2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics

Whose Album is this?

Abhinav Goel, C. V. Jawahar
Centre for Visual Information and Technology
International Institute of Information Technology
Hyderabad, PIN 500032, India
Email: abhinav.goel@research.iiit.ac.in, jawahar@iiit.ac.in

***Abstract*—We present a method to identify the owner of a photo album taken off a social networking site. We consider this as a problem of prominent person mining. We introduce a new notion of prominent persons, and propose a greedy solution based on an eigenface representation. We mine prominent persons in a subset of dimensions in the eigenface space. We present excellent results on multiple datasets downloaded from the Internet.**

***Index Terms*—Eigenface, Clustering, Frequent Itemset Mining, Apriori, Prominent Person Mining**

I. Introduction

Look at Figure 1. Whose Photo Album is this? The answer is obvious – Amitabh Bachchan. If you want a machine to do this task, there are two fundamental problems to be solved. (i) Find the person (face image) who owns this album from a large number of faces present. (ii) Label the person with a name based on apriori knowledge. In this work, our aim is limited to finding the person (face image) to whom a photo album belongs. We focus our attention on finding the template or representative of the person in the form of a face image.

In photo sharing sites, we often find the owner posing for pictures with other people, and posting them as an album. In most cases, he/she occurs the most number of times in these pictures, and is the most prominent person. Hence we call our technique as that of prominent person mining. This could be a step towards extracting useful information from unannotated large collection of images and videos available on the Internet. Such methods will also have applications in processing surveillance videos for mining social and behavioural patterns of people the camera frequently observes.

The notion of prominent person mining is different from that of Frequent Itemset Mining (FIM) [8]. In case of FIM, the itemsets are discrete and well defined. However, in our case the "items" or "persons" appear with wide variations. This uncertainty has to be accomodated into the mining algorithm. In addition, the prominent person is not just the frequent person. Prominent person often occupy a significant area of the image. He/She is also the centre of attraction of the image.

There have been earlier attempts at using data mining techniques in computer vision. Sivic and Zisserman used video mining in [1] for obtaining the principal objects, characters and scenes by measuring the occurence of spatial configurations of viewpoint invariant features. Frequently occuring spatial configurations were found using clustering algorithms, rather than any frequent itemset mining schemes. Quack et al. find frequently occuring objects and scenes in [2], where

Fig. 1. Whose Photo Album is this? The answer is obvious to a human being. For a machine, this involved mining of the prominent face from a large number of faces present.

frequent itemset mining schemes have been used. Data mining algorithms have also been used to assist tasks like object detection [4] and mining objects and events from community photo collections [5]. The characteristic persons are mined in a movie in [7] by first mining the characteristic scenes, and then the persons present in those scenes.

We differ from previous works in multiple dimensions. (i) We focus our attention on prominent person mining rather than frequent itemset mining, (ii) We refine the existing data mining algorithms to suit the problem of prominent person mining, and (iii) Our mining algorithm works under the uncertainty of feature representation rather than vector quantizing the feature descriptor to obtain a crisp representation of visual entitites.

We formulate our problem as that of finding the most prominent person in the given set of images. We use the eigenface representation. In Section 2 we mine prominent persons using an Apriori-like algorithm which generates prominent candidiates in k dimensions from prominent persons in $k-1$ dimensions in the eigenface space. We conduct extensive experiments on multiple datasets with images captured in real world scenarios and achieve excellent results. We test our approach on 200 photo albums consisting of over 3000 images download from the Internet and achieve an accuracy of over 90% The results have been presented in Section 3. We conclude in Section 4 and give some future research directions.

978-0-7695-4599-8/11 $26.00 © 2011 IEEE
DOI 10.1109/NCVPRIPG.2011.26

II. PROMINENT PERSON MINING

Our goal is to find the most prominent person in a given album. Faces are extracted from the images in the album using the OpenCV implementation of the Viola-Jones Face detector [10]. A set of eigenfaces are obtained by performing principal component analysis (PCA) on the faces extracted [9]. For the process of mining, each face image is represented in the form of a feature vector, in which each value corresponds to the contribution of one eigenface. Our database consists of several transactions, each being a face in the set of extracted faces from the album images. Each face is a point in the K dimensional eigenface space. We propose a greedy algorithm to find the most prominent person in a subset of dimensions. We first mine the most prominent person in a single dimension and move in a heirarchical manner to mine prominent persons in more than one dimensions. Our approach is based on a popular data mining algorithm for mining frequent itemsets. Here we briefly summarize the terminology and methodology of frequent itemset mining.

A. Frequent Itemset Mining

Let $I = \{I_1 \dots I_p\}$ be a set of p items. Let D, the task-relevant data, be a set of database transactions where each transaction T is a set of items such that $T \subseteq I$. Each transaction is associated with an identifier, called TID(T). A transaction contains X, a set of items in I, if $X \subseteq T$. The support of an itemset $X \subseteq I$ is given by $support(X) = \frac{|\{T \in D | X \subseteq T\}|}{|D|}$. If the support of the itemset X satisfies the minimum support count threshold $(support(X) > s)$, then it is said to be frequent, where s is the minimum support threshold. Several methods have been proposed to mine frequent itemsets efficiently. The Apriori algorithm is a classic algorithm to find the frequent itemsets in a given database for a given minimum support value.

In traditional frequent itemset mining, the itemsets are well defined and discrete. This is not the case with faces, as "face" representations can vary. We use an approach similar to the Apriori algorithm, where we change the notion of frequent itemsets, and refer to them as prominent persons, taking into account the uncertainty in faces. Prominent "persons" or "itemsets" are faces of the same person who is prominent in the given set of images. Prominent persons often occupy a significant area of the image. He/She is also the centre of attraction of the image. We first find the most prominent person in a single dimension, and then move towards more number of dimensons in the eigenface space. Here each dimension corresponds to one eigenface used for computing the weight values. Our algorithm takes three input parameters:-

- K (Number of dimensions): Minimum number of dimensions across which we want the person to be prominent.
- D (Diameter): Diameter for the terminating condition of the agglomerative hierarchical clustering method used to find prominent persons in a single dimension.
- S (Minimum Count): Minimum support threshold used in frequent itemset mining schemes to determine whether an itemset is frequent or not.

B. Finding prominent persons in 1 dimension

In this step, we wish to identify the prominent persons in each dimension. We perform semi-supervised agglomerative hierarchical clustering in each dimension. This method builds the hierarchy from individual elements by progressively building clusters. In every step, clusters closest to each other are merged until some terminating criteria is met. The distance between two clusters can be defined in several ways. We perform complete linkage clustering, where the maximum distance between elements of each clusters, defined as $max\{d(x, y) : x \in A, y \in B\}$ where A and B are the clusters to be merged, is used. All clusters with diameter less than D are computed. Semi-supervision comes in the form of cannot-link constraints. We ensure that faces extracted from the same image cannot belong to the same cluster. This constraint is justified as faces extracted from one image are bound to be of different persons, and only one of them can be the most prominent person. The elements belonging to the cluster with the largest number of items are considered as prominent if the number of elements in the cluster is greater than S. We assume that the person to which this profile belongs is present in every image in his album. Hence if we have N test images, and of those $M(\gg N)$ face images are extracted, our minimum count theshold is set to a value $\frac{N}{3}$ to take care of missed detections by the face detector. In an ideal scenario, where the face detector gives 100% results, and the environment in which the face images have been captures is controlled, the value of S can be set closer to N.

C. Finding prominent persons in k dimensions

Each dimension of the eigenface space is represented by a bit vector, where the bits corresponding to the prominent persons are set to 1. Say, if we have 10 transactions in our database, and the bit vector for the i^{th} dimension d_i is $\vec{d_i} = 1100010000$, then the 1^{st}, 2^{nd} and 6^{th} persons are prominent in dimension d_i. We progressively compute prominent persons in k dimensions. This is done in an Apriori fashion, where candidiates for prominent persons are generated in k dimensions from prominent persons in $k-1$ dimensions. Hence, if a person is prominent in dimensions d_1d_2, d_1d_3, d_2d_3 and d_2d_4, candidate triplet of dimensions which are generated are $d_1d_2d_3$ and $d_2d_3d_4$. The bit vectors are obtained by taking and of the bit vector of each individual dimension.

$$d_1 d_2 \dots d_i = \vec{d_1} \wedge \vec{d_2} \wedge \dots \vec{d_i}$$

In the pruning step, we eliminate candidate dimensions in which the person is not prominent in at least one of the immediate subset of dimensions. For example, we eliminate $d_2d_3d_4$ from the candidate set since it's immediate subset d_3d_4 does not contain a prominent person. Our algorithm returns persons who occur at least S times in atleast K dimensions.

We successfully incorporate semi-supervision into the mining process. Each face extracted from an image is assigned a weight :-

$$w(f_i) = f_i.presence + f_i.ratio - f_i.deviation$$

where $w(f_i)$ is the weight assigned to face f_i. $f_i.presence$ encodes information about the number of persons present in the image. $f_i.ratio$ encodes information about ratio of the size of a person's face to the size of the image. $f_i.deviation$ tells how far the person is from the centre of the image. A prominent person often occupies a large area and is the centre of attraction in the image. Each component of the weight term is normalized to ensure equal contribution in the mining process. This is done using the *min-max normalization* technique, which performs a linear transformation on the original data. The count of a candidate prominent person in a single dimension or combination of dimensions is given by $Count(p_i) = \sum_{i=1}^{I} w(f_i)$. The weight ensures higher priority is given to persons in the mining process who are the centre of attraction and occupy a large area in an image.

We might get multiple results if S is low. This happens in cases where each image has a large number of people. We assign a primary score, $PScore(p_i) = k_i$ to each prominent person, where k_i is the number of dimensions in which person p_i is prominent with frequency s_i. For the same primary score, we compute a secondary score, $SScore(p_i)_k = s_i$ to each prominent person in the same dimension, where person p_i is prominent with frequency s_i in k dimensions. Hence we prefer those results in which the person is prominent in the dimensions whose eigenvalues are larger. We compute a score similar to the weights assigned to each face in the prominent person mining method. The score of each face is computed as $TScore(f_i) = f_i.presence + f_i.ratio - f_i.deviation$ For the result with maximum ($PScore(p_i)$ and ($SScore(p_i)$), we return the face with the maximum value of $TScore$.

III. Experiments and Results

We conducted extensive experimentation on several datasets to demonstrate the effectiveness of our approach in real world scenarios. The first set of experiments were conducted on face images taken in a controlled environment, with small variation in appearance and background. For this purpose, we used the ORL face database. The ORL face database consists of ten images each of 40 different subjects. The size of each image is 92×112 pixels. For some subjects, the images were taken at different times, varying lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/ no glasses). All images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for side movement). We generated 100 photo albums from this set, with the most prominent person having 8-10 photos and the other persons with 1-4 photos. In the eigenface computation, the eigenvectors corresponding to the top 10 eigenvalues were used for the face representation. Using more number of eigenvectors does not affect the results as most of the the information about the distribution in the data is stored in the first few eigenvectors. Out of the 100 photo albums we tested our approach on, we found 100 prominent persons. For 90 of these photo albums the person we found as most prominent was the actual owner of the photo album. The results of these experiments are summarized in Table I Column II.

Now that we have estabilished the correctness of our approach on images captured under controlled conditions, we needed to test on a challenging real world scenario. Characteristic of real world images are change in appearance, variation in pose and background. We downloaded images of the Labeled Faces in the Wild (LFW) dataset. The LFW Database [11] consists of more than 13000 images collected from the web. 1680 of the people pictured have two or more distinct photos in the data set and only 143 have more than 10 images. We downloaded this subset of images of 143 people. The database has been designed for the task of unconstrained face recognition and is much more challenging than the ORL face dataset. Most of the images of the same person have been taken from different event gatherings, often separated by a large time gap, leading to change in apperance, pose and illumination. The OpenCV implementation of the Viola-Jones face detector was used to detect faces, which were resized to a fixed size of 100×100. We generated 92 photo albums in a similar fashion as in the case of ORL face dataset and conducted experiments to find out the most prominent person in each album. Results are summarized in Table I Column III.

TABLE I
RESULTS
N: NUMBER OF PHOTO ALBUMS NPP: NUMBER OF PROMINENT PERSONS FOUND CPP: NUMBER OF CORRECT PROMINENT PERSONS FOUND A: ACCURACY

	ORL	LFW	IMDB1	IMDB2
N	100	92	100	50
NPP	100	92	100	50
CPP	90	70	91	47
A (in %)	90	76	91	94

Our next task was to test our algorithm on real world photo albums. We crawled pages of 25 celebrities on IMDB and downloaded 100 photo albums. Each celebrity page has a link to multiple pages for each event he/she has been a part of. On that page are photos taken during that particular movie premiere or award show. On an average, each album contains 16 photos. Faces were extracted using the OpenCV implementation of the Viola-Jones Face Detector. 29 faces were detected per photo album on an average. The faces in each photo album were used for the computation of eigenfaces for that album. Out of the 100 photo albums, the most prominent person mined in 91 albums was the actual owner of the album as shown in Table I Column IV. We use the term IMDB1 to denote this set. The value of N is lesser than that used for the ORL dataset to account for false positives and false negatives by the Face detector.

The decrease in accuracy from the ORL face dataset to the Labeled Faces in the Wild dataset and albums downloaded from IMDB can be attributed to the variation in background and pose of the actor. In few of the cases we detected the second most prominent person in the photo album, which was

Fig. 2. Prominent Persons mined on photo albms of six celebrities downloaded from IMDB. A few images frome each photo album are shown on the left. The corresponding most prominent person mined is shown on the right (a) Actor: Nicholas Cage, Movie: Matchstick Men (b) Actor: Jennifer Anniston, Movie: Marley and Me (c) Actor: George Clooney, Movie: Ocean's Eleven

usually when the number of face images in the album were small and the number of images of the most prominent person comparable to that of the second most prominent person.

In the Eigenface representation, eigenvectors corresponding to top few values capture most of the distribution in the data. We observed that this was the case and the dimensions in which a person was prominent were from 1-5. Hence, using dimensions 5-10 were proving to be redundant in most cases as they were not contributing to the final result. To reinforce this theory and ensure that all the 10 dimensions played a role in finding the most prominent person, we crawled pages of 200 celebrities and downloaded photo albums. Of these, we kept those which had more than 50 face images after running the Viola Jones Face detector and conducted experiments. On 47 out of the 50 albums we tested on, we successfully mined the correct prominent person. The results on this dataset (IMDB2) have been shown in Table I Column V, and are in accordance with our theory.

Time Performance: Photo albums on the Internet usually contain a few hundred images. Efficient frequent mining implementations can process a large number of transactions in seconds. In our case, where the number of transactions is in the order of few hundreds, the most prominent person can be mined in real-time.

IV. Conclusion and Future Work

We have introduced a new method of finding out the owner of a photo album, who is often the most prominent person in the set of album images. For this, we have extended the frequent itemset mining technique to mine prominent persons based on our new definition of prominence for a person. We have also developed a semi-supervised version of the algorithm, where we take into account several factors which define prominence. The algorithm has been thoroughly evaluated on different challenging scenarios to ensure mining of the correct owner of the photo album.

References

[1] Josef Sivic and Andrew Zisserman, *Video Data Mining Using Configurations of Viewpoint Invariant Regions*. In IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, 2004

[2] Till Quack, Vittorio Ferrari and Luc Van Gool, *Video Mining with Frequent Itemset Mining Configurations*. In Proceedings of CIVR 2006, Phoenix, July 13-15, 2006

[3] Mark Everingham, Josef Sivic and Andrew Zisserman, *"Hello! My name is... Buffy" - Automatic Naming of Characters in TV Video*. In Proceedings of the British Machine Vision Conference, 2006

[4] Till Quack, Vittorio Ferrari, Bastian Leibe, Luc Van Gool, *Efficient Mining of Frequent and Distinctive Feature Configurations*, In Proceedings of ICCV 2007

[5] Till Quack, Bastian Leibe, Luc Van Gool, *World-Scale Mining of Objects and Events from Community Photo Collections*. In Proceedings of CIVR 2008

[6] Josef Sivic, Mark Everingham and Andrew Zisserman, *"Who are you?" - Learning person specific classifiers fom video*. In Proceedings of the IEEE Conference on Comupter Vision and Pattern Recognition, 2009

[7] Mihir Jain and C.V. Jawahar, *Characteristic Pattern Discovery in Videos*. In Proceedings of ICVGIP 2010

[8] Rakesh Agrawal and Ramakrishnan Srikant, *Fast Algorithms for Mining Association Rules in Large Databases*. In Proceedings of International Conference on Very Large Databases, 1994

[9] M. Turk and A. Pentland, *Face Recognition using Eigenfaces*. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1991

[10] P. Viola and M. Jones, *Robust Real-Time Face Detection*. International Journal on Computer Vision, 57(2), 2004

[11] Labeled Faces in the Wild Dataset, http://vis-www.cs.umass.edu/lfw/

[12] Kiri Wagstaff, Claire Cardie, Seth Rogers and Stefan Schrodl, *Constrained K-means Clustering with Background Knowledge*. In Proceedings of the Eighteenth International Conference on Machine Learning, 2001

[13] Feng Tao, Fionn Murtagh and Mohsen Farid, *Weighted Association Rule Mining using Weighted Support and Significance Framework*, KDD, 2003