A Comparison Study of Image Descriptors on Low-Resolution Face Image Verification

Gittipat Jetsikut, Sasipa Panthuwadeethorn and Suphakant Phimoltares
Advanced Virtual and Intelligent Computing (AVIC) Center
Department of Mathematics and Computer Science,
Faculty of Science, Chulalongkorn University, Thailand
Email: gittipatr@gmail.com, sasipa.c@chula.ac.th, suphakant.p@chula.ac.th

Abstract—Face images are commonly used in various tasks such as face detection, facial feature extraction, and face recognition. In the security applications, the images are used as input of the identification or verification systems due to its usability, availability, and reliability. Since the capacity of memory of face database system is limited, all images retained in the database should be of small size or low resolution. This paper proposes low-resolution face image verification based on image similarity measure as well as five image descriptors—Color Histograms, Auto Correlograms, MPEG7-EHD, CEDD, and Tamura’s Texture. The results of image comparison were depicted under five image types: regular image, overexposed image, underexposed image, non-frontal image, image with facial expression. Lastly, the discussion of our results was provided to present the relationship between the descriptors and the conditions. With our comparison study, the results showed that for a low resolution image the important factors of image are edge, color and gradient. This information can be used in the further various studies.

I. INTRODUCTION

The widespread use of internet leads to the need of accessing online data and applications in different levels. Many identification approaches are proposed to ensure the security of the data and systems. Although password and personal identification number are standard approaches in the identification systems, they are still vulnerable. Since biometric identifiers are unique to individuals and more reliable in identity verification, the use of biometric for system identification increases. Examples of physiological features extracted for the biometric identification system are fingerprint, face, palm print, hand geometry, iris, etc. Among these features, face image is one of the most popular used features by reason of its availability and usability.

Basically, many biometric applications using face image are face detection and face recognition. For face detection, a face in an image is detected and located and then the face can be also used in the next face application. For face recognition, the objective of the recognition system is to identify a person from a digital image or a video frame containing a face. The system extracts a face from the image and search for the matching faces stored in the database. There are two subclasses of face recognition, i.e., face verification and face identification. Nowadays, the face recognition system is not only used for security purposed, but also is used for other applications. For example, Kinect motion gaming system [1] uses face recognition to differentiate among players, social media using face recognition to automate user tagging in photograph. Moreover, using face image as an important feature also plays an important role in Content Based Image Retrieval. The CBIR is an active research area in computer vision. Unlike the traditional Concept Based Image Indexing that searches for an image by relying on metadata of the image (e.g. captions, keywords), CBIR is a technique that uses image contents to retrieve an image. The CBIR utilizes an image similarity concept based on considered image contents. The simple image contents are low-level features (e.g. colors, shapes, textures, spatial layout, etc.) corresponding the properties obtained from images. To compare two images, the features are extracted from both of them and then image distance will be calculated from their features to measure the similarity of the images [2]. For example, a distance of 0 means the exact match. In particular case of face image database, the group of face images that containing low distance values are retrieved after a query face is fed into the retrieval system.

From mentioned benefit of face image, face image similarity is a core concept in both of face recognition [3] and the CBIR with face image database. Usually, any two images that used in the comparison process are in the same high quality in terms of resolution and sharpness. The basic factors that could affect process and cause high error are age, shape of face and color skin [3]. In addition, due to the limitation of the memory size in a computer or the efficiency of low-end digital camera such as a camera embedded in mobile phone, images stored in the image database is of small size or low resolution. In the comparison process, the resolution of a query image is much higher than the others stored in the database. If two face images with different level of resolution are compared, the calculated distance may become worse than those in the same level of resolution. This challenges the face similarity measure process by means of selecting an appropriate feature to overcome this task.

In this research, five descriptors were selected to compare in the experiment – Color Histogram, Auto Correlograms, MPEG-7 Edge Histogram Descriptor (MPEG7-EHD), Color and Edge Directivity Descriptor (CEDD), and Tamura’s
Color correlograms is a method that considers all the combination of pairs of colors. The size of an image specifies the probability of finding a pixel of color correlation of pairs of colors. The color correlograms approach that accounts for the local spatial histogram descriptor, J. Huang et al. [7] proposed color correlograms that provide a way to represent visible colors in an image. The color space is partitioned into several bins to store the distribution of colors in an image. Generally, computers use the RGB color space to represent visible colors in an image. The color space is partitioned into several bins to store the distribution of colors in an image. The space is divided into a fixed number of blocks to store the pixel counts within the color range or frequencies of the occurring colors [5]. The advantages of this approach are easy computation and insensitivity to small changes in viewing positions. However, its drawback is ignorance of other information (e.g. shape, spatial location, texture) [2, 7, 8]. Moreover, the color layout of the image is not considered. In other words, two images with different object contents may have same color histogram.

A. Color Histograms

Color histogram is a common approach in image retrieval [4, 5, 6]. Since the color content extracted from an image is the most basic information that is widely used in comparing images, this approach uses color histograms to represent the distribution of colors in an image. Generally, computers use the RGB color space to represent visible colors in an image. The color space is partitioned into several bins to store the pixel counts within the color range or frequencies of the occurring colors [5]. The advantages of this approach are easy computation and insensitivity to small changes in viewing positions. However, its drawback is ignorance of other information (e.g. shape, spatial location, texture) [2, 7, 8]. Moreover, the color layout of the image is not considered. In other words, two images with different object contents may have same color histogram.

B. Auto Correlograms

To improve the accuracy of image retrieval using color histogram descriptor, J. Huang et al. [7] proposed color correlograms approach that accounts for the local spatial correlation of pairs of colors. The color correlograms of an image specifies the probability of finding a pixel of color correlation at a fixed pixel distance \( k \) from a pixel of color \( i \) in the image. Considering all the combination of pairs of colors, the size of color correlograms is \( O(m^2d) \) which is very large; \( m \) is the number of colors and \( d \) is the distance between pixels. J. Huang et al. [7] also proposed auto correlograms which is a subset of color correlograms. This approach describes only the spatial correlation between identical colors and the size of color correlograms decreases to be \( O(md) \).

C. MPEG-7 – EHD

The MPEG-7 visual standard for content description was designed by The Moving Picture Experts Group (MPEG) to provide standardized descriptions for images or videos. MPEG-7 can be used independently from other MPEG standards. MPEG-7 specified several descriptors such as visual color descriptors, visual texture descriptors, visual shape descriptors, and motion descriptors for video to allow users or agents to search for images or videos. For example, users draw a few lines on the screen to get images that contain similar graphics, users describe actions to get a set of videos with similar actions [9]. In this research, MPEG-7 Edge Histogram Descriptor (EHD) was chosen to test in the experiment. MPEG-7 EHD represents the spatial distribution of 5 types of edges that are vertical, horizontal, 45°, 135°, and nondirectional edge. The descriptor divides an image into 16 nonoverlapping blocks of equal size and utilizes a 5-bin histogram to express the edge information of each image block [9]. The descriptor is scale invariant and supports rotation invariant and rotation sensitive matching operations [4, 9].

D. CEDD

Color and Edge Directivity Descriptor (CEDD) is a descriptor proposed by [8]. It includes both color and texture information in a histogram. An image is divided into predefined number of blocks and the low level features which are color and texture information of each block are extracted independently. The structure of CEDD contains six texture regions and each of which contains 24 color regions. The histogram consists of 144 bins according to the overall 144 color regions within six texture regions. For texture information, five types of edges proposed in MPEG-7 EHD are utilized to create six regions. To extract color information, two fuzzy systems are applied to map the colors in each region to 24 preset colors. A quantization is also used to define the color as three binary digits in the interval [0, 7] so the total size is limited to be 144 x 3 = 432 bits or 54 bytes per image. The small size of the descriptor is the advantage of CEDD.

E. Tamura’s Texture

Texture is an essential feature for human visual perception. Texture in an image is defined over a region or sub image by grey levels rather than at a point [10, 11]. Tamura et al. [10] proposed 6 basic textural features (coarseness, contrast, directionality, line-likeness, regularity and roughness). The psychological experiments were designed to construct psychometric prototypes and compare them to the computational measures of these textural features. The first three textural features yields better results than the others. Among them, coarseness is the most essential factor in the texture.

III. LOW RESOLUTION IMAGE AND INFLUENCE OF IMAGE CONDITIONS

This section shows a description of low resolution image and the effect of image conditions when getting an image from a digital camera.
A. Low resolution image

A face image obtained from a digital camera can be of various sizes. In many applications such as video surveillance or image retrieval, the images collected in the database are in low resolution due to the limitation of a digital camera or storage capacity. There are two main reasons in case of taking the image by the digital camera, i.e., 1) its hardware cannot process and display high-quality image and 2) the speed of network does not support high-quality image communication. Moreover, with the specific size of storage capacity to carry the high number of images, each image should be shrunk into an appropriate size without any change in content and condition. Comparing low resolution image with an input image obtained from user’s camera is a challenging task because the resolution of input image is usually much higher than that of images in database. Since both images are not in the same size, the larger image must be become smaller and then the similarity can be calculated. In other words, the quality of the original image will decrease before operating matching process. In this study, the frontal face has been captured in small size of 180 x 180 with the resolution of 72 dpi. The color model of the low resolution image is a standard RGB that each pixel contains intensity value ranging from 0 to 255. Fig. 1 shows the examples of low resolution image.

B. Image conditions

Image condition affects an appearance of each image under some specific factor in real environment. In this study, image conditions that were considered are separated into 4 classes. The types of input image considered in this study can be depicted as follows.

1. Regular face image

Regular face image is a face image without any image condition. Achieving this image is simple because the image is taken by the digital camera without specifying camera parameters. All parameters including white balance are automatically defined by the camera to yield the image whose appearance is similar to a scene from human perception under typical light condition. In addition, the view of regular face image is frontal with the assumption that users usually takes an image of their own in this view. The example of regular face image is illustrated as Fig. 2a.

2. Overexposed face image

When a face image is taken in outdoor environment, the image will become much brighter. It is difficult to extract image information directly from the overexposed image. To cope with this situation, the image must be darkened with appropriate set of parameters before resizing and applying the matching process. However, the operation of darkening consumes processing time. The matching process should handle this problem without adjusting brightness of image to avoid such time consumption in preprocessing. In our experiment, the matching process based on five descriptors is presented and directly applied to the overexposed face image obtained by a digital camera with pre-defined +2EV exposure compensation in usual scene. The example of overexposed face image is illustrated as Fig. 2b.

3. Underexposed face image

Unlike overexposed face image, in case of insufficient light condition for taking an image, the face image becomes much darker and some information of image becomes less. Again, to solve this problem, the brightening process should be applied to the regular face image before resizing and using matching process. Moreover, the time consumption of this process is also required. Hence, to reduce overall time consumption, the descriptor should be directly extracted from the regular face image without adjusting image brightness. This condition is as important as overexposed face image. The suitable descriptor should support both types of light condition. In our experiment, the underexposed face image obtained by a digital camera with pre-defined -2EV exposure compensation in usual scene. The example of underexposed face image is illustrated as Fig. 2c.

4. Non-Frontal face image

In many face applications such as face recognition and face detection, the most popular and usual view of image appropriate for the system is frontal image. It contains all image features, i.e., eyes, nose, and mouth, with symmetrical information. In case of capturing non-frontal image, some image features might be partially occluded. There are two types of pose variations, that is, horizontal and vertical variations. The horizontal variations more often occur than the vertical variations in usual behavior of human. This is a reason why most methods focus on the horizontal variation more than the other. If the method can handle non-frontal face image, the range of angle for pose variation should be also limited. Some face detection can support ±90° horizontal variations whilst some supports only frontal image. In our experiment, ±45° horizontal variations are considered for face matching. The example of non-frontal face image is illustrated as Fig. 2d.

5. Face image with facial expression

Facial expression is a representative for personal emotion. Basically, face database used in real-world applications contains face images without facial expression. Comparing the low resolution face image with face image with facial expression is quite a difficult task due to the difference of image appearance. There are five main facial expressions corresponding to emotions, which are happy, surprise, anger, sadness, fear, and disgust [12]. In this experiment, facial expression is a factor to be concerned. Face images with either happy or surprise emotion were collected to be used in the face matching process. The example of face image with facial expression is illustrated as Fig. 2e.
Fig. 1 Example of low resolution image

Fig. 2 Regular face image and face image with conditions. a. Regular face image. b. Overexposed face image. c. Underexposed face image. d. Non-frontal face image. e. Face image with facial expression.

IV. EXPERIMENT

A. Experiment Setting

The specification of both software and hardware in our experiment was defined as follows.

1. Hardware
   - DSLR digital camera. This camera was used for constructing a dataset of images of size 3456x5184 with conditions. The camera lens used is 18-135 mm.
   - Phone camera. Low-quality camera with resolution of 0.3 megapixels embedded in digital mobile phone was selected to capture low resolution face images of size 180x180.
   - Computer. The computer used in a matching process includes CPU core i5 with 2.40 GHz speed and 4 GB RAM.

2. Software
   - Img (rummager). A program for calculating image distance developed by S.A. Chatzichristofis, Y.S. Boutalis and M. Lux[13].

In our experiment, the face images were collected from 11 persons by DSLR camera and low-quality camera embedded in mobile phone. Then, five groups of face images obtained from DSLR camera, i.e., regular face images, overexposed face images, underexposed face images, and images with facial expressions will be compared with low resolution face images obtained from phone camera by using Img (rummager) program with one of five descriptors, i.e., color histograms, auto correlograms, MPEG7-EHD, CEDD, and Tamura’s texture. Next, an image distance was calculated to measure a similarity between two images. If the distance is low, it can be concluded that the considered image with condition is similar to the low resolution image. Subsequently, two scenarios were considered in our experiment of a person’s face. The first scenario shows the similarity between a low resolution image from a person and face images with conditions from the same person. This scenario indicates how different two images from the same person are. For this scenario, the average distance $\overline{d_{s1}}$ between two images was calculated. In contrast, the other scenario shows the similarity between a low resolution image from a person and face images with conditions from different persons. This scenario presents how different between two images gathered from the different persons. For this scenario, the average distance $\overline{d_{s2}}$ was also computed. The mid-point of these two average distances was calculated as follows.

$$d_{mid} = \frac{\overline{d_{s1}} + \overline{d_{s2}}}{2}$$

After getting the mid-point, the standard deviation (SD) of all distances was derived so that we can define the appropriate range of threshold used for deciding whether the face image with condition belongs to that person. In this experiment, 21 thresholds starting from $d_{mid} - SD$ to $d_{mid} + SD$ with step size of 0.1 SD were considered. Each threshold value was used for comparing the low resolution face image with the face image with condition. If the distance between two images is less than the chosen threshold, then the image with condition will be accepted as a verified person. Otherwise, the image will be rejected. Then, the False Rejection Rate (FRR) and the False Acceptance Rate (FAR) were calculated as follows.
\[ FRR = \frac{\text{the number of false rejections}}{\text{the number of comparisons in Scenario 1}} \]  
(2)

and

\[ FAR = \frac{\text{the number of false acceptances}}{\text{the number of comparisons in Scenario 2}} \]  
(3)

Both the FRR and the FAR present the performance of verification system. Lower FRR and FAR yields higher system performance. However, for the best configuration in this experiment, harmonic mean of these two rates was calculated and the derived form can be shown as follows.

\[ H = \frac{2 \cdot FRR \cdot FAR}{FRR + FAR} \]  
(4)

The most suitable threshold of each person is the value causing minimum harmonic mean calculated from FRR and FAR. After the suitable threshold, harmonic mean, FRR and FAR were obtained for a person, the average FRR (FRR) and the average FAR (FAR) will be calculated to evaluate the overall performance of the system as follows.

\[ FRR = \frac{\sum_{i=1}^{N} FRR_i}{N} \]  
(5)

and

\[ FAR = \frac{\sum_{i=1}^{N} FAR_i}{N} \]  
(6)

where \( N \) is the total number of persons in the experiment, \( FRR_i \) and \( FAR_i \) are the FRR and the FAR belongs to person \( i \), respectively.

B. Experiment Results

After computing the average FRR (FRR) and the average FAR (FAR), the harmonic mean of these two values was also calculated to indicate the best descriptor for low resolution face image verification as shown in TABLE 1.

<table>
<thead>
<tr>
<th>Image Condition</th>
<th>FRR</th>
<th>FAR</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular face Image</td>
<td>0.0909</td>
<td>0.6909</td>
<td>0.1607</td>
</tr>
<tr>
<td>Overexposed face Image</td>
<td>0.0909</td>
<td>0.5818</td>
<td>0.1572</td>
</tr>
<tr>
<td>Underexposed face Image</td>
<td>0.7273</td>
<td>0.1000</td>
<td>0.1758</td>
</tr>
<tr>
<td>Non-Frontal Face Image</td>
<td>0.0909</td>
<td>0.7818</td>
<td>0.1627</td>
</tr>
</tbody>
</table>

C. Discussion

According to the experimental result in Table 1, the MPEG7-EHD is the best descriptor in term of harmonic mean. Besides, it was found that CEDD is the interesting descriptor because CEDD achieved the lowest FRR while its FAR is not much different to the lowest one. One common factor causing these descriptors better than the others is dividing an image into blocks while the distance from the other methods is calculated from the entire face images. Moreover, in a view of features used in this study, the results can be concluded that using edge histogram as in MPEG7-EHD or a combination of color and texture as in CEDD accomplishes the face matching task.

For the further investigation with the best descriptor as MPEG7-EHD and various conditions of a face image, the descriptor shows the highest performance in overexposed face image. Under this condition, most of information in background region is discarded because such region is too bright while the focused region still contains the significant face information. This reason might cause the distance between two images derived only from the actual focused face. However, under the consideration of CEDD in Table 3, the condition affecting the lowest harmonic mean is face image with facial expression while the lowest performance is obtained from the underexposed image that the information of the focused region like a face is suppressed as shown in Fig. 2c. However, there is no significant difference among the rest conditions in term of harmonic mean. Additionally in the results of each descriptor, the harmonic mean of all five image conditions seems to be higher than that of each condition. This is caused by the selected threshold used for the experiment in Table 1, which must support image variance in five conditions while those thresholds selected for the
experiments in Tables 2 and 3 are used for the verification system with only one possible image condition. In the other word, the harmonic mean is proportion to the variance of image conditions.

Lastly, there are some aspects to be concerned in our future work. From the experiment, the background is a common factor that affects the distance between two images. If two face images are taken in two different places, the backgrounds of both images will become different. This causes higher distance than it really is. Moreover, various types of non-frontal face images and face images with facial expression should be considered as well.

V. CONCLUSIONS

This research is to find the best descriptor for low-resolution face image verification. The difficulty of this task is comparing two images with different levels of information. From our experiment with five image descriptors and five conditions, it is indicated that MPEG7-EHD is the best descriptor because of using edge as the main feature. In other words, although the information of low resolution image is quite low, the edge gathered from the gradient of an image is still sufficient to be used to measure the difference between two images. Additionally, the future study with other conditions and more individuals should be also provided.

REFERENCES