Recognizing faces in news photographs on the web

Hilal Zitouni
Computer Engineering Department
Bilkent University
Ankara, Turkey
zitouni@cs.bilkent.edu.tr

Muhammed Fatih Bulut
Computer Engineering Department
Bilkent University
Ankara, Turkey
fbulut@ug.bilkent.edu.tr

Pınar Duygulu
Computer Engineering Department
Bilkent University
Ankara, Turkey
duygulu@cs.bilkent.edu.tr

***Abstract*— We propose a graph based method in order to recognize the faces that appear on the web using a small training set. First, relevant pictures of the desired people are collected by querying the name in a text based search engine in order to construct the data set. Then, detected faces in these photographs are represented using SIFT features extracted from facial features. The similarities of faces are represented in a graph which is then used in random walk with restart algorithm to provide links between faces. Those links are used for recognition by using two different methods.**

Keywords- face, names, photographs, random walk, face detection

I. INTRODUCTION

Searching for people on the web is an important task, especially for news pages such as Yahoo! news. In order to get the pictures of a person from web pages, the common approach is to search for the name of the person in the textual information associated with the picture. However, such an approach may not always give the desired results. For instance, a news photograph may only have the face of *George W. Bush* while the text may have other names such as *Saddam Hussein*, and as a result a query for *Saddam Hussein* may return an image with *George W. Bush* (see Figure 1).

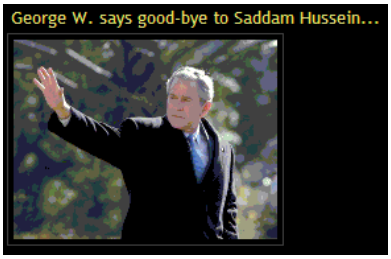

Figure 1: A query result returned for *Saddam Hüssein.*

On the other hand, while face recognition is a long-standing and well-studied problem in computer vision [13], recognition of faces in news pages is difficult. Traditional face recognition systems are likely to fail in recognizing faces on the web pages due to variations in poses, illumination and sizes, and due to several other factors such as occlusion, aging, clothing and make-up.

Recently, as an alternative to purely text based or purely visual based methods, integration of face and name information is proposed and the face recognition problem is turned into face naming problem [3, 6, 7].

In the following, first we discuss some of the recent studies in this direction and then we explain our proposed method for recognizing faces in news photographs on the web.

II. RELATED WORK

Name – face association is first studied by Satoh et al. [7]. They collect the textual and visual information from videos, and then a possible name-face association is extracted. This association is determined by finding the extracted names from the scripts and the extracted faces from the frames to appear at the overlapping time periods. The best N associated faces for a name is determined by the co-occurrence factor *C(N,F)* which is calculated by finding the occurrence rate of a face around a name in videos. In order to decide on the associated resulting face for a name, the most similar face in the association set with the dataset of that name is found.

Liu et al. in their study [4] also focus on naming faces in web images. Although naming faces is not considered as a face recognition problem generally, they claim both of them are the same problems. First they collect their face data from web search engines, and then the correct images are considered to be their recognition dataset. Once this dataset is formed, the naming faces problem is a face recognition problem, therefore the faces on the images are detected and for representation of the faces they adopt the Gabor feature approach [10] and match faces using a threshold value approach.

Berg et al. propose a method for naming faces of images that are taken in uncontrolled environments. The recent researches show that combining textual and visual information increases the correctness of face-name association [1, 2] and this method is also used in this study in order to name the faces in their large dataset. Images that are taken from news images with their associated captions are named with simple natural language and clustering

978-1-4244-5023-7/09/$25.00 ©2009 IEEE

techniques.[2] All face images are put into a pool and they are clustered according to their names, however this approach produces poor results in case of a small variation on environment conditions.

Another related problem on naming faces is studied by Ozkan et al. in [6]. They combine both textual and visual information to construct their similarity graph. In this graph based approach the nodes in their graph, represent the images, and the edges represent the weight of similarity. In order to find the most similar images they find the densest subset on this graph. Their similarity graph is constructed by comparing the interest points between two faces. However, their interest points are not only taken from particular points of the face, other detected interest points are taken into account, as well. On the contrary, selecting more interest points rather than having particular locations for facial features do not always give the best result.

Guillaumin et al. in their study [8] propose a method to name the faces using captions. They divide their work under two problems, one is to find faces for a single query, and they name all the faces in their database. In order to find images of a single query, they used the method explored in the study of Ozkan et al. [6]. Using their method, they find the densest set for a single query. In order to name all the faces, they used two approaches on a graph based method. They construct their similarity graph where the nodes represent the images and the edges represent the similarity weights. Their first approach is a kNN method with a threshold and their second approach differentiates between neighbors, but this time their similarity weights are constructed by extracting 9 facial features and using different similarity measures which are calculated by these features. Compared to [6], selecting 9 particular facial features, unlike Ozkan et al., they get rid of the matching interest points problem.

Satoh et al., in their study [9] propose an unsupervised method to annotate the faces from the web. Their study is based on two steps, first one is to mine the data from the web and find the densest set, and label the output query as query or non-query person, and in the second step they strengthen the labeling process via ranking by Bagging of SVM Classifiers. To find the densest set they use density based estimation, unlike the method used in Ozkan et al., their method does not require a threshold value.

III. PROPOSED METHOD

In our study, we focus on recognition of faces in news photographs appearing on the web. The similarities of the detected faces are found by matching SIFT features extracted from facial features. A similarity graph whose nodes represent the faces and edges represent the similarities between them is then formed. We apply a random walk algorithm on this graph in order to rank the similarity of all nodes. Below the steps of the algorithm is explained in detail.

A. Constructing the data set

We use the name information to construct our data set. Relevant pictures are collected from Google and Yahoo! web search engines via a web crawler by querying the names of the desired people in a text based search engine.

B. Detection and representation of faces

The faces on the images are then detected and represented by nine facial features -which are found to be robust to different poses and illumination conditions- using the method of Everingham et al. [12]. These nine features, as shown in Figure 2, are the left and right corners of each eye, the two nostrils and the tip of the nose, and the left and right corners of the mouth.

Figure 2. Detected faces and corresponding facial feature points.

The facial features are then represented by SIFT descriptors [5], which code the distribution of gradients in the neighborhood of a point.

C. Similarity of faces

In order to find the similarity between two faces, the Euclidean distance between the descriptors is calculated. Since there are 9 facial features, first the Euclidean distance between corresponding features are found and then their average is taken as the dissimilarity between two faces as;

$$d(A,B)=\sum_{i=1}^{N}\frac{D(i)}{N} \quad (1)$$

where *D(i)* is the Euclidean distance between the SIFT descriptors of a feature point on the two faces A and B, and N is the number of facial features (in this case it is 9). Distance values for all faces in the data set are computed and then normalized to be in the range [0-1].

In order to find the similarity between the faces, the distance, d(A,B), between the faces A and B, is then subtracted from 1.

$$sim(A,B)=1-d(A,B) \quad (2)$$

The similarities between all pairs of faces are computed to construct a similarity graph which will be used for recognition.

D. Labeling and Random Walk with Restart

It is shown in the literature that random walk with restart algorithm [11] can be used to rank the similarity of all nodes in a graph given a start node. In this study, we apply random walk with restart algorithm on the face similarity graph to list the similar faces to each face. In the following, first we explain the random walk with restart method and then describe how we use the results for recognizing faces.

Let G= (V, E) be our graph, where V represents the nodes, namely the faces and E represents the undirected weighted edges, namely similarity values among the faces. The random walk with restart algorithm is explained in Figure 3.

First we define *rs(V)*, a *1xn* vector called restart vector, where n is the number of nodes. The elements of this vector should be sum up to 1. We choose to set this restart vector in such a way that only the index of the start node will be set to 1, and the others' to 0. Each node is considered as start node in a sequence . The output for one node will be a *1xn* vector, *ps (V)*. The elements of this vector will have the probability weight for other nodes that the start node will visit. Therefore, the indices with high probabilities will be considered to belong to the nodes that are in the same category with the start node. This iterative algorithm is proved to converge and the convergence is determined by checking the L1 norm between the current and previous output vector, *ps (V)*, to be smaller than a certain value. The restart probability *c*, on the other hand, is also determined empirically, as *c* increases, the iterations for convergence gets smaller. Taking a small value for *c* refers to an increase in the field of possible visiting nodes.

```
Input :
G = (V,E) : the  nxn similarity graph.
            (n : number of nodes)
c:  restart probability

Output:

ps(V) : 1xn matrix to hold the updated
similarities of all other nodes given a start
node

Algorithm:

rs(V) : 1xn restart vector with "1" for start
node and "0" elsewhere
A : column normalized adjacency matrix of the
similarity graph

// initialize rs to ps
ps(V) = rs(V)

while not converged update ps(V) as
       ps(V) = (1-c)*A*ps(V) + c*ps(V)
end while
```

Figure 3. The algorithm for Random Walk with Restart on a graph.

After having a result vector, *ps (V)*, for every node, the labeling process begins. In order to label the nodes, a training set for each category is formed. Then the unlabeled data is labeled according to their closeness to the training sets, as will be explained in the next section.

E. Naming the Unknown Faces on the Images

In order to label the unknown data, two different approaches are proposed. One depends on finding the closest category to an unknown face, by selecting the category of the training set with the highest average of the similarity values. And the other one is a label propagation method, where the unknown image is labeled with the label of the closest previously labeled face. Therefore, at each iteration, a node is labeled; until there are no more nodes left to be labeled. In the following the details of these two methods are explained.

1) Average of the labeled similarity values: For every node, we find a *1xn* resulting similarity vector, *ps(V)*, also the index numbers of the pre-labeled nodes, which are formed to be the training set elements, are kept. Therefore, the average of the similarity values at these indices of the resulting similarity vector are calculated for each category. The category of the highest value among these averages, is assigned as the corresponding node's category. Once the unknown node is labeled, it is added to the labeled data set. This procedure will continue, until there is no more unlabeled data left. The algorithm for the labeling procedure is given in Figure 4.

Input:
ts: the training set for the images
Let *ts(i)* be the training set for the i^{th} category

size(ts): size of the training set for one category

ps: the resulting vector of a node
k : no of categories

Output:

label: label for the corresponding node

/* avg: (1 x *k*) matrix to hold the elements of the average values for each category */

for each *i* from *1* to *k*

$$avg(i) = \frac{\sum_{i=ts(1)}^{ts(10)} ps(i)}{size(ts)}$$

end for

label is assigned as the index of the maximum element of *avg*

Figure 4. The algorithm for the labeling process with average of labeled data method .

2) Label Propagation: Here again, the labeled node indices, in other words the indices of the training set elements, are kept. In every iteration, each node is labeled

with the label of its closest neighbor, and then the labeled data is updated. In Figure 5 the algorithm for label propagation is given.

```
Input:

labels : 1xn matrix to hold the labels for each
face(nodes), initially only the pre-determined
indices are labeled.

psAll : psAll(i) is the resulting vector( ps(V) )
for the i^th node

Output:

labels: 1xn matrix with all entries labeled

for each i from 1 to n
   if i is not labeled then
      labels(i) = max index of psAll(i)
   end if
end for
```

Figure 5. The algorithm for the labeling process with label propagation method

IV. Experiments And Naming

In our experiment, 20 categories of politician names are used for constructing the dataset. The images to form the dataset are gathered from Yahoo! and Google web image search engines for the queries, *Angela Merkel*, *Ariel Sharon, Ban Ki Moon, Basescu Traian, Benjamin Netanyahu, Blair, Bush, Chirac, Colin Powel, Gerard Schroder, Giorgio Napolitano, Gordon Brown, Hugo Chavez, Junichiro Koizumi, Kofi Anan, Nicolas Sarkozy, Obama, Pascal Couchepin, Putin, Silvio Berlusconi.* Figure 6 illustrates the examples from the 20 categories.

Figure 6. The categories used in the experiments

45 frontal face images for each category are selected among the query results. From each category 10 random images are selected for the training set, and the rest of the data is used for test. In order to get an average success result, 10 fold cross validation is applied. That is the procedure of randomly selecting 10 examples for constructing training set, for each category and then running the program is performed 10 times. The minimum, average and maximum success rates are recorded for evaluation. Success rates are found as the ratio of the number of correctly labeled faces to the number of all faces the faces in that category.

Empirically, the restart probability, c is selected as 0.9 and the L1 norm distance for the convergence limit is determined to be 0.1.

Two methods for labeling are used as explained in Section (III.E). For the first method, the maximum, average and minimum success rates obtained from 10 different experiments with 10 random training sets for each category, are shown in Figure 7.

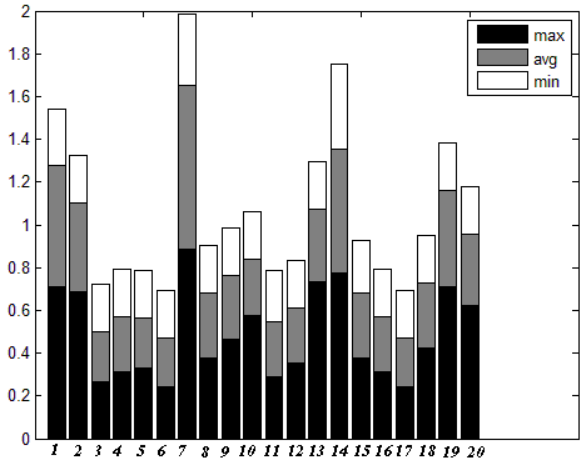

Figure 7. Bar Chart belonging to minimum, average and maximum success rates for the *Average of Labeled Data* approach.

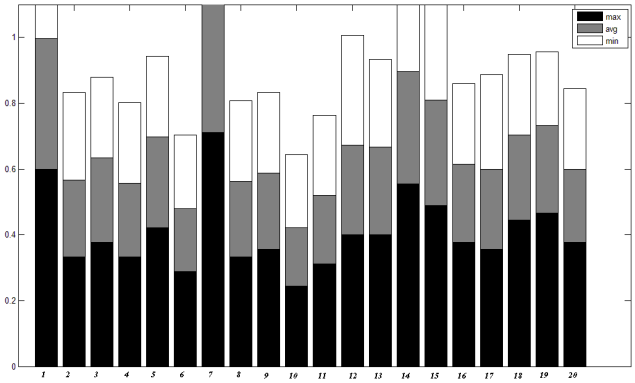

Figure 8. Bar Chart of minimum, average and maximum success rates for the *Label Propagation* approach.

Figure 8 illustrates the maximum, average and minimum success rates for the second method.

In both approaches, it can be observed that the first, seventh and the fourteenth categories, which are the faces of *Angela Merkel, George W. Bush* and *Junichiro Koizumi,* have the highest performance. In Figure 9, some of the correctly labeled *George W. Bush* images are illustrated. Here, out of 45 *George W. Bush* images, 38 of them are labeled correctly, and this gives us 84.44% of a success rate. Hence, the remaining 7 images from *George W. Bush* dataset are incorrectly labeled, and those are the images shown in Figure 10.

Figure 9. Correctly labeled *George W. Bush* images

Figure 10. *George W. Bush* images with incorrect labels

Figure 11. Correctly labeled *Angela Merkel* images

Figure 12. Incorrectly labeled *Angela Merkel* images (average of labeled data approach).

In Figure 11, fifteen of the correctly labeled images of *Angela Merkel* are shown. In this experiment, the maximum of the average labeled data method is used, and the success rate is calculated as 71.11%; 32 of the 45 images are labeled correctly. Other 13 images of *Angela Merkel* which are confused with different people are shown in Figure 12. Observing both the *Angela Merkel* and *George W. Bush* samples given in the figures above, the results of the false labeling can be interpreted as the selected test images to be inappropriate for face detection. Comparing the correctly named images (Figure 9 and Figure 11), with the images that are labeled incorrectly (Figure 10 and Figure 12), the images with incorrect labels are not as frontal images as the correctly labeled ones. Also the sizes of the wrong labeled images are smaller than the other images, which results in having low quality facial features. Moreover, some of the images are too dark to obtain correct facial features. All the correctly named images, on the other hand, are frontal images of the corresponding person, and the sizes of the images are better to have high quality feature extraction.

For one of our experiments where the 10 training set has been selected randomly, we get an overall success of 35% when average of labeled data method is used. The corresponding confusion matrix is given in Figure 13.

V. DISCUSSION

The two methods for labeling do not make a considerable difference in the overall result. In the *Average of Labeled Data* method, the overall success is found to be 34,69%, whereas, in the *Label Propagation* method, 33,98% is the overall success result. Although these results seem to be low compared to the current face recognition techniques, we propose a method for recognizing the images that are not taken in a controlled environment. In our experiments, *10-fold cross validation* is used with 10 training examples in each run. Throughout our experiments, we have recorded that the lowest success rate for any category is found to be 22%, while the highest success is 89%. Highest success rates come out of the categories which have more frontal images in their dataset. Figure 10, explains this situation clearly. The wrong labeled images of *George W. Bush* are

either too small in size, or too dark, and most of them cannot be even considered as frontal images. Besides, considering that we have 20 categories, while the chances to achieve a correct labeling result for an image randomly is 5%, this method improves the success rate up to 35% in overall data. Random walk with restart method is used rather than simply using the similarity graph, because recent research explores that random walk with restart method helps emphasizing the strength or weakness of the weights of the edges. With this iterative method, the edges of the graph with high similarity weights become stronger; meanwhile the edges with low similarity weights tend to get even weaker. Therefore, this method helps to achieve a classification with more reliable results.

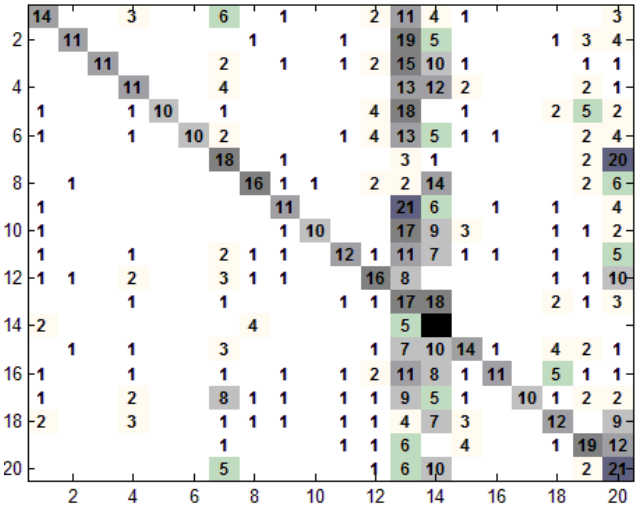

Figure 13. Confusion Matrix

VI. CONCLUSION

In this study, we try to recognize the unknown faces on the photographs using the known faces. Using a graph based method, the possibility of each image being in the same category with other images has been found. In order to label the unknown data, two methods have been used. One of them is labeling the unknown data with the label of the maximum value among the average similarity value calculated for the indices of labeled nodes for each category, at the result vector. The other one is to label the unknown data with its nearest labeled node. As a graph method, random walk with restart is used; having an outcome of strengthening the ability to classify the data. This method converts the similarity graph into a graph that the similar nodes become more strongly bound, and the nodes with weak bindings become weaker. Hence, the classification becomes easier and more reliable.

ACKNOWLEDGMENT

This research is partially supported by TUBITAK Career Project, grant number 104E065. We would like to thank Dr. Tolga Can for his valuable discussions.

REFERENCES

[1] J. Yang, M.-Y. Chen, and A. Hauptmann., "Finding person x: Correlating names with visual appearances.", *In IEEE Conf. on Computer Vision Pattern Recognition (CIVR'04), Dublin 2004.*

[2] N. Ikizler and P. Duygulu. "Person search made easy. ", *In The Fourth International Conference on Image andVideo Retrieval (CIVR 2005), Singapore, 2005.*

[3] T. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y.- W. Teh, E. Learned-Miller, and D. Forsyth. "Faces and names in the news.", *In IEEE Conf. on Computer Vision Pattern Recognition (CVPR, 2004.)*

[4] C Liu, S Jiang, Q Huang, "Naming faces in broadcast news video by image google", *ACM Multimedia Conference, 2005*

[5] D. Lowe. "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision(*IJCV), *60(2):91–110,2004.*

[6] D. Ozkan and P. Duygulu. "A graph based approach for naming faces in news photos." *In IEEE Conf. on Computer Vision Pattern Recognition (CVPR), pages 1477–1482, 2006.*

[7] S. Satoh and T. Kanade. Name-it: Association of face and name in video. *In IEEE Conf. on Computer Vision Pattern Recognition* (CVPR), 1997.

[8] M. Guillaumin, T. Mensink, J. Verbeek, C Schmid. "Automatic Face Naming with Caption-based Supervision" *In IEEE Conf. on Computer Vision Pattern Recognition (CVPR, 2008)*

[9] D.D. Le, S. Satoh, "Unsupervised Face Annotation by Mining the Web", *2008 Eighth IEEE International Conference on Data Mining*

[10] Y. Su, S. Shan, X. Chen and W. Gao., "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition." , *In Proc. of IEEE Int'l Conf. on Computer Vision, 2007.*

[11] T Can, O. Çamo□lu, A K Singh, "Analysis of Protein-Protein Interaction Networks Using Random Walks" , *ACM Multimedia Conference, 2005*

[12] M. Everingham, J. Sivic, and A. Zisserman. " 'Hello! My name is... Buffy' – automatic naming of characters in TV video". *In BMVC, pages 889–908, 2006.*

[13] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. "Face recognition:A literature survey*", ACM Computing Surveys, 35(4):399–458, 2003.*