Pose-Independent Identity-based Facial Image Retrieval using Contextual Similarity

Islam Almasri

King Abdullah University of Science and Technology 4700, Thuwal, Saudi Arabia

Abstract. Pose-independent face recognition is a challenging computation problem; this is because the visual variation in between two images of the same person with different face orientation can be much greater than that between two images of two different people with the same pose. We devise a contextual algorithm to improve the retrieval rates on such databases, and obtain significant improvements of 96% up from 84% in terms of AUROC on the Sheffield Face Database.

1 Introduction

A number of techniques have been developed to address the problem of shape deformations in objects in an image for image matching, a number of these have been proposed and tested for simplified silhouettes of shapes such as MPEG [1], Kimia [2], Swedish Leaf [3], Natural Silhouette database [4] and Gesture Database [5]. Because of the simplicity of this kind of images, the corresponding proposed algorithms perform favorably in identifying images depicting the same object despite of significant deformation and geometric transformations.

For other images with more complicated objects, the shape identification algorithms tend to become more application specific, an example of this is the facial detection problem where the objective is to identify if a certain image contains faces, and if so, detecting their location in the image, this problem is important in the applications such as human computer interaction [6] and video surveillance [7], in this field it is important to detect these faces despite noise in the pictures and different orientations of the face.

This problem becomes more pronounced in the facial recognition problem where the identity of the person in the picture is to be determined; this is done by comparing the detected face against a database of known faces and extracting the most visually similar one. In this regard, a number of algorithms have been proposed for image recognition that seek to identify faces across different poses [8, 9], lighting conditions [10, 11], age [12], facial expression [13] and make up [14], however, the problem of a pose-invariant face recognition is identified as one of the more challenging among these variations [8, 15].

To further enhance the image retrieval problem, people have been looking into what is called contextual information of the images[16], the idea is to use not only the mutual distance between objects separately, but consider all distances between objects as a whole. Yang et al.[16] developed a contextual algorithm

based on label propagation which improved the retrieval rates on object benchmark databases considerably, later, more algorithms were devised to consider the context in different ways [17, 18]. Jegou et al.[19] benefited from the contextual information to provide more relevant results for image search.

In this paper we test the prospects of using context aware algorithms to improve the retrieval of facial images beyond conventional means of feature extraction. We develop an iterative knn ranking approach (ICS) that considers the context to provide more accurate relevant results for facial image retrieval. We test the technique on pose variant facial images which is one of the more challenging in facial recognition, the retrieval rates improve considerably from 84% to 96% in terms of AUROC on the Sheffield Face database [20]. We also test other contextual algorithms on the same database and show that it provides similar enhancements.

2 Pose-Independent Face Recognition

Face recognition algorithms can be used to for matching of a query image to a database of images of known people; this is to identify the person in the pictures as the closest match (or one of the closest matches), such as in surveillance cameras. Another use of these algorithms is to query a database of unlabeled images to extract all visually similar ones, ideally, the all the extracted images are of the same person in the query image, like the online face searches or the clustering of images of people sampled from a video feed.

Because of the various fields of face recognition, the algorithms are required to work on pictures in uncontrolled conditions, these conditions result in multiple variations in the faces depicted in the pictures such as lighting, pose, and facial expressions, also, depending on the time difference of the pictures, facial hair, age and makeup. The variation in pose proved to be one of the more important problems that are difficult to solve, this is because of the average visual variations in different poses are larger than those under other varying conditions [8, 15].

Face recognition algorithms that deal with different pose variations belong to 3 categories [8]. In first category are those generic face recognition algorithms that are meant to recognize faces across different variations not just the pose, those include the EigenFaces [21], FisherFaces [22], LEMs [23] and LBP [24]. These algorithms, however, are very sensitive to pose variations [8]. The other two categories are algorithms designed for pose variations based on 2D and 3D techniques. These techniques attempt to model and compensate for the variation in pose.

The algorithms designed to be pose invariant, or try to compensate for the post variation in during the image matching stage, are designed (and ideal) for the applications that require finding the most similar image from a set of labeled images, this is to assigned a label (the person) to the new image. However, they can still be used for other applications, such that those that require the finding the most relevant subset of pictures from a database of unlabeled images of various people in different poses. But in those cases, these algorithms have

shortcomings this is because the variation between images of the same person in different poses can be much larger than the variation between images of different people in the same pose.

3 Contextual Information

The shape retrieval problem is usually benchmarked by using a query image depicting a certain object, then searching a database of images, and retrieving images that are believed to depict the same object as the query image. The quality of the shape retrieval is measured by the successful retrieval rates like the bulls eye rate defined in [17], or more generally the area under the Receiver Operating Characteristic Curve (AUROC). Examples of these benchmark databases include the MPEG7 [1], Kimia [2], Swedish Leaf [3], Natural Silhouette database [4] and Gesture Database [5].

To retrieve the matching images a distance metric is required to quantify the similarity (distance) between the query image and each of the images in the database, then ranking them and retrieving the higher ranking results. These metrics attempt to first extract the important features of the object depicted in the image, and then those features are compared between the two images, those features are called shape descriptors.

There are two main categories of shape descriptors: Contour based and Region based [25]. Contour based descriptors extract the a number of points to describe the outline of the object and they differ in how they describe these points and how to compare these outlines between two images, examples include Shape Context (SC) [26], IDSC [27], shape tree [28] and Height Function [25]. Region based descriptor extract the pixel information in the shape; these descriptors capture information inside the shape itself unlike the contour descriptors which only hold information about the exterior. Examples of these descriptors include Fourier descriptors [29] and Zernike moments [30].

Good shape descriptors improve the retrieval rates for shape retrieval, trying to obtain images depicting the same object regardless of any deformations or geometric transformations in the object itself, however, researchers have looked into techniques to improve those further beyond shape descriptors. Those techniques are context sensitive techniques that exploit not only the mutual distance between the search query and each of the images in the database, but also the distances between the images in the database itself. Examples of these techniques are LP [16], LCDP [17], CDM [19] and Kontschieder et al. method [18].

Those context-aware techniques improve the retrieval rates of shapes because they capture information beyond the visual similarity between two images. For example, if the query image A is visually similar to image B, and B is similar to image C, then A and C are more likely to depict the same object, even if they are visually dissimilar, which might be attributed to a big deformation in the object that the descriptor cannot capture. For example, see Figure 1 from [16], which is result of a query on the MPEG7 dataset, the second shows the results of the context aware algorithm in fetching the first 10 similar results from the

dataset. It manages to identify more images including bugs in them similar to the query image.

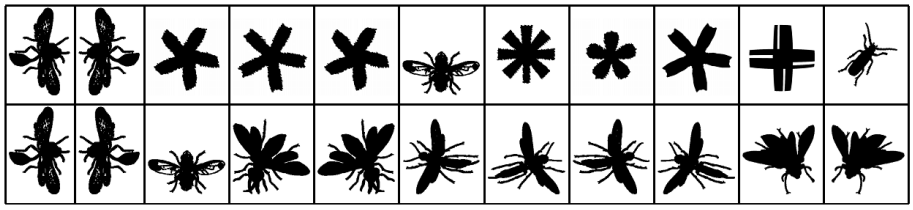

Fig. 1. Retrieval Results on shapes with (bottom) and without (top) context. (Picture Courtesy of [16]).

The context-aware techniques substantially improve the retrieval rates; for example, the LCDP [17] techniques improve the retrieval on the MPEG7 dataset to 93% up from 85% using the IDSC descriptors. These techniques also have the advantage of being general, and can be applied in combination of any shape descriptor or image similarity measure. To make use of this idea, Jegou et al. [19] applied the context idea to another application which is content-based image retrieval; in their paper they propose a context-aware algorithm (CDM) that can improve the ranking of image search results. They test their approach on the N-S dataset [31] and images sampled from different video feeds, these are not simple silhouettes, but rather full resolution complex images showing different objects from different views. See figure 2 that demonstrates the improvement in retrieval.

Fig. 2. Retrieval Results on Images with (top) and without (bottom) context (Image Courtesy of [19]).

4 Contextual Similarity for Identity-based Image Retrieval

The success of context aware algorithms in improving both shape retrieval and content-aware image search is because of certain similarities between the fields. In both a feature extraction method is used to describe the contents of a query

image, then the same method is applied to images in a database, and what is required is to extract the set of images that are most similar to the query image but not according to the visual data only, but to the content of the image. For this reason, we extend the use of Contextual Similarity for identity-based image retrieval, where an image of a person is compared to a database of unannotated images of various faces in different poses. The goal is to identify the images that show the same person regardless of the pose. For this purpose use the Sheffield Face Database and the EigenFaces technique (which is the most widely used in face recognition [8]) to extract the visual information, then we used two techniques for contextual similarity; the technique described in Kontschieder et. al [18], and our own ranking technique (ICS) described in the next section. Figure 3 shows the ROC curve for the retrieval results with and without context information.

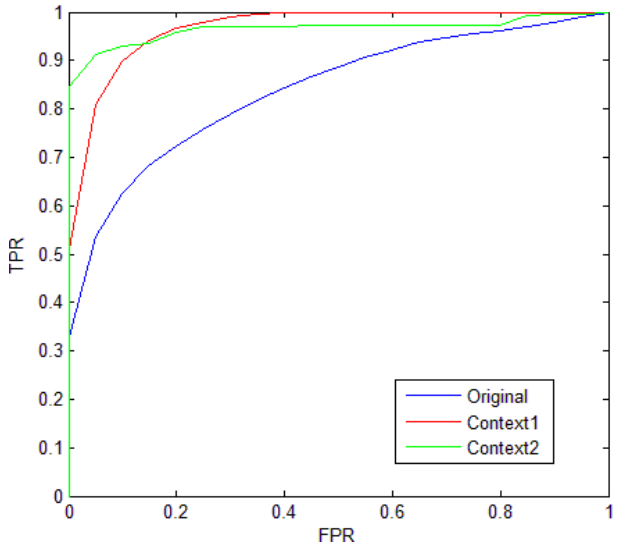

Fig. 3. ROC curves for retrieval on the pose database, with (red (Kontschieder et al.) and green (ICS)), and without (blue) using context information

Table 1. Retrieval Results (AUC).

EigenFaces	EigenFaces+ Kontschieder	EigenFaces+ICS
83.93%	96.74%	96.48

The retrieval results improve substantially from AUROC 83.9% to 96.7% (see Table 1) using contextual similarity. The difference in results is demonstrated in Figure 4, which shows the retrieval results of an image of one person with and without the contextual similarity. For this example ICS was used for context enhancement. It can be seen that using the contextual information enables the retrieval of images of the same person in poses that are very different from the query image. Without the use of context, the images of different people are visually more similar to the images of the same person in two very different poses. As you can see the 5th retrieved result is an image of a different person but it's visually more similar than the image of the same person with more pose angle.

Fig. 4. The query image (first column) and the retrieval results of the best 10 matching images before (first row) and after (second row) including context information using ICS, notice that the other people other than the one in the original image for the query are blurred out because we're not authorized to publish them.

5 Iterative Contextual Similarity

It can be seen from the results in Figures 1, 2 and 4 that the improvement in retrieval, especially for images very visually different from the query image, is partially due to the existence of intermediary images between them. In this section we show that the advantage of contextual similarity can also be achieved using a simple iterative ranking approach that we call the Iterative Contextual Similarity. First, assume we have N images, and let X^0(i, j), i=1,2,..,N, j =1,2,..,N be the initial distance matrix, where X(i,j) is the distance (dissimilarity) between the two images numbered i and j, defined using any of the image similarity techniques. Let a be the query image. We define an iterative algorithm, in each iteration, the image the most similar to the query image is selected and is added to the set S^T, which is initially empty (S^0={}):

$$b^T = \ argmin\left\{X^T\left(a,\ j\right)\right\},\ j \neq a,\ j = 1, 2, \ldots, N,\ j \notin S$$

$$S^T = \ S^{T-1}\ \cup b^T$$

This is the original definition of the KNN algorithm, now our suggested addition is that the after the selection of each image, the distances of all the

other images are updates according to the new selected image b^T:

$$X^T(a,j) = f\left\{X^{T-1}(a,j),\ \ X^{T-1}(b^{T-1},\ j)\right\},\ j \neq a,\ j = 1,2,\ldots,N,\ j \notin S$$

Where f is the combining function, in the first iteration, the distance of each node to the query image, will be updated according to the distance to the new selected first hit, the image b, as defined by the combining function. The combining function can be tailored according to the problem characteristics, for example, the Sheffield Face Database, it was chosen to be a simple minimum function:

$$X^T(a,j) = min\left\{X^{T-1}(a,j),\ \ X^{T-1}(b^{T-1},\ j)\right\},\ j \neq a,\ j = 1,2,\ldots,N,\ j \notin S$$

By using the minimum function, the resulting behavior is that, each iteration, the image that is selected is the image most similar to any of the images already matched to the query image or the query image itself. The testing results for this algorithm on the faces database is shown in Figure 3 in green, and the AUC in Table 1, despite this very simple approach, the results are comparable to the increase of performance of the more sophisticated context similarity techniques.

Table 2. Bulls eye rates on MPEG7 dataset.

IDSC	**IDSC+ICS**	HF	**HF+ICS**
85.7%	**87.4%**	89.6%	**91.1%**

We also test this new iterative procedure against well-known silhouette image databases. Table 2 shows the overall improvement in the retrieval results on the MPEG7 dataset, based on two different shape descriptors; the IDSC [27] and the Height Function [25]. The combining function chosen for those two dataset was an averaging function with a decaying weight.

6 Conclusion

Context-aware algorithms achieve great improvements of retrieval rates over conventional mutual distance techniques. In this paper we show how such algorithms can be used for applications beyond the simple shape database, giving significant improvements in the pose-independent face image retrieval which is one of the more challenging problems in facial recognition.

We also show that the benefit from contextual information can be achieved with alterations to conventional knn for ranking, simple yet flexible enough technique to adapt to different applications. The use of such contextual techniques is a promising prospect for other Image Retrieval fields to improve them beyond visual similarity.

References

1. Latecki, L.J., Lakamper, R., Eckhardt, T.: Shape descriptors for non-rigid shapes with a single closed contour. In: Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on. Volume 1., IEEE (2000) 424–429
2. Sharvit, D., Chan, J., Tek, H., Kimia, B.B.: Symmetry-based indexing of image databases. In: Content-Based Access of Image and Video Libraries, 1998. Proceedings. IEEE Workshop on, IEEE (1998) 56–62
3. Söderkvist, O.: Computer vision classification of leaves from swedish trees. (2001)
4. Gorelick, L., Galun, M., Sharon, E., Basri, R., Brandt, A.: Shape representation and classification using the poisson equation. Pattern Analysis and Machine Intelligence, IEEE Transactions on **28** (2006) 1991–2005
5. Petrakis, E.G.M., Diplaros, A., Milios, E.: Matching and retrieval of distorted and occluded shapes using dynamic programming. Pattern Analysis and Machine Intelligence, IEEE Transactions on **24** (2002) 1501–1516
6. Bartlett, M.S., Littlewort, G., Fasel, I., Movellan, J.R.: Real time face detection and facial expression recognition: Development and applications to human computer interaction. In: Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on. Volume 5., IEEE (2003) 53–53
7. Collins, R.T., Lipton, A., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P., et al.: A system for video surveillance and monitoring. Volume 2. Carnegie Mellon University, the Robotics Institute Pittsburg (2000)
8. Zhang, X., Gao, Y.: Face recognition across pose: A review. Pattern Recognition **42** (2009) 2876–2896
9. Gross, R., Baker, S., Matthews, I., Kanade, T.: Face recognition across pose and illumination. In: Handbook of Face Recognition. Springer (2005) 193–216
10. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. Pattern Analysis and Machine Intelligence, IEEE Transactions on **23** (2001) 643–660
11. Lee, K.C., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. Pattern Analysis and Machine Intelligence, IEEE Transactions on **27** (2005) 684–698
12. Anastasi, J.S., Rhodes, M.G.: Evidence for an own-age bias in face recognition. North American Journal of Psychology (2006)
13. Chang, K.I., Bowyer, W., Flynn, P.J.: Multiple nose region matching for 3d face recognition under varying facial expression. Pattern Analysis and Machine Intelligence, IEEE Transactions on **28** (2006) 1695–1700
14. Chen, C., Dantcheva, A., Ross, A.: Automatic facial makeup detection with application in face recognition. In: Biometrics (ICB), 2013 International Conference on, IEEE (2013) 1–8
15. Tan, X., Chen, S., Zhou, Z.H., Zhang, F.: Face recognition from a single image per person: A survey. Pattern Recognition **39** (2006) 1725–1745
16. Yang, X., Bai, X., Latecki, L.J., Tu, Z.: Improving shape retrieval by learning graph transduction. In: Computer Vision–ECCV 2008. Springer (2008) 788–801
17. Yang, X., Koknar-Tezel, S., Latecki, L.J.: Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 357–364

18. Kontschieder, P., Donoser, M., Bischof, H.: Beyond pairwise shape similarity analysis. In: Computer Vision–ACCV 2009. Springer (2010) 655–666
19. Jegou, H., Schmid, C., Harzallah, H., Verbeek, J.: Accurate image search using the contextual dissimilarity measure. Pattern Analysis and Machine Intelligence, IEEE Transactions on **32** (2010) 2–11
20. (http://www.shef.ac.uk/eee/research/iel/research/face)
21. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of cognitive neuroscience **3** (1991) 71–86
22. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Pattern Analysis and Machine Intelligence, IEEE Transactions on **19** (1997) 711–720
23. Gao, Y., Leung, M.K.: Face recognition using line edge map. Pattern Analysis and Machine Intelligence, IEEE Transactions on **24** (2002) 764–779
24. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on **28** (2006) 2037–2041
25. Wang, J., Bai, X., You, X., Liu, W., Latecki, L.J.: Shape matching and classification using height functions. Pattern Recognition Letters **33** (2012) 134–143
26. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. Pattern Analysis and Machine Intelligence, IEEE Transactions on **24** (2002) 509–522
27. Ling, H., Jacobs, D.W.: Shape classification using the inner-distance. Pattern Analysis and Machine Intelligence, IEEE Transactions on **29** (2007) 286–299
28. Felzenszwalb, P.F., Schwartz, J.D.: Hierarchical matching of deformable shapes. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE (2007) 1–8
29. Zhang, D., Lu, G.: Generic fourier descriptor for shape-based image retrieval. In: Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on. Volume 1., IEEE (2002) 425–428
30. Kim, W.Y., Kim, Y.S.: A region-based shape descriptor using zernike moments. Signal Processing: Image Communication **16** (2000) 95–102
31. (http://vis.uky.edu/ stewe/ukbench/)