Image Recognition Technique using Local Characteristics of Sub-sampled Images

Group 12: Do, Hyungrok

***Abstract*—An image recognition technique utilizing a database of image characteristics is introduced. This technique is different from eigenimage method which requires a large amount of information of training set images in terms of the size of each image and the number of images in the database. Especially, this technique is useful for recognizing images which have fixed shape and structure such as paintings and documents. In this study, images of 33 different classic paintings taken by a common camera phone are used to construct the database and a MATLAB code is written for image recognition. 66 different images of the 33 paintings are tested and approximately 80 % of them are recognized correctly.**

In the code, low pass filters for noise reduction, morphological operators such as dilation and majority filter for clear and smooth boundaries and Haralick corner detector to find characteristic points are used. To construct a database which consists of small size images, original images are trimmed to be a smaller image which contains only the region of interest. Furthermore, each image is sub-sampled to be a fixed size gray image which is 200 pixels by 200 pixels. Sub-sampling can reduce discrepancy of trimming position and angle/position of camera. Finally, using Haralick corner detector, 100 corner points which have large cornerness are selected per image. The points are positioned on a 200 by 200 binary image, which is a reference image in database.

Some conventional image processing techniques are applied to an input image. The resulting image is also converted to a binary 200 pixels by 200 pixels image and compared with the 33 reference images in the database being shifted and warped.

I. INTRODUCTION

RECENTLY, image recognition techniques have been studied for many applications. Especially, those techniques are useful for computer based automated recognition systems and mobile camera phones. Among them, one of the most well-known techniques is the eigenimage method. Eigenimage method is capable of recognizing complex objects such as human face [1]. Furthermore, by collecting information from a large group of training set images possessing same characteristics, this novel algorithm can classify and recognize an enormous number of different groups. In the other word, if enough number of training set images can be obtained, then this intelligent algorithm can discriminate any sophisticated difference of images.

However, if the number of groups is small, then the characteristics of images required to discriminate groups can be much simpler. Therefore, in this study, I tried to use the minimum amount of information for image recognition. In the next section, two different numbers of characteristic points per reference images in database are used and compared.

Generally speaking, perception of an object does not require whole detailed information of it, but only some characteristics which are discernable. However, as the number of discernable groups increases, the characteristics of each group should be more sophisticated to be recognized correctly. For example, at least 33 different characteristics (e.g. point) are required to recognize images from 33 different groups. Fortunately, an image which has complex patterns or structure has many kinds of characteristics. Among them, corner points or cornerness are selected as the representative characteristic points in this study.

Sub-sampling can be used when the number of pixels in the reference images is much larger than the number of groups to be discriminated. Needless to say, information in the image would be lost by sub-sampling. However, if number of pixels in the sub-sampled image is still much larger than the number of groups, then the information in the sub-sampled image is usually enough to distinguish different groups.

A crucial merit of sub-sampling, except the small data size, is reduction of the errors or discrepancy which can be frequently caused when the region of interest is trimmed and detected. For example, let's assume that a region of interest is shifted by 10 pixels and the trimmed image is sub-sampled by a factor of 5. Then the discrepancy or error of position would be reduced to 2 pixels. Similarly, rotation or warping in a proper range can be neglected in the sub-sampled image. Furthermore, by dilating characteristic points (e.g. corner points) in the sub-sampled image, this kind of problems can be easily resolved in a larger range of unexpected shifting, rotating and warping.

II. PROCEDURE

A. Pre-processing

The first step of the pre-processing is excluding gray wall region utilizing ratio of intensities of blue color over red color. If the ratio of a pixel is around one, one can assume that the pixel has gray tone color, although all of the intensities of red, green and blue colors should be the same to be gray. These pixels are set to be black. The other ratios (e.g. green over red, etc.) are also used to perform the same processing for comparison.

Do, Hyungrok, Ph. D. Student in Mechanical Engineering Department of Stanford University, Bldg 520 room 520i Duena St. Stanford CA, 94305 (phone: 650-353-1243; e-mail: hyungrok@stanford.edu)

Fig. 1. Gray color regions and wall region are set to be black.
(Original image: 13_3.jpg)

Fig. 2. Trimmed image of the region of interest.
(Original image: 13_3.jpg)

As shown in fig. 1., the wall region is completely black, although, unfortunately, some parts of paintings also become black. One can observe noisy regions near the bottom of the figure. However, the region of interest, the painting in the center, has relatively clear boundaries compared to the noisy region. To clarify and strengthen the boundary of the painting, the figure is dilated by 3 by 3 structuring element. Additionally, to reduce the noise out of the region of interest, a low pass filter is applied before thresholding. In the procedure, the resulting image (fig. 1.) should be a gray image. Therefore, red color is selected as a representative color. Average intensity of the three colors is also used for comparison.

The second step is trimming the region of interest. Presumably, the picture in the center should be the dominant painting. Therefore, the algorithm probing the dominant painting is designed to start from the center point of the original image. To probe the region of interest, variable size of frame is used to scan the image. The gradient of number of nonzero pixels in the frame as a function of frame size is set as a criterion. After the first trimming job is completed, further trimming is executed from four edges of the trimmed image. Second trimming is to exclude noisy regions near the frame of the painting.

Fig. 2. shows the resulting image of trimming procedure. The edges can be determined at different positions due to the curved shape of the frame or the shading below the frame, which depend on images. However, as discussed in the introduction, this kind of problem can be resolved by sub-sampling. Additionally, the frame region is cut because frame has quite different characteristics from paintings. A fixed ratio of width and height is cut for that reason.

The third step is sub-sampling to be 200 pixels by 200 pixels image. The trimmed images have various sizes due to different resolution, angle, camera position, scale of original painting, etc.. However, by the sub-sampling procedure with variable sampling factors, every image can be transferred and projected to a same space. In this smaller common space, the following procedures become faster and easier.

The fourth step is corner detection using Haralick corner detector. Corner points are detected by thresholding the determinant of normal matrix (cornerness) which is calculated from horizontal and vertical gradient using Sobel operator. The cornerness is independent of rotation and shift. Therefore, a set of corner points can define local characteristics of sub-sampled images. In the code, 100 corner points per sub-sampled image which are ranked in 100 biggest cornerness are selected to represent the characteristics of the image. Using these points, a binary 200 pixels by 200 pixels image is constructed. In the binary image, all pixels have 0 intensities except the corner points. At the first trial, 50 corner points per image are used because 50 is bigger than 33 (number of distinguishable groups) and not much bigger than 33. However, the percentage of correctly recognized images is only approximately 60 %. After raising the number of corner points, the percentage is increased to 80 %. Fig. 3. shows the corner points on the gray color image.

The fifth step is dilation of the resulting binary image of corner points. For the image recognition, the resulting binary image will be compared with reference images to find the most well matched one. Therefore, any small errors or discrepancies generated in the trimming procedure can cause a serious problem because the corner points can miss them in the reference images by a few pixels. Therefore, the corner points are dilated by a 3 by 3 structuring element. In fig. 4., three

Fig. 3. Corner points on the gray scale image.

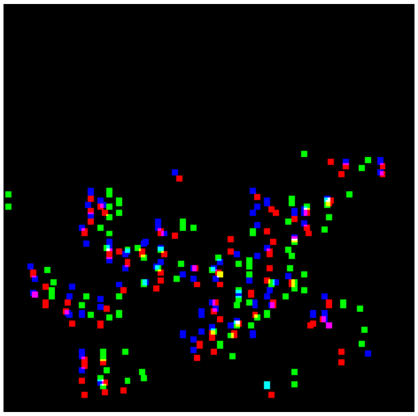

Fig. 4. Dilated corner points of 13_1.jpg (red points), 13_2.jpg (green points) and 13_3.jpg (blue points).

colors (red, green and blue) represents the corner points of 13_1.jpg, 13_2.jpg and 13_3.jpg. As shown in the figure, dilated regions share more pixels than original corner points. This will help the matching algorithm work better.

B. Database Construction

99 images of 33 paintings are given for this study. Among them, 33 different painting images are selected as reference images to be used for database construction. The criteria of the selection are simple: better focusing, lower noise level and correct posture. Then the 33 reference images undergo the pre-processing procedure. As a result, 33 binary 200 pixels by 200 pixels images are obtained. This database can be compressed to be much smaller because they are binary and have only 100 nonzero points per each image: the information of 100 point positions is enough. However, to expedite the speed of image recognition procedure, the compression procedure is omitted in the code.

Additionally, perspective view of the binary images are calculated and added to the database. The transformation is linear and simpler than affine transform because the whole region of interest is projected on 200 by 200 space. Nevertheless, most of the images in the training set could be recognized correctly without the reference images of perspective view. This means that the rotation and warping of the given test images are in the error range of sub-sampled and dilated characteristic pixels. In the point of view, the selection of the size of the projected space (200 by 200) is appropriated for this specific application although the algorithm will not work if the region of interest is smaller than the space.

C. Image Recognition

By comparing the pre-processed input image with the reference binary images in the database, paintings in the input image can be recognized. To guarantee the better matching, 8 shifted images per one input image are generated. Toward the four ways(left, right, up and down), the input image is shifted by 3 and 6 pixels. Therefore, total 9 input images are ready to be match to the 33 reference images in the database.

Comparing job follows a simple algorithm. A reference image is subtracted from the 9 input images. Then the minimum number of survived pixels, whose intensities are bigger than zero, among the 9 input images is chosen. The same routine succeeds 33 times with the reference images and the minimum number of the survived pixels can be obtained. The reference image of the minimum number is most well matched to the input image. Finally, the title of the reference image will be presented.

III. Result and Discussion

(a)

(b)

(c)

Fig. 5. Original images: (a) 13_1.jpg, (b) 13_2.jpg and (c) 13_3.jpg.

(a)

(b)

(c)

Fig. 6. Original images: (a) 3_1.jpg, (b) 3_2.jpg and (c) 3_3.jpg.

Except the 33 images used for constructing database, 66 remaining images in the training set are tested by the code. Approximately 80 % of the images are recognized correctly. Most irrecognizable images are too noisy or too much warped. Intermediately warped, rotated or noisy images are recognizable using the code. As shown in fig.4., although the images (13_1.jpg, 13_2.jpg and 13_3.jpg) are taken at quite different angles and distances: original images are in fig. 5., the positions of corner points or the patterns of points are not so different and all images are recognizable by the code. Furthermore, defocusing problem is not so crucial in this case. As shown in fig.5., 13_1.jpg is a defocused image compared to the other two. However, it is also recognized correctly by the reference image, which is the pre-processed image of 13_2.jpg. Different resolution problem is also resolved by the sub-sampling algorithm. The resolution of trimmed image of 13_1.jpg and 13_3.jpg must be quite different. However, in the sub-sampled space, both images are pretty well matched to each other as shown in fig. 4..

One more example is presented in fig. 6. to investigate the problem of perspective view. The three images in fig. 6. are taken at three different angles. In fact, this problem is partly resolved when the trimmed image is sub-sampled to a same space because the factor of sub-sampling depends on the perspective angle. Therefore, the image can be partly corrected by the fixed ratio of height and width of the sub-sampled space (200 by 200). As shown in fig.7., these three images are projected onto the 200 by 200 space. One can observe that the corner points from the three images have a common pattern although they are slightly shifted. However, as explained in the procedure section, 8 shifted images are generated from each input images. As a result, these 3 images are recognized correctly by the code.

The most important merit of this technique is that it does not need transforms to correct warp, rotation and shift of input image. This is mainly due to the use of sub-sampled and fixed small size 2D space. The relatively small errors caused by the preceding algorithms can be minimized by this algorithm.

Secondly, this technique requires only a small size of memory for the database storage. Only one reference image from each group is usually enough to define the characteristics of the group. Furthermore, each representative image contains only 100 nonzero pixels which can be compressed to a simple index.

However, the limitation of this technique is also evident. At first, the region of interest should be bigger than 200 pixels by 200 pixels. It is not a simple problem to change the size of sub-sampled 2D space because smaller space would lose more information of the original images. Especially, smaller 2D space can not be used when there are many groups to be discriminated because more sophisticated characteristics should be provided for the recognition.

Moreover, if rotation or warp exceeds the acceptable range,

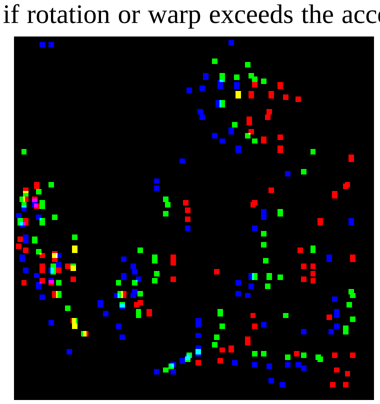

Fig. 7. Dilated corner points of 3_1.jpg (red points), 3_2.jpg (green points) and 3_3.jpg (blue points).

this technique cannot recognize the image at all. Therefore, a transform which can correct the posture of the painting should be added in the pre-processing procedure. To detect the dominant angle of frames of paintings, Hough transform is used in the study. However, correction of dominant angle was not so effective in this application because the warping is always exists due to small camera phone lens. It is almost like a point observer. In the process, more sophisticated and proper transformation model is required to resolve this problem. Affine transform can be a good solution. However, searching for a globally optimized (e.g. global minimum) affine transform coefficients is almost impossible or demands very high cost (e.g. calculation time). Using random variables, I tried to find globally optimized affine transform coefficients. However, the calculation time was much longer than a minute which is the limit of the code running time.

REFERENCES

[1] Matthew Turk and Alex Pentland, “Eigenfaces for Recognition,” The Media Laboratory M.I.T.