well as its variations over time. We also address the problem of multiple person re-identification in the absence of calibration data or prior knowledge about the camera layout. Multiple person re-identification is a open set matching problem with a dynamically evolving and open gallery set and an open probe set. Re-identification is posed as a rectangular assignment problem and is solved to find a bijection that minimizes the overall assignment cost. Open and closed set re-identification is tested on 30 videos collected with nine non-overlapping cameras spanning outdoor and indoor areas, with 40 subjects under observation. A false acceptance reduction scheme based on the developed model is also proposed.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Single and multiple camera person tracking is an important problem relevant to many computer vision applications such as automated surveillance, human–computer interaction, and vehicle navigation (Yilmaz et al., 2006). Consistent tracking of a person in multi-camera scenarios requires the ability to re-identify the person as he/she leaves the field-of-view (FOV) of one camera and reappears in the FOV of another. Re-identification is most commonly defined as a problem of matching videos or images of a person taken from different cameras. This is a rather challenging problem due to changes in the person’s appearance observed across cameras stemming from varying illumination, camera viewpoints, pose and partial occlusions. Further, due to the articulated nature of the human body, person re-identification presents a large range of pose variations (Wang et al., 2007). Biometrics such as face or gait can be used for matching but usually are difficult to recover due to the resolution or frame rate constraints of typical cameras (Gheissari et al., 2006). In multiple camera tracking, invariant feature based appearance models have been most commonly used for re-identification and spatio-temporal relationships between cameras are used to reason about false matches (Javed et al., 2008; Makris et al., 2004). Typically, proposed solutions deal with extracting invariant color, texture, shape or local features that impart discriminative ability to the person model to handle geometric and photometric variations (Cai and Pietikinen, 2010; Wang et al., 2007; Schwartz and Davis, 2009). Such features are extracted from the person’s body and no specific features from the head/face region are extracted. In this paper, we propose a part based spatio-temporal person model that combines color and facial features. The model either includes or excludes facial features based on the availability of usable face region images. Fig. 1 shows the model generation framework. In the absence of good face images, the person model is based only on color features. Since the facial feature inclusion is conditional, the model does not entirely depend on person’s face being visible in the video.

Person models are most often studied in closed set experiments where the gallery is fixed and re-identification is established on a single probe (Farenzena et al., 2010; Zheng et al., 2009; Schwartz and Davis, 2009). Moreover, the probe subject is always assumed to be present in the gallery. In contrast, person re-identification in the context of tracking across multiple cameras is an open set matching problem where the gallery evolves over time, the probe set dynamically changes for each camera FOV, and all the probes within a set are not necessarily present in the gallery. In addition, there might be several subjects that co-exist in time and need to be re-identified simultaneously. Thus, re-identification is not a single person but a multiple person matching problem. In this paper, we study the multiple person re-identification problem and treat it as a rectangular assignment problem and solve this combinatorial optimization under a minimum total error constraint.

Re-identification is done across each camera pair independent of their geospatial positions, assuming no prior knowledge about camera layout or temporal relationships between cameras. The
first time a person is seen, his/her appearance model is learned, and the subject is enrolled in the gallery set. Gallery is a set of people IDs previously seen. Thus, all people observed in the second camera form the probe set. After re-identification, all the people observed in the second camera who were previously unseen are enrolled into the gallery. As the re-identification moves to the next camera pair the gallery set is extended.

The following are the principle contributions of this paper. We propose an adaptive part-based spatio-temporal model based on color and facial features to address the multi-person re-identification problem. The model is studied for use in both closed and open set matching and a false acceptance reduction criterion is proposed based on a cost function using the developed models.

Multiple person re-identification experiments are carried out on 30 surveillance videos, involving 40 people, collected from a camera network with nine non-overlapping cameras spanning outdoor and indoor areas. Open set experiments explore the model’s generality and closed set experiments demonstrate its effectiveness. The rest of the paper is organized as follows: Section 2 presents the related work and Section 3 details our approach. Section 4 presents the data set and the experiments performed, and provides a discussion of the results. Conclusions are presented in Section 5.

2. Related work

Majority of prior research in person re-identification is based on enhancing the discriminative ability of models derived from features such as color, texture and shape. Multiple representations of these features have also been exploited. Temporal and spatial distribution of these features has been leveraged for the matching task. Different models have been proposed in literature, and can be categorized based on the features or combination of features used. Purely color based models have been proposed in (Cai and Pietikinen, 2010; Bird et al., 2005). Spatial distributions of self similarities with respect to color words based on Hue histograms are learned and combined into a descriptor in (Cai and Pietikinen, 2010). A linear discriminant method is used for HSL based color feature matching in (Bird et al., 2005). Color based covariance descriptors extracted from different body parts are used in (Bak et al., 2010b). Texture and color features are combined to generate an appearance model in (Bazzani et al., 2010; Farenzena et al., 2010; Bak et al., 2010a). They mainly differ from each other in type and representation of these features to form the model. Color and edge based descriptors are extracted in (Kang et al., 2004; Gheissari et al., 2006). The edge features are incorporated in order to capture the structural information from the person’s appearance. Shape and appearance context models were used in (Wang et al., 2007), where co-occurrences between apriori learned shape and appearance words form the person descriptor. Similar appearance labels and vital visual context using local and global spatial relationships are used to describe an individual in (Zheng et al., 2009). Interest point based local descriptors like PCA-SIFT and SURF are used in (Arth et al., 2007; Hamdoun et al., 2008) to capture the local appearance variations. Implicit Shape Models are used to capture shape properties of a person in (Jungling et al., 2011) and SIFT descriptors are used for matching. In (Gray and Tao, 2008), Ada-boost learning is employed to simultaneously learn discriminant features and ensemble of classifiers for re-identification. Partial least squares technique is employed to not only find discriminative weighting for color, texture and edge features, but also as a means to reduce descriptor dimensionality in (Schwartz and Davis, 2009).

None of these approaches incorporate facial features into the person model. The primary reason being extraction of discriminative face region features is challenging due to the quality of face images obtained from multi-camera systems. Wide camera views render the proportion of the face region very small compared to the entire image and the face is frequently blurred due to motion (Ao et al., 2009). Gabor, wavelet or Fourier features extracted from the face are not as effective in describing the facial characteristics due to the geometric shape of the face (Yang et al., 2007). Facial features that adapt to the variations in the facial characteristics like Eigenfaces (Turk and Pentland, 1991) or Fisherfaces (Belhumeur et al., 1997) are more relevant. Nonetheless, these features need the underlying face images to be perfectly aligned. Due to the resolution constraints of the head region, alignment between face images is difficult to achieve using standard appearance or landmark based registration methods. Further, most of these facial features used for face recognition breakdown due to changes in incident illumination, head pose, expressions and age (Zhao et al., 2003). According to the study presented in (Yang et al., 2007), pixel gray level values of low resolution images as features can achieve a high face recognition rate. Such facial features are better suited for re-identification scenarios and we leverage this feature.

The model proposed in this paper is a part-based spatio-temporal appearance model that combines color and facial features. A histogram of oriented gradients (HOG) (Dalal and Triggs, 2005) based part detector (Felzenszwalb et al., 2010) is used to extract four stable human body parts: head, left torso, right torso and upper legs. The part based description of an individual implicitly incorporates the spatial relationships between different body parts into the model. The appearance of the torso and legs is characterized by two color features; color histograms and representative color descriptors. In a multi-camera scenario typically a video or sequence of frames of a particular individual are captured. Our model leverages the temporal nature of the data by meaningfully combining these color features over time. The pixel gray level values extracted from the head region images are used as facial features. Since the head region images could be blurred or no face could be present, facial feature extraction is not always possible.
Low level image cues are used to select usable face images. If usable face images are present, the person’s face model is built by using the facial features. As a result of the selection process, the face images are no longer temporally adjacent, even if they come from the same video. Thus, multiple models are learnt for the same person’s face to keep model inaccuracies due to feature misalignment to a minimum. Depending on the presence of usable face images our model can decide to include the facial features or exclude them, this giving our model flexibility in an unconstrained environment. Our person model is distinct from the previously discussed models in three ways. First, we incorporate facial features. Second, our model generation does not require a training phase for feature selection or codebook learning. Thirdly, the model implicitly incorporates spatial information through body-parts and the temporal information using multiple frames. Hence the proposed model is truly a spatio-temporal representation.

3. Approach

Re-identification can be thought of as a two step process, extracting corresponding regions between images and matching them. A part-based person representation is used that implicitly encodes the structural information as well as determines corresponding parts from the images to be matched. These body parts are in turn described by visual features that are used to build the person model, which is then used to establish a match between corresponding parts.

3.1. Spatio-temporal appearance model generation

The spatial distribution of the appearance is captured by the body parts. The features extracted from each body part are combined temporally into a model that describes the appearance along the spatial as well as temporal dimensions.

**Body Part Extraction.** The appearance of a person is characterized by stable body parts that can be well described. These body parts do not necessarily align with anatomical body parts thus imparting partial pose invariance to the model since human body pose variations do not result in drastic changes in the detected parts. These body parts were extracted using the model proposed in (Felzenszwalb et al., 2008), which models the human body as a collection of parts arranged in a deformable configuration. We used the 6-part person model trained on the VOC 2008 pedestrian dataset (Everingham et al., 2008). Of the six parts we retain only four stable parts: head, left torso, right torso and the upper legs, since they encapsulate the area of the body that provides maximum distinguishing appearance information. The model is based on color features extracted from the torso and the legs and facial features extracted from the head region.

### 3.1.1. Color features based model

Color is the most expressive and powerful cue for object recognition and is leveraged in our model to characterize appearance. We extract two different color descriptors for each body part except the head. Fig. 2 shows the color feature extraction and spatio-temporal model generation pipeline.

**Active Color Model (ACM).** A 2D color histogram based on the H and S color channels of the HSV color space is used to characterize the chromatic content of each body part as it provides a good balance between photometric invariance and discriminative power. Each channel is quantized into five bins, thus we have a $5 \times 5$ element HS histogram. In order to capture the appearance variations of the chromatic content over time, we build a probabilistic model following the idea of Active Appearance Models (Cootes et al., 1998) for each body part based on the underlying color histogram. The sequence of frames of a person are used to extract the 2D color histograms used to build the ACM. The active color model is given by $g = \mu + A g \cdot b_g$ and it captures variations in the 2D histograms across the sequence of frames. Here, $g$ is the mean 2D histogram and $A g$ is the matrix describing the modes of variations in the color histograms within the sequence of frames. Vector $b_g$ is the parameter set of the ACM. In our experiments, we only use columns of $A g$ that retain 75% of the variations. This not only helps in capturing the maximum variations but also helps to eliminate redundant information and outliers.

**Representative Meta Colors Model (RMC).** Every part is also characterized using a set of representative colors extracted by fitting finite mixture models to the color vector using the method proposed in (Figueiredo and Jain, 2000). The clustering is done in the HSV space but the final clusters are represented by RGB triplet. Each representative color cluster is described using the average color. The representative colors descriptor (RCD) is defined as,
3.1.2. Facial features based model

**Face Image Selection.** The body part detector gives the head region of a person and in order to extract facial features only the head regions with faces are retained. All the images are converted to gray scale and a two step selection process is used to retain usable face images. First, a threshold ($\tau_1$) on RMS contrast is employed to reject incorrectly detected head region images as well as low contrast faces. The RMS contrast is computed as per Eq. (1),

$$\text{RMSContrast} = \sqrt{\frac{1}{w \cdot h} \sum_{i \in w} \sum_{j \in h} (I_{ij} - \bar{I})^2}$$

where, $I_{ij}$ is the pixel intensity extracted from the face region image of size $w \times h$, $\bar{I}$ is the mean intensity of the image pixels. In the next step, the retained images are then subjected to canny edge detection to detect prominent edges. If more than half the number of pixels in the ellipse fitted region are above a threshold ($\tau_2$) then the head image is retained and used to extract the facial features. If $n_{total}$ is the original head region images, then the number of images retained after the selection process is $n_{ret}$ and these images are not all temporally adjacent. If $n_{ret} < 2$, then the facial model cannot be generated. In our experiments, $\tau_1 = 0.04$ and $\tau_2 = 0.6$.

**Facial Model Generation.** We use the selected images to extract the facial features. All the head region images are resized to a fixed dimension of $24 \times 20$. The face region images are vectorized by stacking its columns, these vectors are used as the facial features. The vector length is thus $m = w \cdot h = 480$. Thus, the facial feature based model is simply a matrix $F = [v_1, ..., v_n]$, where $v_n$ is the column corresponding to the $k$th face region. The matrix $F$ will be of size $m \times n$, where $n$ is the number of selected face images.

$$F_{m,n} = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,n} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m,1} & f_{m,2} & \cdots & f_{m,n} \end{bmatrix}$$

To minimize the errors in the model the underlying features should be as aligned as possible. We make an assumption that among the selected face images temporally adjacent frames do not change drastically and are reasonably aligned. Thus, within $F$, $F_1$ is a submatrix if size $m \times n_1$, such that the columns of $F_1$ are temporally consecutive face images. The submatrix is treated as a sub class, i.e. even if all columns from $F$ belong to the same subject, the submatrices within $F$ are treated as visually different instances of the same subject. $F$ is referred to as the facial feature model (FFM). Fig. 4 demonstrates the face model generation process.

3.2. Multiple person re-identification by matching

Multiple person re-identification is treated as a rectangular assignment problem, where the cost matrix is populated using the matching cost between a gallery subject model $G$ and probe subject set $P$. The gallery subject model is compared to $N$ number of randomly selected frames of the probe subject. The model matching cost is simply the minimum cost among all the probe frames. This matching cost is computed using Eq. (2),

$$\text{dist}(G, P) = w_{\text{color}} \cdot d_{\text{color}}(G, P) + w_{\text{face}} \cdot d_{\text{face}}(G, P)$$

where, $d_{\text{color}}$ is the color model matching and $d_{\text{face}}$ is the facial feature model matching cost. If usable face region images are available for both gallery and probe subject then, $w_{\text{color}} = 0.7$ and $w_{\text{face}} = 0.3$ else $w_{\text{color}} = 1$ and $w_{\text{face}} = 0$. This means that in the absence of usable face region images re-identification is established using only color features.

The color model matching cost is computed as per Eq. (3),

$$d_{\text{color}}(G, P) = w_{\text{ACM}} \cdot d_{\text{ACM}}(G, P) + w_{\text{RMC}} \cdot d_{\text{RMC}}(G, P)$$

where, $d_{\text{ACM}}(G, P) = \|g - g^*\|$ is the reconstruction error obtained by projecting the probe frame histogram $g$ into the ACM space of the gallery model $g - g + A_{g} \cdot b_{g}$. $g_m = g + A_{g} \cdot b_{g}$ where $b_{g} = A_{g}^{-1}(g - g)$, $d_{\text{RMC}}(G, P)$ is the cost based on the matching of the gallery RMC model to the RCD extracted from the probe frame. The gallery RMC and the probe RCD are both treated as signatures and the matching is...
treated as a transportation problem. The matching cost between the gallery RMC and probe RCD is calculated using the Earth Mover’s distance (EMD) (Rubner et al., 2000). \(d_{\text{color}}\) is computed for each body part and is of the form \(d_{\text{color}} = \sqrt{\sum_{p=1}^{N_{bp}} \left( d_{\text{color, } p} \right)^2} \), where, \(N_{bp}\) is the total number of body parts, in our case \(N_{bp} = 3\). And \(w_{\text{ACM}} = w_{\text{RMC}} = 0.5\) which means that both sets of color features contribute equally to the matching cost.

The facial feature matching is established using sparse representation in the context of face recognition (Wright et al., 2009). The underlying implication is that given a dictionary matrix built using labeled training images of several subjects, a test image of a subject is the linear combination of the training images of only the same subject from the dictionary. In our case, a given gallery FFM contains all usable images of the gallery subject but the images are arranged into subsets depending on the temporal adjacency. Given a gallery FFM \(F_G\), if the probe ID is same as the gallery ID, the probe image will lie approximately in a linear span defined by only a subset of the images that constitute the FFM. This implies that given a probe face image \(f_P\), it can be expressed as \(f_P = F_G \cdot \alpha\) and the intent is to find the sparsest \(\alpha\) that generated \(f_P\) in \(F_G\). Thus among all possible solutions of \(\alpha\), we want the sparsest. This amounts to solving the following \(l_1\)-minimization:

\[
\hat{\alpha} = \arg \min \| \alpha \|_1 \quad \text{s.t. } f_P = F_G \cdot \hat{\alpha}
\]

This optimization is solved using linear programming that leverages augmented Lagrange multipliers (Yang and Zhang, 2009). Thus, \(d_{\text{face}}\) is given by Eq. (5) and is an estimate of how well \(\hat{\alpha}\) reproduces \(f_P\).

\[
d_{\text{face}}(G, P) = \| f_P - F_G \cdot \hat{\alpha} \|_2
\]

The cost matrix \(C\) is used for assignment, which is populated using \(c_{ij} = \text{dist}(G_i, P)\). Here, \(i = 1, \ldots, N_p\), \(j = 1, \ldots, N_G\), and \(N_p\) is number of probe subjects and \(N_G\) number of gallery subjects. Since all the probe objects come from the same video we know their IDs are distinct so a one-to-one assignment between the probe and gallery sets is needed from multiple person re-identification. The combinatorial optimization is solved using the Munkres algorithm (Munkres, 1957). Given a cost matrix \(C\), the one-to-one assignment is such that the objective function \(\sum_{i=1}^{N_G} \sum_{j=1}^{N_p} C(i, j)x_{ij}\) is minimized. Here, \(x_{ij}\) represents assignment of element \(G_i\) of the gallery set to element \(P_j\) of the probe set, taking value 1 if assignment is done and 0 otherwise.

4. Experiments and results

For a given camera pair, the gallery model for each subject is built using all the available frames of a subject. Only \(N\) randomly selected frames from a probe sequence are used to calculate the matching cost. In all experiments performed, we chose \(N = 10\) and results presented are based on analysis of 50 independent trials.

4.1. Dataset

In order to obtain real world surveillance data, we setup a camera network consisting of nine cameras in and around a building. The camera network has cameras placed on the first floor and fifth floor of the building. The re-identification data consists of 40 subjects out of which 19 are seen in multiple cameras. The data is split into three scenarios based on the difference in environments between the cameras in a pair on which re-identification is to be established. The scenarios are Outdoor–Outdoor, Indoor–Indoor and Outdoor–Indoor.

4.2. Closed set experiment results

Only a subset of the probe set or closed probe set; i.e. an intersection between the probe and gallery set IDs is used to establish re-identification. This is a closed set experiment, in the sense that it is only possible to have correct matches or mismatches. This experiment is intended to test the sensitivity of the proposed model. The closed set results are evaluated using matching accuracy, i.e. number of probe subjects matched correctly. In other words, it is the percentage of rank-1 correct matches. The effectiveness of our model was compared to the SDALF model proposed in (Farenzena et al., 2010).

Fig. 5 shows the closed set re-identification performance using two variants of our model, one based only on color features and the other using both the color and facial features. From the figure the value of incorporating facial features into the spatio-temporal model is clearly evident. In both the Outdoor–Outdoor and Outdoor–Indoor scenarios, we observe a significant improvement in the re-identification accuracy. In the Indoor–Indoor scenario the performance remains unchanged by addition of the facial features but does not have an adverse effect on the re-identification rate. This implies that even low resolution face regions with varying illumination and pose can contribute towards improving the discriminative ability of our model. In the Outdoor–Outdoor
scenario adding facial features causes the accuracy to increase from 62% to 80%, which is a considerable improvement. The overall accuracy increases to 83% from 75%. Both variants of our model outperform the SDALF model in all three scenarios. In all the three scenarios our color only model gives 75% accuracy compared to 70% using SDALF.

![Graph showing re-identification performance](image1)

**Fig. 5.** Closed set re-identification performance, the bar graph shows the results obtained by our color model, color and face model and SDALF (Farenzena et al., 2010).

![Graphs showing accuracy vs. FAR](image2)

**Fig. 6.** Open set results: Accuracy vs. FAR curves obtained using (a) optimal assignment, and (b) suboptimal assignment. The top, middle and bottom row are results obtained on the Outdoor–Outdoor, Indoor–Indoor and Outdoor–Indoor scenarios respectively.
4.3. Open set experiment results

The entire probe set is used for re-identification wherein all the subjects in the probe set might not be present in the gallery set. This implies that in addition to correct matches and mismatches we will also have false positives. In the case of re-identification, true positives (TPs) are the number of probe IDs that are correctly matched. Mismatches (MMs) are the number of probe IDs that are incorrectly matched to gallery, when that probe ID does exist in the gallery. False positives (FPs) are the number of probes IDs that are matched to the gallery when the probe ID does not exist in the gallery. This experiment is designed to test the model based false acceptance reduction criteria and a secondary objective is to test the generality of the model.

In order to reduce the false acceptance, a threshold is imposed on the cost matrix $C$ and then multiple person re-identification is established. The open set results are presented in terms of Accuracy vs. false acceptance rate (FAR) curves. The accuracy and FAR are defined as $\text{Accuracy} = \frac{TP}{TP + FP}$ and $\text{FAR} = \frac{MM}{N_p}$, respectively, where $N_p$ denotes the total number of probe subjects. The curves are obtained by varying a threshold imposed on the matching cost during the computation of the cost matrix. The threshold is varied from 0 to 0.9 in increments of 0.05. Two different assignment techniques are employed; optimal, which is the Munkers algorithm, and the sub-optimal assignment. The sub-optimal assignment technique is usually used when the cost matrices have many forbidden assignments. The assignment is suboptimal in the sense that the overall assignment cost is not the minimum possible value. In case of re-identification, the gallery is ever increasing and most likely the intersection between gallery and probe is small compared to the size of the gallery. Thus, the possibility of incorrect assignments increases and suboptimal assignment technique could be better suited for such cases.

Fig. 6 shows that in all three scenarios it is possible to find a threshold that yields the best possible trade off between accuracy and FAR. In the Outdoor–Outdoor scenario a threshold of 0.5 yields the best possible Accuracy/FAR ratio of 60%/33% using the color and facial features model and optimal assignment technique. In the Indoor–Indoor and Outdoor–Indoor scenarios as well, color and facial features model gets the best Accuracy/FAR ratio possible. The exact same observations are applicable to the suboptimal assignment technique as well. Except in the case of the Outdoor–Indoor scenario under suboptimal assignment, our color only model is able to find a threshold that gives better Accuracy/FAR ratio compared to SDALF as well. In the exception case, SDALF finds a threshold that gives nominally better Accuracy/FAR ratio compared to our color only model. Thus, overall in both variants of our model it is possible to find a threshold that gives better Accuracy/FAR ratio than SDALF which implies that our model can generalize well to handle open set re-identification.

In order to compare the open set performance of the three models we use the discrimination measure as an indicator of model’s accuracy and false rejection ability. The discrimination measure is simply the area between the curve and the non-discrimination line. Table 1 shows that in all three scenarios the color and facial features model has improved discrimination over the other two models evaluated. In the Outdoor–Outdoor scenario and Indoor–Indoor scenario using optimal assignment technique, and Indoor–Indoor scenario using suboptimal assignment technique SDALF has a higher discrimination than our color only model. We surmise this is because SDALF also has a texture component in its model that is absent in ours. Overall, all the models suffer a slight drop in discrimination using the suboptimal assignment technique. Thus, we can deduce that Munkres algorithm is a better fit for re-identification assignments.

In general, closed and open set experiments both suggest that using color and facial features has a distinct advantage in the re-identification over using only color or even combined color and texture features. From the shape of these curves we can deduce that a global threshold on the matching cost is a crude yet sensible strategy for minimizing the false acceptance rate. As the threshold increases the accuracy values start increasing as well with a less drastic increase in the FAR. The most important conclusion that we can draw from these experiments is that the proposed model can be used for false acceptance reduction during re-identification. Specifically, open set experiments can be used to select the parameters of false acceptance reduction criteria, in our case a suitable threshold.

5. Conclusions

In this paper, we have presented a spatio-temporal model based on color and facial features that captures complementary aspects of a person’s appearance. A face model based on useful facial features was exploited using sparse representation and applied towards a multi-feature model. A low level image feature based face region selection module was designed to select usable face region images. A strategy for multiple person re-identification based on the rectangular assignment problem was presented. Experiments were performed based on data collected from a wide area surveillance system comprising of nine cameras and results presented indicating that the proposed model has the necessary balance of discriminative ability and photometric invariance.

We would also like to point out that, to the best of our knowledge, this is the first study that offers strategies for principled spatio-temporal appearance model generation, multiple person re-identification, and appearance model based criterion for reduction in false acceptance rates. Specifically, in this paper we have addressed the relatively unexplored problem of multiple person re-identification as a true open set matching problem. Future work will focus on enhancing the model’s discriminative and invariant properties. Overall, exploring better solutions to the multiple person re-identification and false acceptance reduction will lead to improvements in overall performance of vision applications leveraging multiple cameras with non-overlapping field-of-views.
Acknowledgements

This work was supported in part by the US Department of Justice 2009-MU-MU-K004. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of our sponsors.

References


