International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 4, Issue 4, July - August 2015 ISSN 2278-6856

A Survey of Similarity Measures in Web Image Search

Yossra H. Ali[1] , Wathiq N. Abdullah[2]

[1] University of Technology, Computer Sciences Department,
Sinaa'a Street, Baghdad, Iraq

[2] University of Technology, Computer Sciences Department,
Sinaa'a Street, Baghdad, Iraq

Abstract

The rapid development of internet has brought explosive growth in information of World Wide Web. Searching for images of people is an essential task for image and video search engines. However, current search engines have limited capabilities for this task since they rely on text associated with images and video, and such text is likely to return many irrelevant results. Large-scale face image retrieval is the enabling technology behind the next generation search engines (search beyond words), by which web users can do social search with personal photos. This paper surveys various methods of similarity measurements used for image retrieval techniques and face recognition applications. Each method is differentiated with other surveyed method and comparative measures of methods are presented which provides the significance and limitations of web image retrieval techniques with correspond to query.

Keywords: similarity Measures, Hamming distance, Image search.

1. INTRODUCTION

With the rapid growth of digital images on Internet, images have been increasingly used in content expression[19]. Image search engines are currently dependent on textual metadata. This data can be in the form of filenames, manual annotations, or surrounding text. However, for the vast majority of images on the Internet (and in peoples' private collections), this data are often ambiguous, incorrect, or simply not present [14].

Technically, it is very challenging to find a person from a very large or extremely large database which might hold face images of millions or hundred millions people. A highly efficient image retrieval technology is needed[8].

In the actual Internet application, there are a lot of demands for face image search, such as searching celebrities, criminals, and popular photos. Hence, it becomes an important issue for information sharing about how to retrieval face images available on web effectively and accurately. Traditional search engines only rely on keyword retrieval on the texts around the images with lower accuracy. Much progress has been made in content-based image retrieval. There are also a small amount of content-based general image retrieval systems[19].

Huge efforts have been devoted to face recognition technology and remarkable results, noticed. Such advances will provide us the possibility to build a new generation of search engine: persons photo fetching [8].

At the same time, the developments of face detection and recognition technologies become relatively mature. There are some face recognition systems for users to search face image by uploading a query image. The related knowledge about face detection and recognition can help to provide more accurate information in the images. If the materials can be combined together to retrieve face image, it will be more effective and accurate [19]. Figure (1) shows an image search system overview.

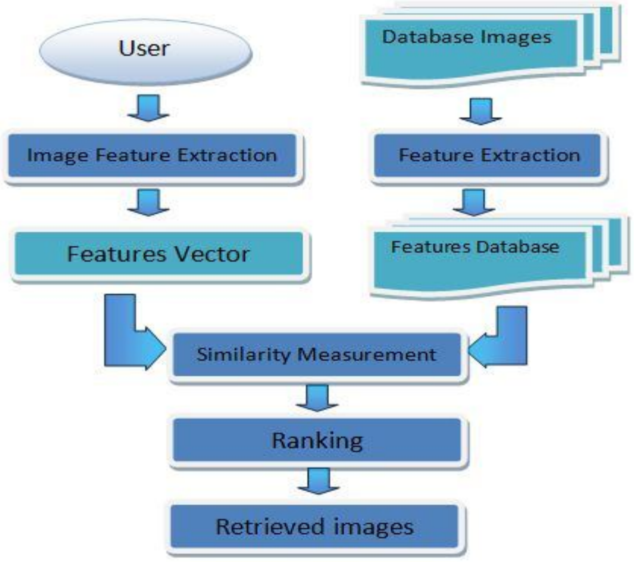

Figure 1 Image Retrieval System

One way to improve the retrieval performance is to take into account visual information present in the retrieved faces. This task is challenging for the following reasons:

- Large variations in facial appearance due to pose changes, illumination conditions, occlusions, and facial expressions make face recognition difficult even with state-of-the-art techniques [15,20,12]
- The fact that the retrieved face set consists of faces of several people with no labels makes supervised and unsupervised learning methods inapplicable[5].
- This paper surveyed several image retrieval systems focusing on image similarity measures.

2. SIMILARITY MEASURES

Similarity measures are typically used for quantifying the affinity between objects in search operations, where the user presents an object (query) and requests other objects “similar” to the given query. Therefore, a similarity measure is a mathematical abstraction for comparing objects, assigning a single number that indicates the affinity between the said pair of objects. The results of the search are typically presented to the user in the order suggested by the returned similarity value. Objects with higher similarity value are presented first to the user because they are deemed to be more relevant to the query posed by the user [13].

2.1 Histogram similarity

In many application histogram is used as object model .Thus to compare an unknown object with a known object we can compute the similarity between their histograms. The histogram of image shows the frequency distribution of different pixel intensities.

Z. Zhang et al. in 2000 propose a realistic approach to effectively indexing images in an image library to facilitate image retrieval is to first manually classify all the images into different "categories" or "domains", and to design search engines for each of the domains. This is referred to focused image retrieval.

Experimental results show that the proposed approaches have significantly improved the image retrieval precisions than the existing search engines in these focused application domains. In particular, the research focuses on a face detection system as well as two application areas: image retrieval of querying human beings and image retrieval of querying similar background. In first application area, it assumes that there is certain collateral textual information available accompanying the images, such as caption. only face detection is conducted to verify the existence of human beings in the image.

The paper presents an efficient face detection system. Next, this face detection system is combined with a conventional color histogram based similarity matching system to solve for the second application of image retrieval. These methods are evaluated separately.

The system uses two color features: hue and chrominance. Intensity is not used so that the proposed method can be independent of lighting condition to a certain degree. The technique applies face detection to the retrieved documents from the first round of search of the whole database based on text indexing. Also, due to this filtering, the precision is significantly improved. The system applies face detection to the image first to crop out the areas with human bodies. Then the rest of the image (i.e. the background) may be ported to the conventional similarity based matching for image retrieval based on the background similarity.

The proposed system shows two application examples of this face detection system in querying and retrieving images, one being image retrieval with indexed collateral text, and the other being image retrieval with background similarity [23].

X.Wang et al. in 2009 propose two ideas. First, it proposes to use individual bins, instead of whole histograms, of Local Binary Patterns (LBP) as features for learning, which yields significant performance improvements and computation reduction in the experiments. Second, it presents a novel Multi-Task Learning (MTL) framework, called Boosted MTL, for face verification. The effectiveness of the system is verified with a large number of celebrity images/videos from the web. The system addresses the problem of giving a small set of face images of a celebrity, verify whether the celebrity is in other images and videos that are returned by the search engine.

The proposed LBP bin features is compared with direct LBP based recognition and LBP histogram feature based AdaBoost. The LBP bin feature based AdaBoost approach significantly outperforms the approach which compares the distances of local histograms of LBP. It also outperforms the similarly trained AdaBoost classifier based on LBP histograms of local regions.

Experimental results show that when using local histograms as features AdaBoosting can marginally improve the performance compared with directly using LBP, but it is much worse than using counts of individual bins as features. It significantly outperforms the approach that directly uses LBP for face verification in terms of both accuracy and speed [21].

2.2 Similarity Matrix

A similarity matrix is a matrix of scores that represent the similarity between a number of data points. Each element of the similarity matrix contains a measure of similarity between two of the data points.

L. Gu et al. in 2007 approach achieved promising accuracy and proposed a totally automated approach for organizing consumer photos based on the combination of unsupervised face clustering with supervised face model training. Based on this, as well as contextual information, a semi-supervised agglomerative clustering is conducted, and the collection is divided into groups by face. Then, larger clusters are modeled as frequently appearing people. Finally, clusters are consolidated by matching faces with each of these face models. Especially, faces in smaller clusters are merged into larger clusters. Many small clusters are merged into large clusters and the average recall rate is improved substantially with only minor degradation in the precision rate.

Faces are detected in every picture and a semi-supervised clustering algorithm is employed on the facial similarity matrix to group faces into clusters while at the same time incorporating spatial constraints. Dominant clusters are modeled as significant people and small clusters are recognized against them to further improve the grouping performance. Promising results have been achieved in consumer photo datasets with high accuracy and fast speed[10].

2.3 Visual Similarity

Visual Similarity Search retrieves images that "look similar", regardless of the image content. Retrieved images have similar visual attributes as the chosen images, such as comparable colors, textures, and shapes. The retrieval does not use knowledge about the actual subjects in the images.

D. Le and S. Satoh in 2008, propose a method for retrieving relevant faces of one person by learning the visual consistency among results retrieved from text correlation-based search engines. The method consists of two steps. In the first step, each candidate face obtained from a text-based search engine is ranked with a score that measures the distribution of visual similarities among the faces. The second step improves this ranking by treating this problem as a classification problem; and the faces are re-ranked according to their relevant score inferred from the classifier's probability output.

The system plans a general framework to boost the face retrieval performance of text-based search engines by visual consistency learning. The framework seamlessly integrates data mining techniques such as supervised learning and unsupervised learning based on bagging. A comprehensive evaluation on a large face dataset of many people was carried out and confirmed that our approach is promising. The system framework includes an estimation the ranked list of these faces using Rank-By-Local-Density-Score, and improve this ranked list using Rank-By-Bagging-ProbSVM.

The technique uses the idea of density-based clustering described by to solve this problem. Specifically, the local density score (LDS) of a face is defined as the average distance to its k-nearest neighbors. The faces retrieved from the different name queries were merged into one set and used as input for ranking. The retrieval performance is evaluated with measures that are commonly used in information retrieval, such as precision, recall, and average precision. The approach works fairly well for well-known people, where the main assumption that text-based search engines return a large fraction of relevant images is satisfied [5].

N. Kumar et al. in 2011, focus on images of faces and the attributes used to describe them. Examples of face attributes include gender, age, jaw shape, nose size, etc. The system shows how one can create and label large data sets of real-world images to train classifiers which measure the presence, absence, or degree to which an attribute is expressed in images. These classifiers can then automatically label new images.

Prior to feature extraction, the background is masked out to avoid contaminating the classifiers. Different types of information can be extracted. The types of pixel data to extract include various color spaces (RGB, HSV) as well as edge magnitudes and orientations. To remove lighting effects and better generalize across a limited number of training images. All types of low-level features could simply be extracted from the whole face, and let a classifier figure out which are important for the task and which are not. The approach illustrates how attribute-based face verification is performed on a new pair of input images. In all experiments, not only are the images in the training and test sets disjoint, but there is also no overlap in the individuals used in the two sets. In addition, the individuals and images used to train the attribute and simile classifiers are disjoint from the testing sets.

Search queries are mapped onto attribute labels using a dictionary of terms. This approach is simple, flexible, and yields excellent results in practice. Furthermore, it is easy to add new phrases and attributes to the dictionary, or maintain separate dictionaries for searches in different languages [14].

2.4 Cosine Similarity Measure

Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. Cosine similarity is particularly used in positive space, where the outcome is neatly bounded in [0,1]. The inner product of the two vectors (sum of the pairwise multiplied elements) is divided by the product of their vector lengths. This has the effect that the vectors are normalized to unit length and only the angle, more precisely the cosine of the angle, between the vectors accounts for their similarity [11].

In 2008, Robbie Lamb & Rafal Angryk and Piet Martiens s' system discusses solar images and the results of our investigation of techniques that can be used to identify solar phenomena in images from the TRACE satellite for the creation of a search engine. The images are first segmented into smaller regions and texture information is extracted. The segmentation technique breaks the image into 128 by 128 pixel blocks for extracting features. This technique is called Grid Segmentation.

The values in the attribute vector reflect different types of texture information extracted from the intensity of the images and sub-images. Representation of images in the form of attribute vectors also allows us to evaluate the similarity of two images by using a cosine similarity measure to calculate the angle between two image vectors.

The Information Retrieval module is responsible for analyzing the submitted sample image(s) and retrieving similar images from the TRACE image repositories.

Distinct features are extracted from the sample image(s) during Image Preprocessing. Classification of the sample image(s) is performed based on the extracted information. After each sample image has been classified, we select similar images from the data catalogs related to the query and order them using a cosine similarity function with the extracted image vector.

The Searching & Ranking component uses the classification results and extracted attributes from the images to quickly select and rank images that are similar to the query. The results show that for a given query image, we can expect more than half of the returned images to be relevant. The accuracy of our systems seems to be phenomenon specific, as some solar phenomena have more distinctive features than others [16].

In 2015, C. Ding, C. Xu, and D. Tao propose a novel face identification framework capable of handling the full range of pose variations. The proposed framework first transforms the original pose-invariant face recognition problem into a partial frontal face recognition problem. A robust patch-based face representation scheme is then developed to represent the synthesized partial frontal faces. The face matching is performed at patch level rather than at the holistic level. Directly matching faces in different poses is difficult. One intuitive solution is to conduct face synthesis so that the two facial images can be compared in the same pose.

The scheme of flexible Patch-based Partial Representation (PBPR) and a learning method called Multi-Task Feature Transformation Learning (MtFTL) are proposed, where the length of face representation is related to the pose of the face. The system matches a probe face image of one pose type to a frontal gallery face image. To calculate the similarity of each patch pair, the cosine metric is utilized to and the similarity scores of all patch pairs are fused by the sum rule.

Although the adopted matching scheme is simple compared to existing methods [3], [2], it is still expected that the proposed PBPR-MtFTL framework will achieve stronger performance, since the recognition ability of PBPR-MtFTL has been enhanced by exploiting the correlation between poses.

Similarity scores between all un-occluded patch pairs are averaged as the similarity score of the face image pair [4].

2.5 Similarity Graph

A Similarity Graph is a data structure that can be used to express this dissimilarity or difference. Is a weighted graph in which vertices correspond to elements and edge weights are derived from the similarity values between the corresponding elements. Hence, the similarity graph is just another representation of the similarity matrix.[18]

H. Zitouni et al. ,in 2009, 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. A similarity graph whose nodes represent the faces and edges represent the similarities between them is then formed. A random walk algorithm is applied on this graph in order to rank the similarity of all nodes. The images to form the dataset are gathered from Yahoo! and Google web image search engines.

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, a new method is proposed for recognizing the images that are not taken in a controlled environment.

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 [6].

2.6 Hamming Distance

The Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different. In another way, it measures the minimum number of substitutions required to change one string into the other, or the minimum number of errors that could have transformed one string into the other. The normalized Hamming distance between two binary vectors is the ratio of Hamming distance to the size of the vectors; this metric only takes values in the range[0,1][7].

J. Kouma and H. Li, in 2009, found that the art face identification rate is only around 70%. The computing complexity of face identification is linearly related to the number of individuals N. For large-scale face image retrieval the efficiency of face identification is a key issue. The paper focuses on the efficiency aspects of face identification. In this technique, all face images in a gallery are transferred into lower resolution used for feature vectors (called face signatures).

The image retrieval problem is treated as a source coding problem and the rate distortion theory is used to characterize retrieval quality and retrieval speed. The approach introduces a similarity measure where the key is using normalized Hamming distance. At the retrieval phase, the. Normalized Hamming distance is performed between the query and every template. The templates is then ranked according to their distance to the query. In the experiment the ORL Database of Faces is used. In the database there are 10 different images of each of 40 distinct subjects, taken at different times, varying light conditions and facial expressions [9].

J. Kouma et al., in 2010, Developed the system in [9] and used the same signature image. The approach's contribution is 1) to treat the image retrieval problem as a source coding problem and the rate distortion theory is used to characterize retrieval quality and retrieval speed; 2) to view compression of signature images it as a typical

"Wyner-Ziv Coding" problem, which circumvents the problem that the query images are not available until we decompress the signature image. 3) to develop a distributed coding scheme based on LDPC codes to compress face signature images. The system minimizing the Hamming distance over the corresponding search area. The system uses the same ORL Database of Faces as in [9]. It is a totally new retrieval paradigm in the sense that the insight of the image coding is directly applied to image retrieval problem. Thus merging two different areas into one [8].

2.7 Euclidean (L2) Distance

Euclidean Distance is the most common use of distance. It is the square root of the sum of squared differences between coordinates of a pair of objects [17]. The Euclidean is also called the "L2 distance".

Yushi Jing et al. ,in 2013, adopt a system of a hybrid search approach in which a text-based query is used to retrieve a set of relevant images, which are then refined by the user. We propose scalable solutions to learning query-specific distance functions by 1) adopting a simple large-margin learning framework, 2) using the query-logs of text-based image search engine to train distance functions used in content-based systems. The query-specific distance functions can be applied to only the most popular search queries and still service large portion of the overall search engine traffic.

The work also demonstrated that learning query-specific image distances produces more accurate measurement of image similarity than the state-of-the-art Google similar image search system. The results demonstrate that query-specific distance functions outperform the L2 distance function used in Google image search.

In a single image search session, if image xi and image xj are both clicked by the user, they are said to be co-clicked. Only images similar to the target image are selected while others are seen but ignored. Therefore if we aggregate the co-click statistics over all search sessions conducted within a sufficient period of time, then images that are clicked more often are more similar to each other. The goal is to derive reliable measurement of image similarity from such aggregated co-click statistics, and use it to train query-specific distances. The co-click statistics can also be combined with other types of distances, such as Euclidean distance (L2) derived from image features.

Using query-specific distance functions will produce more accurate image comparisons than query-independent distances. The results show that L2 distances are sufficiently accurate when two images contain the same objects or share dominant visual cues [22] .

3. Conclusions

The survey presents various methods used for image searches. Each surveyed method is significantly efficient in image retrieval process and ranking of images. This paper shows the similarity measurement method used in each system. The efficiency of the surveyed method can be measured in terms of retrieval accuracy and computational time. The merits of each method can be taken into account and further these techniques can be enhanced for large scale web image searches and re-ranking mechanism efficiently.

References

[1] A. Holub, P. Moreels, and P. Perona, "Unsupervised Clustering for Google Searches of Celebrity Images", IEEE , 2008.

[2] A. Li, S. Shan, and W. Gao, “Coupled bias–variance tradeoff for crosspose face recognition,” IEEE Trans. Image Process., vol. 21, no. 1, pp. 305–315, 2012.

[3] A. Li, S. Shan, X. Chen, and W. Gao, “Maximizing intra-individual correlations for face recognition across pose differences,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 605–611, 2009.

[4] C. Ding, C. Xu, and D. Tao, "Multi-Task Pose-Invariant Face Recognition", IEEE Transactions on Image Processing, 2015.

[5] D. Le and S. Satoh, "Unsupervised Face Annotation by Mining the Web", Eighth IEEE International Conference on Data Mining , pp 383-392, 2008.

[6] H. Zitouni, M. F. Bulut, and P. Duygulu, "Recognizing faces in news photographs on the web", , IEEE, 2009.

[7] J. Chi, M. Koyuturk, and A. Grama, "A Distributed Tool for Constructing Summaries of High-Dimensional Discrete Attributed Datasets", Proceedings of The Fourth SIAM International Conference on Data Mining, 2004.

[8] J. Kouma , H. Li, and D. Olsson, "Large-scale Face Images Retrieval: A transform coding approach", 2010, IEEE.

[9] J. P. Kouma and H. Li , "Large-scale Face Images Retrieval: *A distribution coding approach", IEEE, 2009.*

[10] L. Gu, T. Zhang, and X. Ding , " Clustering Consumer Photos Based on Face Recognition ", IEEE, 2007.

[11] M. Dittenbach ,"Scoring and Ranking Techniques - tf-idf term weighting and cosine similarity" , 2010.

[12] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, Nov 2002.

[13] M. Vlachos, " Similarity Measures", Encyclopedia of Machine Learning, pp 903-906, Springer, 2010.

[14] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Describable Visual Attributes for Face Verification and Image Search", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 10, pp 1962-1977, Oct 2011.

[15] O. Arandjelovic and A. Zisserman. "Automatic face recognition for film character retrieval in feature-length films". In Proc. Intl. Conf. on Computer

Vision and Pattern Recognition, volume 1, pages 860–867, 2005.

[16] R. Lamb , R. Angryk and P. Martiens , "An Example Based Image Retrieval System for the Trace Repository" , IEEE, 2008.

[17] R. Primicerio, M. Greenacre, "Multivariate Analysis of Ecological Data", first eddition, 2013.

[18] R. Shamir, and R. Sharan," Algorithmic Approaches to Clustering Gene Expression Data", Current Topics in Computational Molecular Biology, 2002.

[19] R. Zheng, S. Wen, Q. Zhang, H. Jin, and X. Xie, "Compound Face Image Retrieval Based on Vertical Web Image Retrieval", Sixth Annual ChinaGrid Conference, 2011.

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

[21] X. Wang, C. Zhang, and Z. Zhang, "Boosted Multi-Task Learning for Face Verification With Applications to Web Image and Video Search" , IEEE , 2009.

[22] Y. Jing, M. Covell, D. Tsai, and J. M. Rehg, "Learning Query-Specific Distance Functions for Large-Scale Web Image Search", IEEE Transactions on Multimedia, VOL. 15, NO. 8, 2013.

[23] Z. Zhang, R.K. Srihari, and A. Rao, " Face Detection and Its Applications in Intelligent and Focused Image Retrieval ", 2000.

AUTHOR

Yossra H. Ali received the B.Sc., MSc., and PhD. degrees in computer sciences from University of Technology, Iraq , in 1996, 2002, and 2006, respectively. Currently she is Assist Professor at faculty of computer sciences at University of technology, The research interests include Agent programming, image processing and Information Security.

Wathiq N. Abdullah received B.Sc., and M.Sc., Degrees in computer science from University of Baghdad, Iraq, in 2001 and 2004, respectively. Currently he is a Ph.D. student in the University of Technology, Department of Computer Science had completed the Ph.D courses on 2013 and now he begins a research on Image Retrieval.
He is a lecturer in the University of Baghdad, College of Education, Department of Computer Science teaching several computer science courses like Structured and Object Oriented Programming , Data Structure, Image Processing, System Analysis, Pattern Recognition, and Information Technology.