Discriminative FaceTopics for Face Recognition via Latent Dirichlet Allocation

Tejas Indulal Dhamecha, Praneet Sharma, Richa Singh, and Mayank Vatsa
IIIT Delhi, India
{tejasd,praneet10061,rsingh,mayank}@iiitd.ac.in

Abstract

Latent Dirichlet Allocation is a widely used approach for topic modeling and it has been successfully applied in several information retrieval applications. In this paper, we introduce this modeling technique for face recognition, by making an analogy between the two domains. We utilize latent Dirichlet allocation to represent facial regions in terms of FaceTopics. Further, linear discriminant analysis is utilized to obtain discriminative FaceTopics which are more suitable for classification tasks. The performance of the proposed approach is evaluated on the CMU-MultiPIE dataset under illumination and expression variations. The evaluation on over more than 50k images shows the effectiveness of the proposed approach. Further, the proposed approach shows improved identification results on e-PRIP dataset for matching composite sketches to photos.

1. Introduction

Retrieving an identity from a database of face images is one of the key challenges in biometrics research. Several approaches are proposed to address different covariates such as pose, expression, and illumination [6]. While the research in face recognition has advanced over the past couple of decades, it still requires strengthening state-of-the-art results. Major efforts such as JANUS[1] program are underway to transform the research to the next level and researchers are not only focusing on improving the results but also developing enriched understanding of face representation. Among these attempts, *learning face representations* has received significant attraction.

It is well understood that face recognition is an interdisciplinary research area encompassing broader areas such as cognition, sensors, pattern recognition, machine learning, image processing, and computer vision as well as allied application areas such as information retrieval and human computer interactions. With proper underpinning and exploration, any transformational development in these areas can also be embraced in face recognition research. Among the allied domains, information retrieval (IR) domain has several similarities with face recognition. For example, generative models have been well explored in both the domains; however, various models such as probabilistic latent semantic analysis [17] and latent Dirichlet allocation [9] are explored in IR (or text-analytics) which can potentially be explored in biometrics (or face recognition).

Topic modeling is an interesting IR approach, where text documents are classified into the associated (possibly unknown) topics [14]. Latent Dirichlet Allocation (LaDiAl)[2] [9] is one such topic modeling approach, which provides a generative model of the document with topics as one of the latent variables. The applicability of LaDiAl in non-textual data, particularly images, has been a stimulating research direction amongst researchers [11, 31, 33]. Wang and Eric [33] present spatial latent Dirichlet allocation for image segmentation and clustering. The authors modify the LaDiAl model for image segmentation and classification tasks. Similarly, Sudderth *et al.* [31] utilize the Dirichlet process based model for representing scenes. Topic modeling based approaches have found their applicability in human action recognition as well [25, 34]. Since LaDiAl is a generative model for discrete data, using it for image data requires mapping image features to discrete visual words. Such mapping has been explored by researchers in various contexts [10, 19, 26, 31]. The objective function of these generative models is based on efficiently capturing the data generation process. Thus, features obtained from such generative models may not be effective for classification tasks. To fine-tune the features for classification tasks, we need to transform them into a usable form, while minimizing intra-class variations and maximizing inter-class variations. For scene classification, Bosch *et al.* [10] propose a generative model of scene images using probabilistic latent semantic analysis, over which discriminative learning is applied. Similarly, Bregonzio *et al.* [11] propose a discriminative learning on modified LaDiAl based approach for hu-

[1] http://www.iarpa.gov/index.php/research-programs/janus

[2] Since, the abbreviation LDA is often used for linear discriminant analysis in biometric community, we chose to use LaDiAl for representing Latent Dirichlet Allocation.

man action recognition. To make LaDiAl more suitable for classification tasks, Blei and McAuliffe [8] propose to incorporate supervised learning in the model. Rasiwasia *et al*. [28] present an extension to handle more complex class structures and demonstrate it's effectiveness for image classification.

In this paper, we step forward by exploring the utility of LaDiAl based topic modeling for face recognition. We utilize LaDiAl to find the *topic feature representation* pertinent to face images. These topic features are transformed into *discriminative topic feature representation* that makes the features more suitable for classification (face recognition) task. The key contributions of this paper can be summarized as:

- exploring LaDiAl for face recognition by making an analogy with text (IR) domain;
- developing LaDiAl topic modeling based generative-discriminative feature extraction approach to represent faces; and
- evaluating the proposed approach in presence of expression and/or illumination variations on ∼50,000 face images from the CMU-MultiPIE face dataset [15]. Further, effectiveness of the proposed algorithm is shown on e-PRIP composite sketch to photo matching database [16, 23].

2. Latent Dirichlet Allocation

Latent Dirichlet Allocation [9] is a widely used topic modeling technique which unravels the structure in a given set of documents. In other words, it can help identify, probabilistically, what all *topics* does a sample belong to. For example, given a set of newspaper articles, we can expect it to appropriately arrange all the articles according to their topics. The basic interpretation of LaDiAl is shown in Figure 1. LaDiAl represents documents $\mathbf{w}$ (sets of words) as a random mixture over unknown (latent) topics $\mathbf{z}$; and topics as the distribution over words. The document generation process is assumed to be governed by latent random variables $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$, $\mathbf{z}$, and $\boldsymbol{\beta}$ as well as observed random variable $\mathbf{w}$. For j^{th} word of i^{th} document, let $w_{i,j}$ and $z_{i,j}$ be the word and topic, respectively. For $i \in \{1, 2, \ldots, M\}$ and $j \in \{1, 2, \ldots, N\}$, it is assumed that:

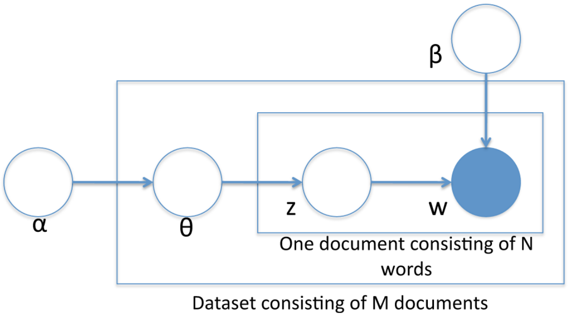

β
α
θ
z
w
One document consisting of N words
Dataset consisting of M documents

Figure 1: Graphical representation of LaDiAl for modeling document generation.

- $\boldsymbol{\theta}_i$ follows a Dirichlet distribution governed by parameter $\boldsymbol{\alpha}$.
- $z_{i,j}$ follows a multinomial distribution governed by parameter $\boldsymbol{\theta}_i$.
- $w_{i,j}$ follows a multinomial distribution governed by parameter $\boldsymbol{\phi}_{z_{i,j}}$ which in turn is governed by Dirichlet distribution with $\boldsymbol{\beta}$ parameter.

Let k and t be the vocabulary (i.e. set of all unique words) size and the number of topics respectively. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are t and k dimensional dataset level parameters receptively; i.e. they are sampled once in the process of the dataset generation. $\boldsymbol{\theta}$ is a t dimensional document level parameter which is sampled once for each of the M documents. $\mathbf{z}$ and $\mathbf{w}$ are N dimensional word level variables representing topics and words of the document, respectively.

The joint distribution of i^{th} document $\mathbf{w}_i$, document level parameter $\boldsymbol{\theta}_i$, and topics variable $\mathbf{z}_i$ is:

$$p(\boldsymbol{\theta}_i, \mathbf{z}_i, \mathbf{w}_i | \boldsymbol{\alpha}, \boldsymbol{\beta}) = p(\boldsymbol{\theta}_i | \boldsymbol{\alpha}) \prod_{j=1}^{N} p(z_{i,j} | \boldsymbol{\theta}_i) p(w_{i,j} | z_{i,j}, \boldsymbol{\beta}) \tag{1}$$

Using this joint distribution, the marginal distribution of a single document $\mathbf{w}_i$ is obtained as:

$$p(\mathbf{w}_i | \boldsymbol{\alpha}, \boldsymbol{\beta}) = \int p(\boldsymbol{\theta}_i | \boldsymbol{\alpha}) \prod_{j=1}^{N} \sum_{z_{i,j}} p(z_{i,j} | \boldsymbol{\theta}_i) p(w_{i,j} | z_{i,j}, \boldsymbol{\beta}) d\boldsymbol{\theta}_i \tag{2}$$

Eq. 2 describes how probable it is for a document $\mathbf{w}_i$ to be generated when parameters are set as $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. It is important to note that the document probability is encoded in terms of topics $z_{i,j}$. Thus, for fixed $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, one can understand the topics by observing words $w_{i,j}$.

2.1. Analogy with Faces

LaDiAl shows an effective way of modeling document generation and document representation. It is our hypothesis that a similar approach can be useful in understanding facial representation and therefore, LaDiAl can be explored in face recognition. However, adaptation of LaDiAl for face images warrants attention to the following details:

- LaDiAl is designed to deal with unstructured textual documents; whereas, faces have a well structured geometry.
- LaDiAl is designed to operate on discrete data, thus continuous data cannot be directly utilized with LaDiAl.

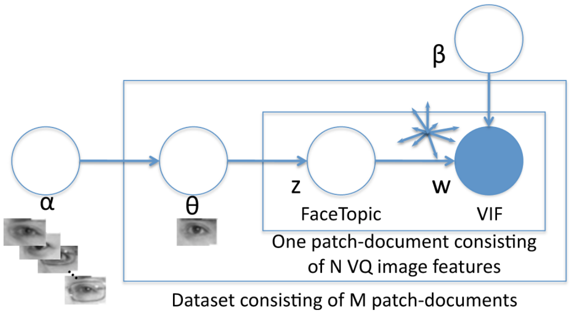

β
α
θ
z
w
FaceTopic
VIF
One patch-document consisting of N VQ image features
Dataset consisting of M patch-documents

Figure 2: Graphical representation of LaDiAl for encoding face patches.

In order to address the first aspect, we propose to apply LaDiAl on local facial patches, rather than the full face. The second challenge can be addressed by performing vector quantization (VQ) prior to applying LaDiAl. Such vector quantization is often termed as codebook learning also. Instead of applying LaDiAl on raw pixels, it is more appropriate to utilize image features such as dense scale invariant feature transform [20], local binary patterns [2], and Weber local descriptor [12]. Table 1 shows the analogy between text and face domains and Figure 2 shows the proposed modeling of LaDiAl for face representation. As shown in Figure 2, Latent Dirichlet Allocation on face images can provide a model describing how a face patch is represented using the latent *FaceTopics*. It is our assertion that similar to the manner in which documents are represented using *topics*, the proposed *FaceTopics* can encode face representation and can be utilized for recognition.

Text	**Face Images**
Word	VQ Image Features (VIF)
Document	Set of VIF from a Face Patch (Patch-Document)
Topics	FaceTopics

Table 1: Analogy between text and face representation.

3. Proposed Approach

The proposed approach is pictorially explained in Figure 3. The input is a pre-processed and registered (with respect to pre-defined eye location and inter-eye distance) face image. The face image is divided into 16 patches by utilizing the golden ratio template [3]; similar image tessellation is also used in [7]. Such a tessellation helps in creating facial patches corresponding to features such as eyes, nose, and mouth. Separate LaDiAL models are learned for each of the 16 facial patches using the training data. For the test data, LaDiAl features are computed from each of the patch image using the corresponding trained model. In order to compute the match score of an image pair, distance between LaDiAl features corresponding to each patch is computed and fused using weighted score fusion [29].

3.1. LaDiAl based Feature Representation

Topic modeling relies on the following assertions:

- full face can be seen as structural arrangement of facial parts and
- each facial part can be described using a set of FaceTopics.

These assertions can be exemplified by an example: types of eyes are limited and any eye can be obtained by appropriate mixture of eye templates. This is analogous to Eigenface [32] and independent component analysis [4] approaches, where it is assumed that every face can be constructed by appropriate mixture of template Eigenfaces. In a similar spirit, we aim to learn *FaceTopics* of each facial part and utilize them to obtain a novel representation. The procedure to obtain the LaDiAl based feature representation consists of three stages:

1. Converting the patch image into a patch-document,
2. Obtaining topic features (aka FaceTopics) from patch-document, and
3. Obtaining discriminative topic features.

3.1.1 Image to Patch-Document Conversion

In this research, dense scale invariant feature transform (DSIFT) [20] features are utilized as image features. From each patch, DSIFT features[3] are extracted from uniformly spaced points. Note that, every element in (normalized) DSIFT is a real number and therefore, the vocabulary size is very large. This makes it challenging to utilize DSIFT features directly as analogous to words in textual data. To address this challenge, we perform vector quantization (VQ) of DSIFT features using K-means approach. Essentially, K-means VQ assimilates similar DSIFT features into corresponding clusters. After applying VQ, the vocabulary size is k (= number of clusters) and the set of cluster association of DSIFTs computed from each key point on the uniform grid becomes the patch-document representation. Thus, a patch-document is a set of vector quantized image features (VIF) and is equal to the number of key points.

3.1.2 Topic Features from Patch-Documents

Here, the goal is to obtain a representation of patch-documents in terms of their FaceTopics. In the LaDiAl model (Eq. 2), the probabilistic relationship between VIFs and FaceTopics can be learned using inference techniques

[3] We have also explored Local Binary Patterns (LBP) as image features. However, finally DSIFT features are utilized and reported because of their promising results.

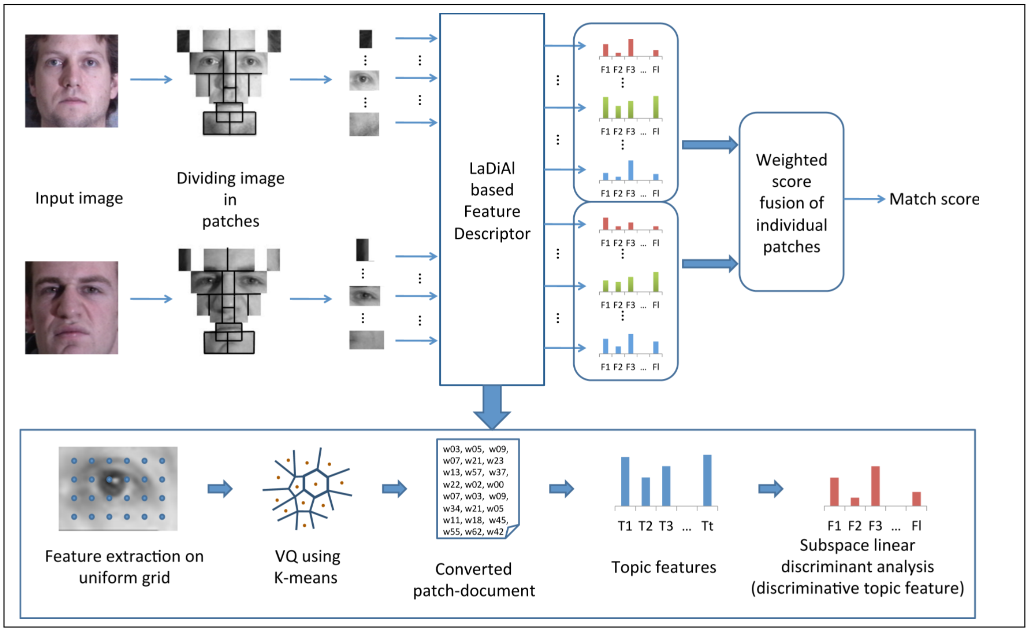


Figure 3: Illustrating the steps involved in the proposed LaDiAl based face recognition approach.

such as variational Bayes approximation [9], Gibbs sampling [14], or expectation propagation [22]. In this research, we utilize Gibbs sampling inference technique [14] for learning this relationship. The technique requires predefined number of FaceTopics t and corpus level parameters ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$). The training set is utilized to learn the generative model, leading to following two matrices:

- $\mathbf{W}$ is a matrix of size $k \times t$, where the element $W_{i,j}$ represents the number of times when the i^{th} vector quantized DSIFT feature (VIF) has been assigned to j^{th} FaceTopic, $i \in \{1, 2, \ldots k\}$ and $j \in \{1, 2, \ldots, t\}$.
- $\mathbf{D}$ is a matrix of size $M \times t$, where the element $D_{i,j}$ represents the number of times a vector quantized DSIFT feature in i^{th} patch-document has been assigned to j^{th} FaceTopic, $i \in \{1, 2, \ldots M\}$ and $j \in \{1, 2, \ldots, t\}$.

Let V_i be the i^{th} VIF in the vocabulary and T_j be the j^{th} FaceTopic. The matrices $\mathbf{W}$ and $\mathbf{D}$ are used to obtain (1) FaceTopic conditional probability of each word V_i in the vocabulary (Eq. 3) and (2) prior probability of each FaceTopic T_j (Eq. 4).

$$p(V_i|T_j) = \frac{W_{i,j}}{\sum_{p=1}^{k} W_{p,j}} \tag{3}$$

$$p(T_i) = \frac{\sum_{p=1}^{M} D_{p,i}}{\sum_{p=1}^{M} \sum_{q=1}^{t} D_{p,q}} \tag{4}$$

For testing, given any patch-document $\mathbf{w}$, the posterior probability $p(T_i|\mathbf{w})$ that it belongs to i^{th} FaceTopic is computed. This probability is proportional to $\mathcal{T}_{\mathbf{w}}^i$, which is defined as

$$\mathcal{T}_{\mathbf{w}}^i = \sum_{v \in \mathbf{W}} p(T_i) p(v|T_i) \tag{5}$$

The topic feature vector representation $\boldsymbol{\mathcal{T}}_{\mathbf{w}}$ of the patch-document $\mathbf{w}$ is created as the concatenation of such t posterior probabilistic forms pertaining to each FaceTopic, i.e.

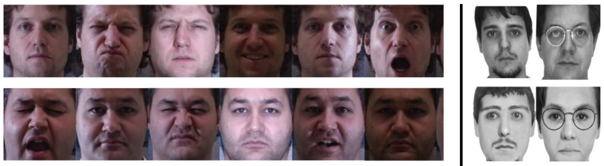
$$\boldsymbol{\mathcal{T}}_{\mathbf{w}} = [\mathcal{T}_{\mathbf{w}}^1, \mathcal{T}_{\mathbf{w}}^2, \ldots, \mathcal{T}_{\mathbf{w}}^t] \tag{6}$$

3.1.3 Discriminative Topic Feature

Topic features obtained from Eq. 6 are appropriate for representing the corresponding patch-documents. However, they may not be well suited for classification tasks (e.g. face recognition). They can be further fine-tuned towards classification tasks by applying discriminative learning. Note that, the first two learning stages, i.e. image to patch-document and then to topic feature conversion are unsupervised in nature. In order to make the topic feature representation more useful for face recognition, they are transformed using supervised learning. Potentially, any supervised learning with objective function as maximization of inter-class and minimization of intra-class variations should be applicable. Here, Linear Discriminant Analysis (LDA) [5] is applied to obtain a discriminative feature representation. We call this final l dimensional discriminative topic feature as $\boldsymbol{\mathcal{F}}_{\mathbf{w}}$, the LaDiAl features.

3.2. Matching using Weighted Score Fusion

The two face images are finally matched by matching the corresponding discriminative topic features followed by



(a) CMU-MultiPIE (b) e-PRIP

Figure 4: Sample images of two subjects from the CMU-MultiPIE [15] and e-PRIP dataset [16, 23].

weighted summation of these per-patch distances. Let I_g and I_p be two face images from which LaDiAl features $\boldsymbol{\mathcal{F}}_{g,i}$ and $\boldsymbol{\mathcal{F}}_{p,i}$ ($i \in \{1, 2, \ldots, 16\}$) are extracted for each of the 16 patches, respectively. The distance d between the two face images is obtained as the weighted sum of the cosine distance between corresponding patches (Eq. 7).

$$d = \sum_{i=1}^{16} \omega_i \left(1 - \frac{\boldsymbol{\mathcal{F}}_{g,i} \cdot \boldsymbol{\mathcal{F}}_{p,i}}{|\boldsymbol{\mathcal{F}}_{g,i}||\boldsymbol{\mathcal{F}}_{p,i}|}\right) \tag{7}$$

where, weight ω_i is a function of the accuracy of i^{th} patch on the training data. Let γ_i be the rank-1 accuracy on training data when only i^{th} patch is used; $\omega_i = \gamma_i / \sum_{j=1}^{16} \gamma_j$.

4. Experiments

Since the proposed approach learns a representation over DSIFT features, it's effectiveness can be evaluated in terms of comparison with the original DSIFT features applied on face images. Therefore, the features are extracted from the blocks of 8×8 key points uniformly spaced on a face image. Baseline performances are shown with two open access face recognition systems, local region principal component analysis (LRPCA) [27] and OpenBR[4] [18], and local binary patters (LBP) based approach [1]. The experiments on LaDiAl are performed with and without subspace LDA in order to substantiate it's application with the proposed LaDiAl feature descriptor. The effectiveness of the last stage of weighted score fusion is evaluated by replacing it with sum rule [29] and majority voting rule. In all the experiments, final matching is performed using cosine distance measure. All the experiments are performed to simulate identification scenario and cumulative match curves (CMC) are used for illustrating the experimental results.

4.1. Dataset and Protocol

The proposed approach is evaluated for covariates of illumination and expression; as well as on the digital photo to composite sketch face matching problems. Therefore, CMU-MultiPIE [15] and e-PRIP composite sketch dataset [16, 23] are used for evaluating the performance of the proposed approach.

Exp.	I	E	Images (Subjects)		
			Train (168)	Test (169)	Total (337)
1		✓	1,260	1,254	2,514
2	✓		11,220	11,980	23,200
3	✓	✓	24,200	26,080	50,280

Table 2: Experimental protocol details on CMU-MultiPIE dataset. I and E stand for illumination and expression, respectively.

Stage	**Dimensionality**
Input Image	22500 (=150×150)
Features from Uniform Grid	2304-7680
VQ	18-60
Topic Feature	70-100
Discriminative Topic Feature	69-99

Table 3: Dimensionality of image patch representation at various stages.

CMU-MultiPIE consists of more than 750,000 images of 337 people with several variations in pose, expression, and illumination. In this research, only the frontal images (camera no. 05_1) with all illumination and expression variations are used. Samples from the dataset are shown in Figure 4. To evaluate the effectiveness of the proposed algorithm, experiments are performed on three scenarios; in presence of (1) expression variation, (2) illumination variation, and (3) both expression and illumination combined. According to these scenarios, we created three subsets of this dataset. For the first experiment, we utilize the subset consisting of face images of varying expression, with constant (frontal) pose and (frontal) illumination. The second subset consists of face images of varying illumination, with constant (frontal) pose and (neutral) expression. For the third subset, we utilize the face images of all the combinations of varying illuminations and expressions. In the first experiment, one image per person is randomly selected to form the gallery set (due to small number of images per person), whereas in the second and third experiments, we use five randomly selected images per person to form the gallery set. All the remaining images form the probe set. The details of these subsets are given in Table 2.

The e-PRIP composite sketch dataset [16, 23], the only publicly available dataset of its kind, contains composite sketches of 123 face images from the AR face dataset [21]. It contains the composite sketches created using two softwares, Faces[5] and IdentiKit[6]. The PRIP dataset [16] originally has composite sketches prepared by a Caucasian user (IdentiKit and Faces) and an Asian user (Faces). Later, the dataset is extended by Mittal *et al.* [23] by adding composite sketches prepared by an Indian user (Faces) which is termed as e-PRIP composite sketch dataset. The experiments are

[4] The pre-trained face recognition engine provided with OpenBR is utilized, which is based on Spectrally Sampled Structural Subspaces Feature (4SF) algorithm.

[5] http://www.iqbiometrix.com

[6] http://www.identikit.net

Approach		Rank-1 Accuracy
Feature	Fusion	$\mu \pm \sigma$ (%)
OpenBR [18]	-	21.8 ± 0.6
LRPCA [27]	-	37.0 ± 2.7
LBP [1]	-	66.6 ± 3.3
DSIFT	-	52.8 ± 0.0
DSIFT+LDA	-	66.0 ± 3.1
LaDiAl (w/o LDA)	Majority	44.4 ± 2.0
	Sum	54.9 ± 2.9
	Weighted	58.7 ± 4.0
LaDiAl	Majority	44.4 ± 1.7
	Sum	66.0 ± 3.2
	Weighted	**67.3** ± **4.1**

(a) Exp. 1 (expression variation)

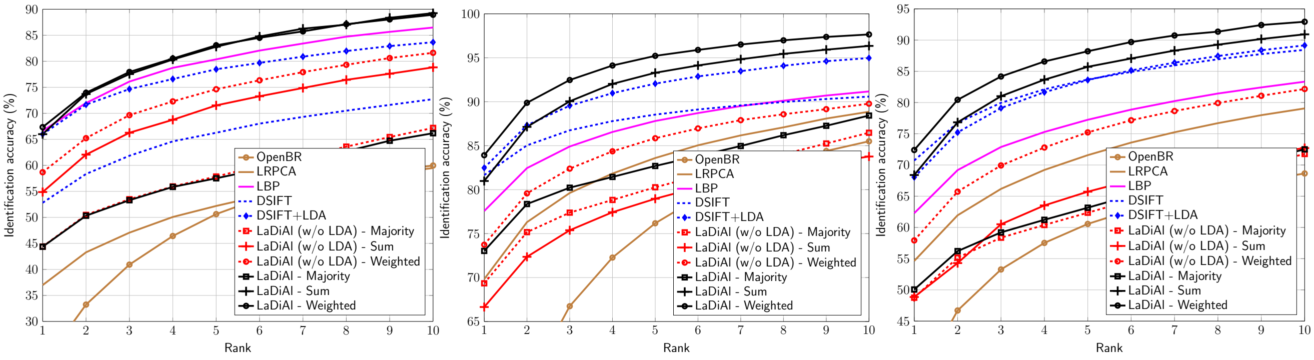
Approach		Rank-1 Accuracy
Feature	Fusion	$\mu \pm \sigma$ (%)
OpenBR [18]	-	41.6 ± 0.2
LRPCA [27]	-	69.8 ± 2.7
LBP [1]	-	77.6 ± 0.1
DSIFT	-	81.7 ± 0.5
DSIFT+LDA	-	82.5 ± 0.8
LaDiAl (w/o LDA)	Majority	69.4 ± 0.6
	Sum	66.7 ± 0.4
	Weighted	73.7 ± 0.1
LaDiAl	Majority	73.0 ± 0.9
	Sum	81.0 ± 0.3
	Weighted	**84.0** ± **0.1**

(b) Exp. 2 (illumination variation)

Approach		Rank-1 Accuracy
Feature	Fusion	$\mu \pm \sigma$ (%)
OpenBR [18]	-	34.1 ± 0.6
LRPCA [27]	-	54.6 ± 0.6
LBP [1]	-	62.3 ± 0.1
DSIFT	-	70.7 ± 0.3
DSIFT+LDA	-	68.0 ± 0.4
LaDiAl (w/o LDA)	Majority	48.7 ± 0.4
	Sum	48.8 ± 0.4
	Weighted	57.9 ± 0.3
LaDiAl	Majority	50.0 ± 0.4
	Sum	68.4 ± 0.4
	Weighted	**72.4** ± **0.3**

(c) Exp. 3 (expression and illumination variation)

Table 4: Mean and standard deviation of rank-1 identification accuracy with the three experimental protocols on CMU-MultiPIE database.



(a)

(b)

(c)

Figure 5: CMCs pertaining to (a) experiment 1: expression variation, (b) experiment 2: illumination variation, and (c) experiment 3: expression and illumination combined.

Approach	**Rank-10 Accuracy (%)**	
	Faces (Caucasian)	**Faces (Indian)**
COTS [23]*	11.3±2.1	9.1±1.9
Mittal *et al.* [24]*	32.4±2.4	30.3±1.7
LaDiAL+Sum	34.1±6.9	36.3±5.6
LaDiAl+Weighted	38.4±5.9	39.7±5.9

Table 5: Rank-10 Identification accuracy for composite sketch (query) to photo (gallery) matching on e-PRIP dataset [16, 23]. $*$ are results reported by Mittal *et al.* [23].

performed with the same protocol as presented by Mittal *et al.* [23]. The dataset is divided into 40% training (48 subjects) and 60% testing (75 subjects), with random sampling based five times cross validation.

4.2. Preprocessing

Face regions are segmented by utilizing the eye locations provided by El Shafey *et al.* [13] and are resized to 150× 150 pixels. Since Gibbs sampling based inference model is used, symmetric Dirichlet distributions are used with $\alpha = 50/t$ and $\beta = 0.01$ [30]. In the experiments, t is varied from 20 to 100 with step of 10; and vocabulary size $k = 128$ and $k = 64$ for experiments pertaining to CMU-MultiPIE and e-PRIP datasets respectively. The dimensionality of features at various stages of the proposed approach are given in Table 3. Depending on the size of each patch, 18 to 60 key points are selected on a uniform grid, which results into a varying length image feature size of 2304 to 7680. The output of vector quantization stage is same as the number of points on uniform grid, which is 18 and 60 for the smallest and the largest patch respectively. According to the length of topic features, the discriminative topic features are of $t - 1$ dimensionality.

4.3. Results and Analysis

Rank-1 identification accuracies of three experiments on the CMU-MultiPIE dataset are reported in Tables 4a, 4b, and 4c. Their mean and standard deviation of rank-1 accuracy over three random cross validations are reported and CMCs are shown in Figure 5. The rank-10 identification accuracies of the experiment pertaining to composite sketch to photo matching on e-PRIP dataset are summarized in Table 5. The key observations from these experiments are as follows:

Vocabulary Size (k)	**Number of FaceTopics (t)**								
	20	30	40	50	60	70	80	90	100
32	52.2±5.9	50.9±5.4	41.5±5.5	**53.1±7.5**	54.3±6.1	40.7±6.8	43.2±5.0	52.2±6.5	52.3±7.2
64	38.3±6.3	54.4±4.4	45.6±4.7	52.1±4.9	**62.3±3.8**	51.0±1.5	39.2±3.3	40.8±3.2	61.2±3.3
128	54.7±3.4	48.7±2.9	63.2±3.7	65.2±3.4	57.7±3.4	**67.3±4.1**	46.8±2.6	54.2±4.0	65.4±3.0

Table 6: Effect of varying the number of topics and vocabulary size on experiment 1 in terms of rank-1 identification accuracy ($\mu \pm \sigma\%$).

Stage	**Time (s)**
Face tessellation	0.0323
Extraction of DSIFT features	0.1315
Patch-Document	0.0270
Topic feature	0.1908
Discriminative topic feature	0.0006
Total time for feature extraction	0.3822

Table 7: Average time required for feature extraction from an image.

- Both the open source face recognition systems, LRPCA and OpenBR, and the LBP based approach yield lower performance compared to the proposed LaDiAL based approach with weighted fusion. Further, in presence of both covariates of illumination and expression, the performance of LBP, LRPCA, and OpenBR is significantly lower than the proposed approach.

- **FaceTopic:** The proposed approach (LaDiAl + weighted) yields better results than using DSIFT only. This may be attributed to FaceTopic based representation and/or to the discriminative learning. However, the proposed approach exhibits better performance than DSIFT+LDA also, which shows the effectiveness of FaceTopic based representation. It is to be noted that in presence of only expression variation, LaDiAl with weighted score fusion and DSIFT with LDA exhibit comparable mean rank-1 accuracy. However, as the rank increases, the proposed approach outperforms DSIFT+LDA.

- **LDA:** Comparing the results of topic features $\mathcal{T}$ (LaDiAl without LDA) and discriminative topic features $\mathcal{F}$ (LaDiAl) show the effectiveness of the later. Similarly, DSIFT+LDA exhibits better performance than DSIFT in experiments 1 and 2. This shows that discriminative learning is helpful.

- **Data Size:** Effectiveness of the proposed approach is more clearly visible in experiments 2 and 3, with performance improvement of ∼2.5% and ∼4.5% in mean rank-1 accuracy, respectively. This suggests that in presence of large intraclass variations, DSIFT features may not yield high performance, however LaDiAl features may still remain robust. These results also suggest that the larger training set is well suited for leaning LaDiAl model.

- **Weighted Fusion:** Consistently in all three experiments on CMU-MultiPIE database, weighted score fusion outperforms majority voting and sum rule. Sum rule fusion is better than majority voting and the training accuracy based weighted score fusion performs almost same or better than sum rule fusion. This shows that different patches have different discriminating power and the information from these patches is effectively combined in the proposed algorithm.

- **Parameters:** For LaDiAl, the results pertaining to the best value of t are reported. For the first, second and third experiments, the best results are obtained at $t = 70, 100,$ and 70 respectively. In addition to the aforementioned experiments, an additional experiment is performed to study the effect of parameter selection. Experiment 1 is repeated with vocabulary size $k = 32, 64,$ and 128 along with varying the number of FaceTopics t from 20 to 100 with a step size of 10. The results reported in Table 6 show that with bigger vocabulary size, larger number of FaceTopics yield better results.

- **Time:** On a server with two Intel(R) Xeon(R) E5-2640 (2.5GHz) processors and 64 GB RAM, under Matlab environment, the average time required at each stage of feature extraction is given in Table 7. Matching two features require an average of 0.0719 seconds. We believe that the computational time can be further improved by native implementations with parallel programming.

- The effectiveness on e-PRIP dataset suggests that the proposed approach may be able to handle the heterogeneous data such as composite sketch and photo. Motivated with these results, we assert that the proposed approach can be further extended to other heterogeneous face recognition problems.

5. Conclusion and Future Work

The main contribution of this research is showcasing the applicability of text analytics inspired approach for face recognition. This paper presents Latent Dirichlet Allocation based approach for facial feature representation. To encode LaDiAl for face recognition, an analogy between face image and textual document is performed. The generative model provides topic features, over which discriminative learning is applied to obtain final feature representation. The evaluation on CMU-MultiPIE and e-PRIP sketch datasets show the effectiveness of the proposed approach. In future, we plan to study its applicability for other challenging covariates and in extended gallery set scenario. We would also like to explore the utility of image features other than DSIFT in the proposed framework.

Acknowledgement

This work is supported by a grant from the Department of Electronics and Information Technology, India. T. I. Dhamecha is partly supported through TCS PhD research fellowship.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *ECCV*, pages 469–481. 2004.

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE TPAMI*, 28(12):2037–2041, 2006.

[3] K. Anderson and P. W. McOwan. Robust real-time face tracker for cluttered environments. *CVIU*, 95(2):184 – 200, 2004.

[4] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE TNN*, 13(6):1450–1464, 2002.

[5] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE TPAMI*, 19(7):711–720, 1997.

[6] H. Bhatt, R. Singh, and M. Vatsa. Covariates of face recognition. Technical report, IIIT Delhi, 2015.

[7] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. Recognizing surgically altered face images using multiobjective evolutionary algorithm. *IEEE TIFS*, 8(1):89–100, 2013.

[8] D. M. Blei and J. D. McAuliffe. Supervised topic models. In *NIPS*, volume 7, pages 121–128, 2007.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *JMLR*, 3:993–1022, 2003.

[10] A. Bosch, A. Zisserman, and X. Muoz. Scene classification using a hybrid generative/discriminative approach. *IEEE TPAMI*, 30(4):712–727, 2008.

[11] M. Bregonzio, J. Li, S. Gong, and T. Xiang. Discriminative topics modelling for action feature selection and recognition. In *BMVC*, pages 1–11, 2010.

[12] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao. WLD: A robust local image descriptor. *IEEE TPAMI*, 32(9):1705–1720, 2010.

[13] L. El Shafey, C. McCool, R. Wallace, and S. Marcel. A scalable formulation of probabilistic linear discriminant analysis: applied to face recognition. *IEEE TPAMI*, 35(7):1788–1794, 2013.

[14] T. L. Griffiths and M. Steyvers. Finding scientific topics. *PNAS*, 101(Suppl 1):5228–5235, 2004.

[15] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *IVC*, 28(5):807–813, 2010.

[16] H. Han, B. F. Klare, K. Bonnen, and A. K. Jain. Matching composite sketches to face photos: A component-based approach. *IEEE TIFS*, 8(1):191–204, 2013.

[17] T. Hofmann. Probabilistic latent semantic indexing. In *ACM SIGIR*, pages 50–57, 1999.

[18] J. C. Klontz, B. F. Klare, S. Klum, A. K. Jain, and M. J. Burge. Open source biometric recognition. In *IEEE BTAS*, pages 1–8, 2013.

[19] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In *CVPR*, volume 2, pages 2169–2178, 2006.

[20] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[21] A. M. Martinez. The AR face database. *CVC Technical Report*, 24, 1998.

[22] T. Minka and J. Lafferty. Expectation-propagation for the generative aspect model. In *UAI*, pages 352–359. Morgan Kaufmann Publishers Inc., 2002.

[23] P. Mittal, A. Jain, G. Goswami, R. Singh, and M. Vatsa. Recognizing composite sketches with digital face images via SSD dictionary. In *IEEE IJCB*, pages 1–6, 2014.

[24] P. Mittal, A. Jain, R. Singh, and M. Vatsa. Boosting local descriptors for matching composite and digital face images. In *IEEE ICIP*, pages 2797–2801, 2013.

[25] J. Niebles, H. Wang, and L. Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. *IJCV*, 79(3):299–318, 2008.

[26] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *ECCV*, pages 490–503, 2006.

[27] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'Toole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer. An introduction to the good, the bad, & the ugly face recognition challenge problem. In *IEEE F&G-Workshop*, pages 346–353, 2011.

[28] N. Rasiwasia and N. Vasconcelos. Latent Dirichlet allocation models for image classification. *IEEE TPAMI*, 35(11):2665–2679, 2013.

[29] A. Ross and A. Jain. Information fusion in biometrics. *PRL*, 24(13):2115–2125, 2003.

[30] M. Steyvers and T. Griffiths. Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7):424–440, 2007.

[31] E. Sudderth, A. Torralba, W. Freeman, and A. Willsky. Describing visual scenes using transformed Dirichlet processes. In *NIPS*, pages 1299–1306, 2005.

[32] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *CVPR*, pages 586–591, 1991.

[33] X. Wang and E. Grimson. Spatial latent Dirichlet allocation. In *NIPS*, 2007.

[34] Y. Wang, P. Sabzmeydani, and G. Mori. Semi-latent Dirichlet allocation: A hierarchical model for human action recognition. In *Human Motion–Understanding, Modeling, Capture and Animation*, pages 240–254. Springer, 2007.