https://doi.org/10.1109/ICIP.2018.8451190

DEEP RESIDUAL NETWORK WITH SUBCLASS DISCRIMINANT ANALYSIS FOR CROWD BEHAVIOR RECOGNITION

Bappaditya Mandal[1]*, Jiri Fajtl*[1]*, Vasileios Argyriou*[1]*, Dorothy Monekosso*[2] *and Paolo Remagnino*[1]

[1]Kingston University London, Surrey, United Kingdom
Email: {B.Mandal, k1410371, Vasileios.Argyriou, P.Remagnino}@kingston.ac.uk
[2]Leeds Beckett University, West Yorkshire, United Kingdom
Email: D.N.Monekosso@leedsbeckett.ac.uk

ABSTRACT

In this work, we extract rich representations of crowd behavior from video using a fine-tuned deep convolutional neural residual network. Using spatial partitioning trees we create subclasses within the feature maps from each of the crowd behavior attributes (classes). Features from these subclasses are then regularized using an eigen modeling scheme. This enables to model the variance appearing from the intra-subclass information. Low dimensional discriminative features are extracted after using the total subclass scatter information. Dynamic time warping is used on the cosine distance measure to find the similarity measure between videos. A 1-nearest neighbor (NN) classifier is used to find the respective crowd behavior attribute classes from the normal videos. Experimental results on large crowd behavior video database show the superior performance of our proposed framework as compared to the baseline and current state-of-the-art methodologies for the crowd behavior recognition task.

Index Terms— Crowd behavior recognition, feature extraction, discriminant analysis, residual network.

1. INTRODUCTION

The use of crowd analysis and management with video data is common practice at public events such as concerts, sport matches, event celebrations and protests, public gatherings at stations. A large number of people die every year in very crowded environments, such as the Mumbai railway station 2017 stampede which killed 22 people and injured 30 people [1] and the New Year's Eve 2015 celebration in Shanghai, where a stampede tragically left 36 people dead and nearly 50 others injured [2]. For human observers, it is extremely difficult to monitor a very large number of individuals, their behaviors and activities from a large topology of cameras. The affected areas are generally highly congested urban areas and extracting useful behavior pattern information has become of paramount importance for public security, safety, crowd management, providing timely critical decisions and support.

According to published reviews [3, 4], a large amount of work exists on tracking, recognizing and understanding behavior of people in videos. Existing research is mainly focused on sparse and mostly staged scenes. However, relatively little effort has been devoted to reliable classification and understanding of human activities in real and very crowded scenes. In heavily crowded scenes, often the detected targets (people and objects of interest) are very small, and the recognition task is very challenging, for instance characterizing people interactions. In general, researchers have proposed two ways of analyzing behavior in such complex scenes. Firstly, considering the crowd and scene targets as a whole, where individual targets such as objects, places, scenes, their actions or interactions are not identified or classified individually, rather they are processed based on their whole appearance [5, 6]. It is often advantageous and simpler to understand the crowd behavior without knowing the actions of the individuals. Secondly, object based approach, where each individuals (human and/or objects) are detected and segmented to perform motion and/or behavior analysis [7, 8]. This kind of complex segmenting and tracking individuals in crowded videos is a very challenging task. In our work we use the former, where individuals are not segmented or tracked, but crowd behavior pattern is perceived holistically, they are fine-tuned with a deeply learned model, features are extracted and subclass regularized discriminative analysis is performed so as to classify crowd behavior attributes.

2. FEATURE EXTRACTION IN CROWD VIDEOS

Recent research has shown deep residual neural networks perform very well on image recognition tasks [9]. Unlike classical sequential neural network architectures, such as VGG [10], a residual network consists of *network-in-network* modules. This kind of architectures solves both the vanishing gradient and over-fitting problems, with a number of layers usually larger than 100. It also provides a clear path for gradients to back propagate to early layers of the network, thus making the learning process faster [9]. Since these residual networks

are trained on large number of images (typically millions), their deeply learned weights are transferable to other problems, such as action recognition [11] and video summarization [12]. This has motivated the use of pre-trained deeply learned residual models for crowd behavior recognition.

In our work, we first fine tune the network using crowd behavior videos and then extract rich representations of the pattern of specific crowd behavior. Spatial partitioning trees are used to create subclasses so as to facilitate extracting intra-class variance information. Features projected from the intra-subclass variances are then scaled using an eigen feature regularization scheme. Finally low-dimensional discriminative features are extracted using total class scatter matrix. Our proposed framework is shown in Fig. 1.

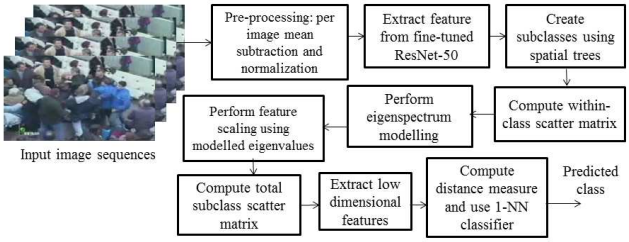

Fig. 1. Our proposed framework.

2.1. Deep Learning Features and Fine-tuning

While training the network using pre-trained ResNet-50 model, all images are resized to 224 $\times$ 224, normalized each channel-wise to zero mean and unit standard deviation. ResNet-50 pretrained network is fine-tuned using crowd behavior recognition database by retraining the whole network. Training only the higher-level layers helps in extraction of features specific to the present dataset, avoids overfitting and increases the robustness, since the number of images are few as compared to ImageNet trained with millions of images. Let $I(width, height, channels)$ represent the resized and normalized input image of size $width \times height$ and number of channels as depth, in our case it is 3. $C(a, b, \upsilon)$ represents the convolutional layer, where a is the filter size, b is the strides and υ is the number of filter banks. $P(\varrho, \iota)$ represents the pooling layer, ϱ is the number of strides and ι is the size of window for subsampling. Each convolutional layer is followed by a batch normalization layer and RELU as a non-linearity function. The summations at the end of each residual unit are followed by a ReLU unit. Each repetitive residual unit is presented inside R. $F(E)$ denotes the fully connected layer where E is the number of neurons. Thus, the fine-tuned ResNet-50 is represented as (1).

The length of $F(E)$ depends on the number of categories to classify, E is the number of classes. P^* refers to average pooling rather than max pooling as used everywhere else. The softmax function or the normalized exponential function is described as: $S(F)_j = \frac{exp^{F_j}}{\sum_{k=1}^{E} exp^{F_k}}, for\ j = 1, 2, ..., E.$ Using the above equations fine-tuning training is performed and a new model is obtained. Rich crowd behavioral pattern representations (features) are extracted from this new deeply learned model, they are partitioned to form subclasses and subsequently our discriminative analysis is performed.

$$
\begin{aligned}
\Theta_R = {} & I(224, 224, 3) \rightarrow C(7, 2, 64) \rightarrow P(2, 3) \rightarrow \\
& 3 \times R(C(1, 1, 64) \rightarrow C(3, 1, 64) \rightarrow C(1, 1, 256)) \rightarrow \\
& R(C(1, 2, 128) \rightarrow C(3, 2, 128) \rightarrow C(1, 2, 512)) \rightarrow \\
& 3 \times R(C(1, 1, 128) \rightarrow C(3, 1, 128) \rightarrow C(1, 1, 512)) \rightarrow \\
& R(C(1, 2, 256) \rightarrow C(3, 2, 256) \rightarrow C(1, 2, 1024)) \rightarrow \\
& 5 \times R(C(1, 1, 256) \rightarrow C(3, 1, 256) \rightarrow C(1, 1, 1024)) \rightarrow \\
& R(C(1, 2, 512) \rightarrow C(3, 2, 512) \rightarrow C(1, 2, 2048)) \rightarrow \\
& 2 \times R(C(1, 1, 512) \rightarrow C(3, 1, 512) \rightarrow C(1, 1, 2048)) \rightarrow \\
& P^*(1, 7) \rightarrow F(e) \rightarrow Softmax
\end{aligned}
\tag{1}
$$

2.2. Subclass Partitioning and Discriminant Analysis

Creating subclasses is apportioning a crowd behavior attribute (class) into multiple partitions. Among spatial partition trees [13, 14], popular random projection (RP) and principal component analysis (PCA) trees are used to partition each crowd behavior class into subclasses. RP trees are built by recursive binary splits, it adapts to intrinsic low dimensional structure without having to explicitly learn crowd behavior pattern structure. In PCA tree, the partition axis is obtained by computing the principal eigenvector of the covariance matrix of the crowd image data. Since crowd image appear very differently under various contexts this kind of partitioning would be advantageous for data that are heterogeneously distributed in all dimensions. In our experiments, we use the implementations of spatial partitioning trees by Freund *et al.* [13], with their default parameters of maximum depth up to eight layers and no overlap in samples splitting.

After the subclass creation, crucial intra-class variance information is learned by computing the within-subclass scatter matrix and optimization is performed using Fisher criterion [15]. This involves between-class (S_b) and within-class ($S_w = \frac{1}{n}\sum_{i=1}^{E}\sum_{j=1}^{n_i}(x_{ij} - \mu_i)(x_{ij} - \mu_i)^T$) scatter matrices, where E is the number of classes or crowd behavior attributes, μ_i is the sample mean of class i, μ is the global mean, $x_{ij} \in \mathbb{R}^l$, where l = feature dimension obtained before the final pooling layer from (1), is the j^{th} sample of class i, n_i is the number of samples in i^{th} class and $n = \sum_{i=1}^{E} n_i$ is the total number of samples. Traditional linear discriminant analysis (LDA) assumes that the class distributions are homoscedastic [16], which is rarely true in practice for complex crowd behavior analysis. We assume that there exist subclass homoscedastic partitions of the data and model each class as mixtures of Gaussians subclasses, whose Fisher objective function is defined as $J(\Psi) = \frac{tr(\Psi^T S_{bs} \Psi)}{tr(\Psi^T S_{ws} \Psi)}$, where tr represents trace of a matrix, Ψ denotes a transformation matrix, S_{bs} is the between-subclass scatter matrix and S_{ws} is the

within-subclass scatter matrix defined as

$$S_{ws} = \sum_{i=1}^{E} p_i \sum_{j=1}^{H_i} \frac{q_{H_i}}{G_{ij}} \sum_{k=1}^{G_{ij}} (x_{ijk} - \mu_{ij})(x_{ijk} - \mu_{ij})^T. \quad (2)$$

However, when the number of classes are low and training data are small, variance from the total scatter matrix outperforms the between-class scatter matrix [17]. So in this work we propose to use Fisher objective function as $J(\Psi) = \frac{tr(\Psi^T S_{ts} \Psi)}{tr(\Psi^T S_{ws} \Psi)}$, where S_{ts} is the total subclass scatter matrix, H_i denotes the number of subclasses of the i^{th} class and G_{ij} denotes the number of samples in j^{th} subclass of i^{th} class. $x_{ijk} \in \mathbb{R}^l$ is the k^{th} image vector in j^{th} subclass of i^{th} class. $\mu_{ij} = \frac{1}{G_{ij}} \sum_{k=1}^{G_{ij}} x_{ijk}$ is the sample mean of j^{th} subclass of the i^{th} class. p_i and q_{H_i} are the estimated prior probabilities. If we assume that each class and subclasses have equal prior probabilities then $p_i = \frac{1}{E}$ and $q_{H_i} = \frac{1}{H_i}$.

2.3. Regularization of the Subclasses

By forming the subclasses using spatial partition trees, we are able to capture robust variance information more closely in the holistic analysis of the crowd behavior recognition. We compute the eigenvectors $\Psi^{ws} = \{\psi_1^{ws}, \ldots, \psi_l^{ws}\}$ corresponding to the eigenvalues $\Lambda^{ws} = \{\lambda_1^{ws}, \ldots, \lambda_l^{ws}\}$ of S_{ws} described by (2), where the eigenvalues are sorted in descending order. The whitened eigenvector matrix $\overline{\Psi}^{ws} = \{\psi_1^{ws}/\tau_1^{ws}, \ldots, \psi_l^{ws}/\tau_l^{ws}\}$, $\tau_k^{ws} = \sqrt{\lambda_k^{ws}}$, is used to project the image vector x_{ij} before constructing the total subclass scatter matrix. This is equivalent to image vector x_{ij} first transformed by the eigenvector $y_{ij} = \Psi^{wsT} x_{ij}$, and then multiplied by a scaling function $\omega_k^{ws} = 1/\tau_k^{ws}$ (whitening process). Truncating dimensions is equivalent to set $\omega_k^{ws} = 0$ for these dimensions as done in Fisherface and many other variants of LDA [18]. The scaling function is thus

$$\omega_k^{ws} = \begin{cases} 1/\sqrt{\lambda_k^{ws}}, & k \leq r_{ws} \\ 0, & r_{ws} < k \leq l \end{cases}, \quad (3)$$

where $r_{ws} \leq min(l, \sum_{i=1}^{E} \sum_{j=1}^{H_i} (G_{ij} - 1))$, is the rank of S_{ws}. Similar scaling function (ω_k^w) with $r_w \leq min(l, n - E)$ are also applicable for S_w scatter matrix.

There are two problems associated with the inverses of τ_k^{ws} and τ_k^w. Firstly, the eigenvectors corresponding to the zero eigenvalues are discarded as the features in the null space are weighted by a constant zero. This leads to the loss of important discriminative information that lies in the null space [18, 19, 20, 21]. Second, when the inverse of the square of the eigenvalues are used to scale the respective eigenvectors, features get undue weighage, noise get amplified and tend to over-fit the training samples. We use a median operator and its parameter similar to that used for face recognition in [22, 23] to find the pivotal point m for decreasing the decay of the eigenspectrum. We use the function form $1/f$, similar to [24, 22], to estimate the eigenspectrum as $\tilde{\lambda}_k^{ws} = \frac{\alpha}{k+\beta}, \quad 1 \leq k \leq r_{ws}$, where α and β are two constants, used to model the real eigenspectrum in the initial portion. We determine α and β by letting $\tilde{\lambda}_1^{ws} = \lambda_1^{ws}$ and $\tilde{\lambda}_m^{ws} = \lambda_m^{ws}$, which yields $\alpha = \frac{\lambda_1^{ws} \lambda_m^{ws} (m-1)}{\lambda_1^{ws} - \lambda_m^{ws}}$, $\beta = \frac{m\lambda_m^{ws} - \lambda_1^{ws}}{\lambda_1^{ws} - \lambda_m^{ws}}$. These are used to generate the model eigenvalues which matches closely as that of the real eigenvalues.

2.4. Feature Selection and Dimensionality Reduction

The noise component is small as compared to crowd behavior pattern components in the principal space but it is dominating in an unstable region. Thus, the estimated eigenspectrum $\tilde{\lambda}_k^{ws}$ is given by

$$\tilde{\lambda}_k^{ws} = \begin{cases} \lambda_k^{ws}, & k < m \\ \frac{\alpha}{k+\beta}, & m \leq k \leq r_{ws} \\ \frac{\alpha}{r_{ws}+1+\beta}, & r_{ws} < k \leq l \end{cases} \quad (4)$$

The feature scaling function is then $\tilde{\omega}_k^{ws} = \frac{1}{\sqrt{\tilde{\lambda}_k^{ws}}}, \quad k = 1, 2, ..., l$. Using this scaling function and the eigenvectors ψ_k^{ws}, training data are transformed to $\tilde{y}_{ij} = \tilde{\mathbf{\Psi}}_l^{wsT} x_{ij}$, where $\tilde{\mathbf{\Psi}}_l^{ws} = [\tilde{\omega}_k^{ws} \psi_k^{ws}]_{k=1}^{l}$. New total subclass scatter matrices are formed by vectors $\tilde{y}_{ij}$ of the transformed training data as

$$\tilde{S}_{ts} = \sum_{i=1}^{E} \frac{p_i}{n_i} \sum_{j=1}^{n_i} (\tilde{y}_{ij} - \tilde{\mu})(\tilde{y}_{ij} - \tilde{\mu})^T, \quad (5)$$

where $\tilde{\mu}_{ij} = \frac{1}{G_{ij}} \sum_{k=1}^{G_{ij}} \tilde{y}_{ijk}$ and $\tilde{\mu} = \frac{1}{E} \sum_{i=1}^{E} \tilde{\mu}_i$, such that $\tilde{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \tilde{y}_{ij}$. In this work, we employ the total scatter matrix $\tilde{S}_{ts}$ of the regularized training data to extract the discriminative features because of its greater noise tolerance as compared to $\tilde{S}_{bs}$. The transformed features $\tilde{y}_{ij}$ will be decorrelated for $\tilde{S}_{ts}$ by solving the eigenvalue problem. Selecting the eigenvectors with the d largest eigenvalues, $\tilde{\mathbf{\Psi}}_d^{ts} = [\tilde{\psi}_k^{ts}]_{k=1}^{d}$, the proposed feature scaling and extraction matrix is given by $\mathbf{U} = \tilde{\mathbf{\Psi}}_l^{ws} \tilde{\mathbf{\Psi}}_d^{ts}$, which transforms a crowd behavior image vector x, $x \in \mathbb{R}^l$, into a feature vector z, $z \in \mathbb{R}^d$, by $z = \mathbf{U}^T x$.

2.5. Video Matching using Dynamic Time Warping

Crowd behavior analysis has inherent varying space-temporal structure. Different crowd groups might show the same behavior (e.g. a protest) differently and even the same group is not ever able to reproduce the same behavior exactly. So, to compare two behavior events of different lengths we use dynamic time warping (DTW) [25], which performs a time alignment and normalization by computing a temporal transformation allowing two behaviors to be matched. In the experiments of this work, Cosine distance measure and the first nearest neighborhood classifier (1-NNK) are applied along with DTW to test the proposed approach for crowd behavior recognition.

3. EXPERIMENTAL RESULTS

We tested our approach on the world's largest crowd behavior recognition database [8], comprising of 10,000 videos from 8,257 scenes. This database is constructed to challenge on the *where, who and why* questions (abbreviated WWW Crowd Database). The WWW has 94 crowd-related annotated attributes, such as stadium, concert, stage, fight, mob, parade, and others, to describe each video in the database. We selected a few normal crowd videos (like waking, skating, graduation, and others) and 4 violent crowd behavior videos, such as fight, protest, mob and protester from this large database. Following the conventional protocol [8], the WWW crowd database is randomly partitioned into training, validation and test sets in the ratio $7:1:2$, videos are converted to images at 25 frames per second. This provides a total of 219,094 images, the distribution for each of the selected attributes is shown in Table 1. In all our experiments we validate our proposed method with the existing ones using the same corresponding database and protocol.

Table 1. Selected attributes and their images from the WWW crowd database.

Attributes	Normal	Fight	Mob	Protest	Protester
# Images	15,631	14,059	14,609	87,241	87,554

We follow the rule described in [16] that every class is partitioned by the same number of subclasses h (equally balanced with $h = 5$), such that $H_i = h, \forall i$. We test our approach using both PCA and RP decision trees [14], as both have similar performance we report only the results with PCA decision trees. Fig. 2 shows the recognition rate (%) versus the number of features used for classification for various crowd behavior recognition. We have also implemented the baseline approach, which uses the features from the ResNet-50 model fine-tuned using the images from the WWW crowd database. It is evident that our proposed learning framework outperforms the baseline method for all the violent activities. For protest and protester, the recognition rate improvement is small, but for fight and mob our approach outperforms the baseline method significantly.

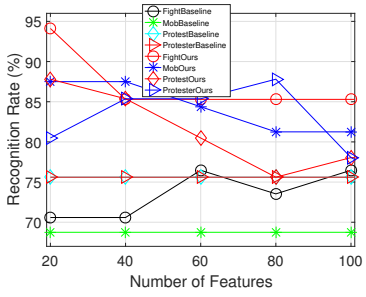

Fig. 2. Recognition rate (%) on WWW crowd database.

The average recognition rates (%) of all the violent behaviors are shown in Fig. 3. It can be clearly seen that our proposed method outperforms the baseline method for both the Cosine (Cos) and Euclidean (Eud) distance measures. The performance gain is higher for small number of features.

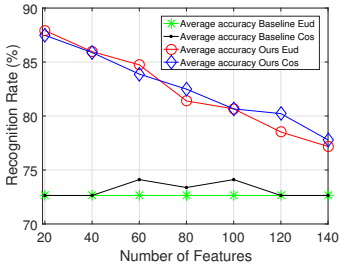

Fig. 3. Average recognition rate (%) on WWW crowd database.

We also compare our results with the current state-of-the-art algorithms and baseline method on this database. Table 2 shows the AUCs of the various reported results as compared to ours. It can be easily seen that our proposed approach outperforms the baseline and the recently proposed ResnetCrowd [6] for both the single task and multi-task crowd behavior recognition significantly. Although Shao *et al.* [8] used several millions of images for their deep learning attributes using computationally expensive motion channels during training, our approach outperforms their methodology for crowd behavior recognition task as shown in Table 2.

Table 2. Crowd behavior recognition AUCs on WWW crowd database. ResnetCrowd^1 and ResnetCrowd^2 represent single task and multi-task respectively.

Methods	Fight	Mob	Protest	Protester	Average
Baseline	0.87	0.82	0.83	0.89	0.85
Shao *et al.* [8]	0.93	0.91	0.95	0.97	0.94
ResnetCrowd^1 [6]	0.62	0.68	–	–	0.65
ResnetCrowd^2 [6]	0.71	0.77	–	–	0.74
Our Proposed	0.95	0.94	0.96	0.96	0.95

4. CONCLUSIONS

This paper proposes a fine-tuned deep convolutional neural residual network framework that creates subclasses in the feature maps of each of the crowd behavior attribute classes using spatial partitioning trees. Eigen feature regularization using eigenmodel is used to weigh the features of the whole intra-subclass eigenspace of the crowd behavior videos. This has helped to model the variance appearing from the intra-subclass variance information. Low dimensional discriminative features are extracted using total subclass scatter matrix to represent the various crowd behavior videos. Dynamic time warping is used on the cosine distance measure to find the similarity measure between the two videos for crowd behavior recognition task. Experimental results on a large crowd behavior video database show the superiority of our proposed framework compared to the baseline and current state-of-the-art methodologies.

5. REFERENCES

[1] The Guardian, "Mumbai railway station stampede kills at least 22," "https://www.theguardian.com/world/2017/sep/29/stampede-mumbai-railway-station-rush/-hour", 2017.

[2] The Guardian, "Shanghai: dozens killed and injured in stampede at new year celebrations," "https://www.theguardian.com/world/2014/dec/31/shanghai-35-people-killed-42-injured/-new-year-crush", 2014.

[3] Teng Li, Huan Chang, Meng Wang, Bingbing Ni, Richang Hong, and Shuicheng Yan, "Crowded scene analysis: A survey," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 25, no. 3, pp. 367–386, 2015.

[4] Julio Cezar Silveira Jacques, Soraia Raupp Musse, and Cláudio Rosito Jung, "Crowd analysis using computer vision techniques," *IEEE Signal Process. Mag.*, vol. 27, no. 5, pp. 66–77, 2010.

[5] Berkan Solmaz, Brian E. Moore, and Mubarak Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems," *IEEE PAMI*, vol. 34, no. 10, pp. 2064–2070, 2012.

[6] Mark Marsden, Kevin McGuinness, Suzanne Little, and Noel E. O'Connor, "Resnetcrowd: A residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification," *CoRR*, vol. abs/1705.10698, 2017.

[7] Bolei Zhou, Xiaogang Wang, and Xiaoou Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in *CVPR*, 2012, pp. 2871–2878.

[8] Jing Shao, Kai Kang, Chen Change Loy, and Xiaogang Wang, "Deeply learned attributes for crowded scene understanding," in *IEEE CVPR*, 2015, pp. 4657–4666.

[9] Zhang X. Ren S. *et al.* He, K., "Deep residual learning for image recognition," in *IEEE CVPR*. IEEE, 2016, pp. 770–778.

[10] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman, "Deep face recognition," in *BMVC*, Sep 2015, pp. 41.1–41.12.

[11] Ahsan Iqbal, Alexander Richard, Hilde Kuehne, and Juergen Gall, "Recurrent residual learning for action recognition," in *Pattern Recognition - 39^{th} German Conference*, 2017, pp. 126–137.

[12] Ana Garcia del Molino, Bappaditya Mandal, Jie Lin, Joo-Hwee Lim, Vigneshwaran Subbaraju, and Vijay Chandrasekhar, "VC-I2R@ImageCLEF2017: Ensemble of deep learned features for lifelog video summarization," in CLEF *Conference and Labs of the Evaluation Forum, Dublin, Ireland*, Sep 2017.

[13] Yoav Freund, Sanjoy Dasgupta, Mayank Kabra, and Nakul Verma, "Learning the structure of manifolds using random projections.," in *NIPS*, 2007, vol. 7, p. 59.

[14] J. Wang, N. Wang, Y. Jia, J. Li, G. Zeng, H. Zha, and XS. Hua, "Trinary-projection trees for approximate nearest neighbor search," *IEEE PAMI*, vol. 36, no. 2, pp. 388–403, 2014.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, John Wiley and Sons, New York, 2001.

[16] M. Zhu and A. Martinez, "Subclass discriminant analysis," *IEEE PAMI*, vol. 28, no. 8, pp. 1274–1286, August 2006.

[17] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE PAMI*, vol. 23, no. 2, pp. 228–233, Febuary 2001.

[18] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE PAMI*, vol. 27, no. 1, pp. 4–13, January 2005.

[19] W. Liu, Y. Wang, S. Z. Li, and T. N. Tan, "Null space approach of fisher discriminant analysis for face recognition," in *ECCV*, 2004, pp. 32–44.

[20] B. Mandal, X. D. Jiang, and A. Kot, "Dimensionality reduction in subspace face recognition," in *IEEE ICICS*, Dec 2007, pp. 1–5.

[21] B. Mandal, Liyuan Li, V. Chandrasekhar, and Joo Hwee Lim, "Whole space subclass discriminant analysis for face recognition," in *IEEE ICIP*, Quebec city, Canada, Sep 2015, pp. 329–333.

[22] X. D. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE PAMI*, vol. 30, no. 3, pp. 383–394, March 2008.

[23] B. Mandal, X. D. Jiang, and A. Kot, "Verification of human faces using predicted eigenvalues," in 19^{th} *International Conference on Pattern Recognition (ICPR)*, Tempa, Florida, USA, Dec 2008, pp. 1–4.

[24] B. Moghaddam, "Principal manifolds and probabilistic subspace for visual recognition," *IEEE PAMI*, vol. 24, no. 6, pp. 780–788, June 2002.

[25] N. V. Boulgouris, K. Plataniotis, and D. Hatzinakos, "Gait recognition using dynamic time warping," in *IEEE Workshop on Multimedia Signal Processing*, Sept 2004, pp. 263–266.