ATRテクニカルレポート表紙

[公開]

TR-M-0053

# The System Development of the Cyber Photographer Project

ジェフリー　チ　イェン　エン　　田中　昭二
Jeffery Chi Yin Ng　　Shoji TANAKA

2000.4.20

ATR 知能映像通信研究所

**Abstract**

The Cyber Photographer is a robot that takes pictures according to the compositional information extracted from masterpieces. My purpose during my coop stay was to develop modules to handle a number of image processing procedures, and in to increase the speed and accuracy of these modules. The Peer-Group Filter was developed to remove random noise, and sudden discontinuality from the image. Colour Image Segmentation was developed so that the subject and the background from a given image could be separated and given to the Attractive Region Extraction algorithm. Rotationally Variant Fourier Descriptors, and Moment Fourier Descriptors were developed for the shape-matching portion of the Cyber Photographer's software, so that the correct compositional information for a given subject could be accurately selected from the predefined Compositional Information Database. Of the developed modules a number were rejected; others modified and enhanced to suit the needs of the Cyber Photographer.

**Contents**

## 1 Introduction

The Cyber Photographer is project that was developed to allow a computer to take pictures based on the compositional information taken from masterpieces, and produce images that are comparable to those taken by professionals.

The purpose of my stay at ATR, was to develop modules to handle a number of image processing procedures, and to increase the speed and accuracy of these modules. These modules were Moment Fourier Descriptors, Rotationally Variant Fourier Descriptors, Colour Image Segmentation, and Peer-Group Filter. A few of these developed modules were rejected for various reasons, while others were modified, or enhanced for the Cyber Photographer's needs.

The rest of this paper is organized as follows: Section 2 will detail the software that makes the Cyber Photographer possible. Section 3 will present the conclusion of this paper. Section 4 will list the references used. Finally, section 5 will provide a brief listing and description of the source code used.

## 2 The Cyber Photographer

A picture is taken by the Cyber Photographer when the operator sends a "Take" command to the robot by clicking on the appropriate button on the Cyber Photographer GUI.

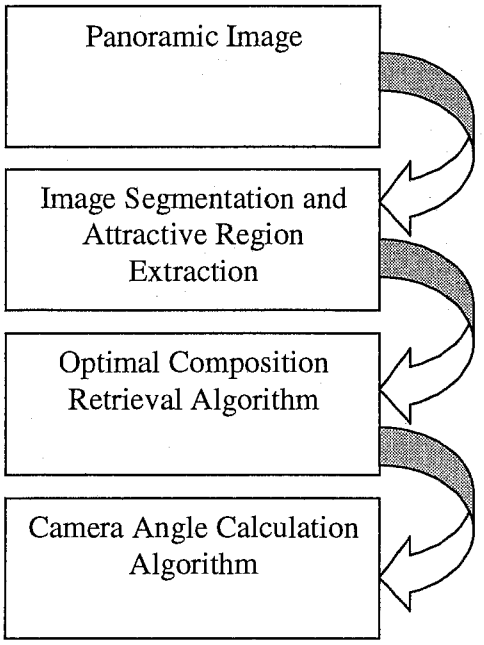



**Figure 1: System Overview**

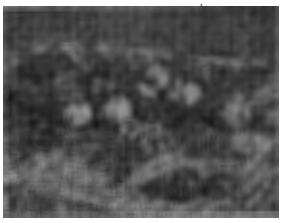
The system flow is presented in the chart above (Figure 1.) First, an initial panoramic image is taken. Afterwards, that image is given to the Image Segmentation and Attractive Region Extraction module, so that objects within the image can be separated from the background, and then the most visually attractive segmented region is selected. The Optimal Composition Retrieval module determines how best to place the chosen subject in a photograph, by finding a Masterpiece that possesses a subject of similar shape, size, and orientation as the selected region. After a masterpiece is chosen, that masterpiece is used as a template for how the subject is best framed with the surrounding background. Finally, compositional information obtained from the Optimal Composition Retrieval Algorithm and the chosen subject is given to the Camera Angle Calculation Algorithm that crops the panoramic image so that the resulting image will contain the compositional

information of the original masterpiece, thus being the optimal composition for the selected subject in the panoramic image.

An example of how the system works is shown in the diagram below. The computer at the remote site generates a panoramic image from the video frames in real-time while the robot rotates its head, (Figure 2 Panoramic Image). The figures and features within the image are then identified, (Figure 2 Figure Extraction Result), and the most attractive region selected, (Figure 2 Extracted Region). The best matched masterpiece for the selected region is chosen, (Figure 2 Retrieved Painting), and the panoramic image is trimmed so that the final image possesses the ideal composition for its subject, (Figure 2 Final Trimmed Image).

Generated Panoramic Image

Extracted Region



Retrieved Painting Final Trimmed Image

**Figure 2: Pictures taken by the Cyber Photographer**

Of these components, I developed the Peer Group Filter and Colour Image Segmentation used in the Image Segmentation and Attractive Region Extraction module, and the

Moment Fourier Descriptors, and Rotationally Variant Fourier Descriptors used in the Optimal Composition Retrieval Algorithm.

### 2.1 Image Segmentation and Attractive Region Extraction

The image segmentation algorithm used by the Cyber Photographer is based on Color Image Segmentation [3]. After an image is segmented the image is passed to the Attractive Region Extraction algorithm [1]. This algorithm selects the region that is most likely to catch the attention of the viewer by performing the following steps:

1. A figure-ground segmentation is performed on the panoramic image by using the Color Image Segmentation as shown in Figure 2 Figure Extraction Result.

2. The heterogeneity attractiveness of each region is then calculated by finding the difference, in descending order of importance, between the colour, texture, shape, and size heterogeneity of each region against those of other regions. The calculated difference in heterogeneities are then weighed according to their importance in the above order, and then summed for the final heterogeneity attractiveness.

3. The feature attractiveness of each region is calculated by finding, in descending order of importance, the colour, spatial frequency, and size attractiveness of each region. The calculated values are then weighted according to their importance in the above order, and then summed for the final feature attractiveness.

4. The heterogeneity attractiveness and feature attractiveness of each region is then accumulated and placed within the attractiveness evaluation function where the final evaluated attractiveness of each region is found. "This function is a Beta function that is an S curve function and can control it's critical points..." [2] The most attractive region is then chosen, as shown in Figure 2 Extracted Region.

#### 2.1.1. Peer-Group Filter

The purpose of the Peer-Group Filter [6] is to pre-process the image fed into the Colour Image Segmentation algorithm so that any random noise produced by the camera or the transmission of the image to the computer are removed allowing for a more accurate segmentation of the image. A side effect of the filter is that it removes sharp discontinuities within the image, which aids the segmentation of the image as well.

The algorithm works by first taking a 5x5 window around each pixel. Every pixel in the window has a vector calculated for it using the colour information of that pixel. In our system the values were simply the RGB values of a given pixel. The vectors are then sorted according to their distance from the origin. The two lowest and highest valued vectors are discarded immediately, the remaining vectors then have the difference between it and its closest neighbour calculated. The average ($\mu$) and standard deviation ($\sigma$) of the difference of vectors in the window are then calculated. The standard deviation is then multiplied by a user-defined value ($\beta$). Starting from the lowest valued vector, the difference between it and its closest neighbour is compared to the value:

$$(1) \quad T = \mu + \beta\sigma$$

If the difference is higher than the above value, the neighbouring vector and all vectors greater than it are discarded. The remaining vectors and their respective pixels are considered part of the "Peer Group" of a given window. The pixels in the window that correspond to the "Peer Group" vectors are then weighed, using a Gaussian distribution with the peak in the centre of the window, and summed. The summed value is the final pre-processed value of a pixel. Pixels which are near the edge and cannot have a 5 x 5 pixel placed around it are either ignored and are not taken into account in the final pre-processed image, resulting in a 4-pixel loss in width and height. Another method to deal with near border pixels is calculate the Peer-Group by using the above method but assume that pixels that are outside the image boundary were discarded immediately.

#### 2.1.1.1 Experimental Results

Figure 3 demonstrates Peer-Group filtering of an image with 20% monochromatic Gaussian noise added to it. As you can see the noise is removed, with only a slight lost in contrast, yet all the edge information remains sharp and relatively undisturbed. Though this module was said to be a requirement for Colour Image Segmentation to work properly, it was found that it provided no substantial aid to the segmentation of the image. Also since the processing time can be quite high for large images, it is an unnecessarily high overhead that does not need be added to the Cyber Photographer.

Original 20% noise added Peer-Group Filtered

**Figure 3: Peer Group Filter**

### 2.1.2 Color Image Segmentation

Colour Image Segmentation [3] is the algorithm that is used to separate objects from the background of the image. It is based on the J-value which essentially measures the distances between different colours, over the distances between identical colours in a given window. For the case when a window consists of several homogeneous colour regions, the colours are further separated from each other and the value of J is large. On the other hand, if all the colours in a given window are uniformly distributed, the value of J tends to be small. Since edges tend to be in areas where several homogenous colour regions are placed together, the J-value becomes an ideal way of identifying edges.

The algorithm works by first quantizing the image to reduce the number of colours in the image to 10–20 colours. Afterwards, it calculates the J-value for each pixel in the quantized image, thus creating the J-image. Regions where the J-values are uniformly low are considered "valleys", while regions where the J-values are uniformly high are considered "mountains". The average ($\mu$) and the standard deviation ($\sigma$) of the J-value are then calculated, and a threshold (T) is calculated by

$$(2) \quad T = \mu + a\sigma$$

where a is chosen from the set of parameter values [-0.6 –0.4 –0.2 0 0.2 0.4]. Pixels with values less than T are considered as candidate valley points. The candidate valleys are then obtained by connecting valley points using 4-connectivity. If the candidate valley has a size larger than the minimum size for the scale being used, the valley is considered a candidate valley, if not the valley remains unassigned. To increase computational speed, the valleys are then grown by averaging the local J-values in the remaining unsegmented part of the region and connecting pixels below the average to form growing areas. If the growing area is adjacent to one and only one valley, it is assigned to that valley. The process of valley determination is then is repeated again for each candidate valley at a smaller scale until the smallest scale is used. Next, the remaining pixels are grown one by one at the smallest scale. Unclassified pixels at the valley boundaries are stored in a buffer. Each time, the pixel with the minimum local J-value is assigned to its adjacent "valley" and the buffer is updated until all the pixels are classified.

After region growing, the initial segmentation of the image is obtained. This object is often over-segmented and because of this, regions with similar colours need to be merged. The average colour of each region is calculated and if the Euclidean difference between a region and a neighbouring region's average colour is below a user defined threshold, the two valleys are merged. The merged image is the final segmented image.

#### 2.1.2.1 Enhancements to Increase Processing Speed

The original algorithm, though accurate, lacks the speed that was necessary in order for it to be used in practical applications. Because of this, several changes were made to the algorithm in order to increase the processing speed. First, to handle the processing time increase caused by large images, large images are scaled down until they are below a user defined optimal area. The images are scaled down using bicubic interpolation, in order to retain as much edge information as possible during the shrinking process. The shrunken image is then segmented normally, and the pixels of each valley are then assigned to the corresponding pixels in the original image that were used to calculate the value of the pixel in the shrunken image.

Secondly, since the only computationally intensive part of the algorithm was the J-value calculation, "tiling" was used to reduce the time for the J-image to be produced. Basically, the J-value of only one pixel in a given $N \times N$ window is calculated and that J-value spread across all the pixels in the window. Unfortunately, for a given large window, a large amount of edge information could be lost, so there is a trade-off between speed and accuracy when utilizing this algorithm. However, for large images where only a rough segmentation of the image is desired, i.e., landscapes, there is not a considerable difference between using a large tile value and a smaller tile value.

Finally, the final pixel growing for an image is done one pixel at a time, and this becomes a time-consuming function. In order to speed up this portion of the algorithm, a rougher approach can be taken. Instead of performing pixel by pixel growing based on the smallest J-value, the established regions are grown by using a 4-point dilation. Any pixel adjacent to an established valley is assigned to the adjacent valley. If a pixel is adjacent to more than one valley, the Euclidean distance between that pixel and its surrounding valley pixels are taken, and the pixel is assigned to the valley that has the shortest distance. Again there is a trade off between speed and accuracy when utilizing this algorithm. Though for large images where only a rough segmentation of the image is desired this algorithm provides an quick and efficient means of segmenting the image.

#### 2.1.2.2 Enhancements to Increase Accuracy of Segmentation

As often the case, there is a trade-off between speed and accuracy. To obtain the processing speed that was necessary for the Cyber Photographer's needs, there was a noticeable drop in accuracy. To combat this, a number of steps were taken. Since the basic assumption of the Colour Image Segmentation algorithm is that the colours between two neighbouring regions are distinguishable, a good colour quantization algorithm is essential for an accurate segmentation. For this reason a K-means algorithm based on the HSV colour space was used to determine the initial colours to which the image is to be quantized. First, the entire image is converted from RGB to HSV colour space, since HSV is a closer representation of how people perceive colour. Afterwards, the K-means algorithm is used to determine the initial colour clusters and recalibrate each cluster until a stable number of colour clusters are found. The average colour of each colour cluster's pixels becomes one of the colours to which the image will be quantized. By using this algorithm, the initial number of colours can be found automatically, and accurately providing the image segmentation algorithm with a good basis to which to work with.

Furthermore, it was found during testing that weak edges were ignored or lost during segmentation. This problem was traced back to the J-value calculation, where two adjacent valleys were considered as one, because of a one or two pixel connection between valleys in the J-image. To solve to problem we added the Sobel value of each pixel to the pixel's J-value, and found the average of the combined value. Since both the J-value and the Sobel value are between 0.0 and 1.0 only the average of the two values are taken. What this does is strengthen the edge information of an image, providing a more accurate segmentation of the image.

It was also found that by adding the Sobel value to the a pixel's J-value, the threshold that was used to determine the valleys and mountains in the J-image became erratic and inaccurate due to the sensitivity of the Sobel value to textures and noise. Instead it was

found that simply using the average J-value and Sobel value of each region was enough to provide a good means of determining the valleys of a given J-image.

As a final enhancement, during the region merging process of the Colour Image Segmentation algorithm, the HSV values of each region were used instead of the RGB values. This is because HSV values that are close to one another are perceptively similar in human vision. Thus, HSV values are a better basis to judge whether two regions should be merged or not.

### 2.1.2.3 Experimental Results

The resulting images (figure 4) were segmented using the default setting of the Enhanced Colour Image Segmentation and the Original Colour Image Segmentation program. Each image is 1280 x 189 in size and was processed on a Silicon Graphics Octane. The processing time of the Enhanced Colour Image Segmentation Program was up to ten times faster than the original, yet the results of the segmentation were very similar, with the Original Colour Image Segmentation being able to pick up more subtle details from the test images. This substantial increase in speed allows the Color Image Segmentation to be practical for the Cyber Photographer's purpose.

Sample3.tiff Enhanced Colour Image Segmentation processed time: 13 seconds

Sample3.tiff Original Colour Image Segmentation processed time: 2 minutes 34 seconds

Sample3.tiff Enhanced Colour Image Segmentation processed time: 19 seconds

Sample3.tiff Original Colour Image Segmentation processed time: 2 minutes 50 seconds

**Figure 4: Comparison of segmentation between the Original Color Image Segmentation and Enhanced Color Image Segmentation**

## 2.2 Optimal Composition Retrieval Algorithm

The Optimal Composition Retrieval algorithm [3] determines how best to place the chosen subject in a photograph. This is done by finding a Masterpiece that possesses a subject of similar shape, size, and orientation as the selection region, and using that masterpiece as a template for how the subject is best framed with the surrounding background.

A database containing the shape and compositional information of subjects from over two hundred masterpieces was created for the Cyber Photographer. The subject of each masterpiece was extracted by hand, and twenty descriptors [4] were calculated for each subject as the shape information to be used by the Cyber Photographer. The Composition Analyzer was used to analyze and extract the compositional information of each masterpiece.

The chosen region from the panoramic image also has twenty descriptors calculated. The descriptors for the selected region and every subject in the database are compared and the Euclidean distance between each pair of Fourier descriptors taken. The subject that has the shortest distance compared to the selected region is the most similar in shape to the

selected region. That image is then used as the template image for how the panoramic image will be trimmed as shown in Figure 2, Retrieved Painting.

### 2.2.1 Rotationally Variant Fourier Descriptors

Normally, Fourier Descriptors are invariant to translation, rotation, and scaling, but sensitivity to rotation was necessary for the Cyber Photographer because the compositional technique used for an object that is horizontally elongated is vastly different from that technique used for an object that is vertically elongated.

It is assumed that the image that will have the Fourier Descriptors calculated is a solid object, does not possess any holes, and is black and white. First, an outline of the object is taken, then the outline is then traced, and the x, y coordinates of the outline are kept in a buffer. Afterwards, the points in the buffer are then normalized by the length of the shape. The coordinates are then fed into another algorithm that computes the Fourier Transform of the coordinates, and returns the complex values of the Fourier Transformed coordinates. Normally, the amplitude of the complex values are taken in order to achieve rotational invariance, but since rotational sensitivity is desired, the amplitude of the values are not taken, and instead the raw values are stored together to be used for comparison with other Fourier Descriptors.

Two programs were developed using this algorithm. One was made to create a database of rotationally variant Fourier descriptors given images that are desired to be in the database. The second program was created to calculate the descriptors of a given object, compare it to the values in the prepared database, find the set of descriptors that had the shortest Euclidean distance to the given object, and return the name of the file that belongs to the selected descriptors.

### 2.2.2 Moment Fourier Descriptors

After the development of the rotationally variant Fourier Descriptors, a variation of Fourier Descriptors called Moment Fourier Descriptors [5] was pursued. It was in the hopes of being able to develop descriptors that were more robust so that they could handle multiple objects, complex objects, and objects that do not necessarily possess smooth curves that MFD were developed.

Firstly, it is assumed that an object is composed of a set of closed regions in the complex plane. Secondly, the centroid of the object is found, and for N number of Moment Fourier descriptors, N angularly equispaced radial vectors are formed. The moments of these vectors are calculated by assuming that points that cross an object count towards the vectors moment, while points that cross the background are not taken into account. The

Fourier Transform is then performed on the vector moments, and the resulting Fourier coefficients are then normalized by the sum of Fourier Descriptors. The resulting descriptors are the Moment Fourier descriptors, and are invariant to rotation, translation, and scaling. Since it is desirable to have the descriptors rotationally variant, the amplitude of the Fourier coefficients are not calculated but instead the raw complex numbers are used, just as it is with the Rotationally Variant Fourier Descriptors.

### 2.2.3 Experimental Results

Since we are simply using the descriptors to find shape similarity and not pattern matching, a more subjective approach was taken during experimentation. Ten patterns that were not in the database were tested, and the shape of the selected image from the database was compared to the original image. It was found that the resulting descriptors lacked the accuracy that the original Rotationally Variant Fourier Descriptors possessed. It required at times up to three times the number of original descriptors in order to achieve the same amount of accuracy. This is most likely because the contour of the shape could not be as preciously tracked as with the original descriptors, and the fact that irregularities in the contour would be less evident when moments are used. Though it did prove useful for multi-part objects, and complex objects, accuracy was more important, and because of this, this module was left out of the Cyber Photographer's software.

## 3 Conclusion

This paper described a robot that can take pictures according to the compositional information taken from masterpieces, and the modules that were developed to improve the speed and accuracy of the image processing portions of it. Of the developed modules Peer-Group Filter was discarded because it was found that it provided no substantial aid to the segmentation of the image, while adding a substantial overhead to the Cyber Photographer's application.

Color Image Segmentation module was enhanced in speed by up to ten times while maintaining an accuracy comparable to the original algorithm by scaling down the image, J-value tiling, pixel growing by region dilation, HSV k-means colour quantization, addition of Sobel information, and merging region using colour information based in the HSV colour space.

Rotationally Variant Fourier descriptors were chosen over Moment Fourier descriptors because Moment Fourier descriptors required up to three times the number of original descriptors in order to achieve the same amount of accuracy. This is most likely because the contour of the shape could not be as preciously tracked as with the original

descriptors, and the fact that irregularities in the contour would be less evident when moments are used.

Future improvements to the Cyber Photographer's image processing modules include combining the Color Image Segmentation algorithm with a gradient based approach to segmentation so that patterns and textures can be used in unison with the existing algorithm to improving the accuracy of the segmentation. There exists also the possibility of using neural network to train the algorithm to recognize patterns, textures, and smooth gradients which would also increase the accuracy of the segmentation.

Further research should also be done on looking for descriptors that can handle multiple objects because by using only using one object in shape matching limits the range of masterpieces that can be applied to the system, and it could prove to be inaccurate since in many masterpieces the compositional information of multiple subjects are taken into account together, not separately.

## 4 References

[1] Shoji Tanaka, Seiji Inokuchi, Yuichi Iwadate, "A Figure Extraction Method based on the Color and Texture Contrasts of Regions," Proc. IEEE ICIAP'99, pp.12-17, 1999.
[2] Shoji Tanaka, Seiji Inokuchi, Yuichi Iwadate, "A Study on an Attractiveness Evaluation Model based on the Physical Features of Image Regions," IEICE Technical Report, Vol. PRMU99, No. 95, pp. 1-8, 1999.
[3] Y. Deng, B.S. Manjunath, "Color Image Segmentation," IEEE CVPR'99, Vol. 2, pp. 446-451, 1999.
[4] T. Crimmins, "A complete set of Fourier descriptors for two dimensional shapes," IEEE Trans, Syst. Man Cybern. 12, pp. 170-179, 1982.
[5] Shueen-Shyang Wang, P-Cheng Chen, Wen-Gou Lin, "Invariant Pattern Recognition by Moment Fourier Descriptor," Pattern Recognition, Vol, 27, pp. 1735-1742, 1994.
[6] Y. Deng, C. Kenney, M.S. Moore, and B.S. Manjunath, "Peer group filtering and perceptual color image quantization", *Proc. of ISCAS*, 1999.

## 5 Appendix

Colour Image Segmentation

CIS.c

Contains the main functions for calculating the J-Value of an image, valley determination, valley growing, pixel growing, and region merging.

color.c

Contains useful function for converting between various colour spaces. Including hsv, rgb, and Lab.

kma.c

Contains the algorithm for k-mean calculation.

ntsc.c

Contains the function to convert a colour image to a greyscale image.

sobel.c

Contains the function to calculate the Sobel values of an image.

scaleimage.c

Contains the function for scaling down an image by using bicubic interpolation.

Peer-Group Filter

PGF.c

Contains the functions to perform a the Peer-Group filter upon a given image.

Rotationally Variant Fourier Descriptors

Fourierd.c

Contains the functions to perform the Fourier transform on a given set of coordinates, and normalize them.

Fourierd2.c
Contains the functions to perform the Fourier transform on a given set of coordinates, and normalize them for a given set of two objects.

Fourierdn.c

Contains the functions to perform the Fourier transform on a given set of coordinates, and normalize them for a given set of any number of objects.

Database.c

Contains the function to build the compositional database.

Matches.c

Contains the function to calculate the Fourier descriptors of a given image and find the best match to that image given a specific compositional database.

Moment Fourier Descriptors

MFD.m

Contains the Matlab code for calculating the Moment Fourier Decriptors.