THE ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
FASCICLE XIV MECHANICHAL ENGINEERING, ISSN 1224-5615
2011

PROCESSING AND ANALYZING OBJECT IMAGES WITH THE VIEW OF OBJECT GRASPING

Prof.Dr.Eng.Staretu Ionel
Dr.Inf.Itu Alexandru
Drd. Inf. Moldovan Catalin
"Transilvania" University of Brasov

ABSTRACT

This paper presents a solution based on image processing techniques used for analyzing images of objects which will be grasped, using existing software technologies completed with specific software modules. Digital image processing, as part of the grasping strategy, is based on a number of Microsoft technologies used to manage the devices attached to a computer, and a series of image processing techniques. The analysis is made directly on the video sequence taken from a video device (web cam). From the video, sequence frames are extracted which are then converted into RGB format and transformed into digital images used for analysis.
Based on the existing technologies review, several necessary adjustments are identified to the proposed objective, and several specific software modules which can be conceived are described in details, and are then used to facilitate the implementation.
The sequence presented in this paper is a part of an ample process, that aims to optimize the grasping strategy used for anthropomorphic prehensors with fingers from robot equipment.

KEYWORDS: image processing, image analysis, grasping, pre-configuration, grippers

1. INTRODUCTION

Without identifying a sufficiently precise position, shape and size of the objects intended to be grasped, a fair and efficient grasping operation cannot be made with great accuracy. This requires an adequate solution for processing and analysis of grasping objects, which is based on existing software technologies complemented with specific software modules. In the context of different researches in this dirrection [1,2,3,5,7], in the current paper it is presented a practical method implemented for image processing and analyzing of objects intended to be grasped. This sequence is a part of a wider process that aims to optimize the grasping strategy used for anthropomorphic prehensors with fingers from endowment of robots[4,8].

2. IMAGE PROCESSING AND ANALYZING

2.1. Digital image processing

Digital image processing, as a main component in grasping strategy is based on a series of Microsoft technologies used for the management of special devices attached to a computer and a series of advanced image processing techniques. The analysis is done directly on the video sequence taken from a video capturing device (webcam). All frames are extracted from the video stream, which are first converted in RGB format, and are used to create digital images which are analyzed later in the process.

The process of working with a video (video stream) is exemplified in Fig. 1.

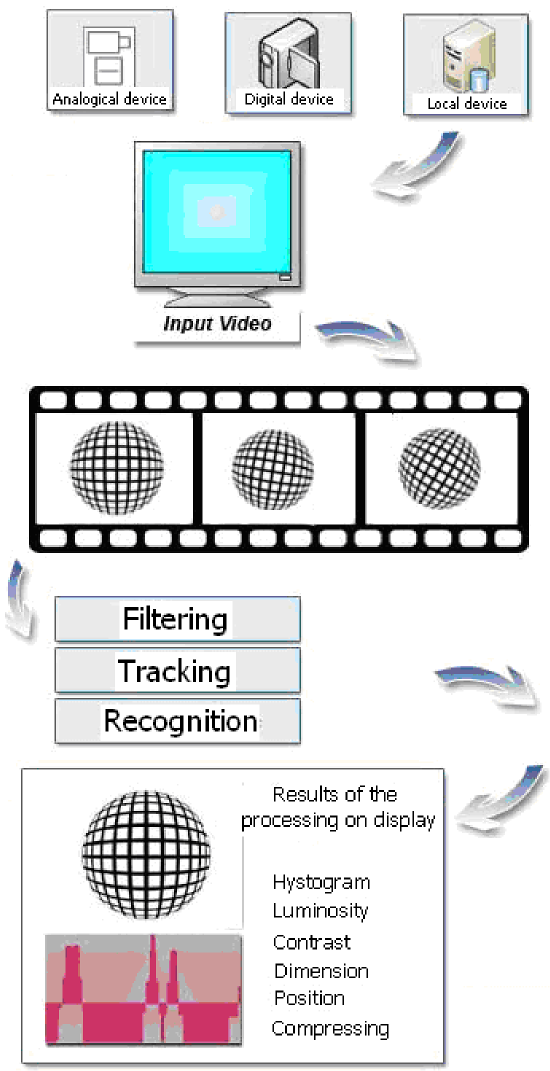

Fig. 1. Video sequence processing and analysis

The video frame analysis is concentrated on two main directions: on the one hand, image filtering, color transformations and optimization techniques for improving the color quality and, on the other hand, shape and form recognition alghorithms of the object which will be grasped. After all this data is known, the control module will pre-configure the prehensor: the shape of the object will define the pre-configuration type, and object dimension will influence the aperture of the gripper.

2.2. General aspects of digital image processing

Due to the extremely rapid evolution of computers performance in recent years especially in the power of computing and information processing speed, technologies for image processing and analyzing were developed and used in representative applications in various fields.

Image processing for subsequent analysis has the following main goals [7]:

- improving the quality of visual aspects of an image taking into account issues related to the exposure on the display or printing to paper;
- preparing the images for the identification and measurement of forms, some traits or characteristics of the structure;
- image compression / decompression for transmission / reception over a network

Image analysis is concerned with evaluations of certain parameters characteristic to images and with data extraction through automatic or semiautomatic methods. These operations are found in the literature under the names of feature extraction [5], [1], scene analysis [2], description of an image [3], automatic image interpretation [9].

The image analysis differs from image processing by the fact that the final result is not an image, but a value or a set of values called features (average, standard deviation, uniformity, entropy, moment invariations, etc.) extracted from specific models and mathematical relationships. This data, in this case, is used as inputs into a feature learning algorithm that identifies the objects shape and then creates classificators for each type of object. For this purpose, a series of filters are necessary with which the software application will identify and measure the general objects to be grasped. Thus, it has been developed a set of routines (commands, tools) for black and white conversion, binarization with variable threshold, color negatives, shape identification and measurement of object size. Likewise, video sequence analysis can be done only on a certain area, called region of interest (ROI). All these operations were optimized for a real time analysis so that the video stream received from the video device will retain the original number of frames per second.

2.2.1. Image Clasification

Depending of the acquiring mode and the presentation mode, images can be divided into two main categories:

• analogue images, usually presented on a physical medium (film camera, film, photographic paper, etc.)

• digital images, typically stored electronically (file).

Images are taken using conventional cameras, digital cameras or camcorders. Classical photo camera uses a thin pelicule from a photo film which is then developed, constituting the support through which the image can be transposed to a photographic paper. Digital cameras store the image information directly to a file. Digital images can be easily obtained from the analog images through a process called scanning [6].

Analogue Images are in fact an image presented on a 2D medium which is modeled mathematically with a function f (x, y) with two variables whose value is called the *intensity*, where x, respectively y are geometrical coordinates associated to the image points. The notion of intensity is encountered under the name *gray level* for monochrome images. Color images are formed by combining (overlapping) single-color images. For example, RGB images are composed from images associated with the three primary colors (Red, Green, Blue). Because of this, all processing and analysis techniques developed for 2D monochrome images can be applied to color images by processing individual components. It is believed that *an image is an analog image if mathematical functions associated to color components are continuous.*

The process of converting an analog image into a digital image involves sampling (quantization) to determine the intensity values in different points in the xy plane, called pixels (picture elements). Consequently, a digital image can be represented by an array of associated values to intensity function, usually two dimensional, with the dimension equal to M x N pixels:

$$[f(x,y)] = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,M-1) \\ f(1,0) & f(1,1) & \dots & f(1,M-1) \\ \vdots & \vdots & & \vdots \\ f(N-1,0) & f(N-1,1) & \dots & f(N-1,M-1) \end{bmatrix} \quad (1)$$

Usually in the programming practice, the matrix associated with a digital image is represented in this way:

$$F = \begin{bmatrix} f(1,1) & f(1,2) & \dots & f(1,M) \\ f(2,1) & f(2,2) & \dots & f(2,M) \\ \vdots & \vdots & & \vdots \\ f(N,1) & f(N,2) & \dots & f(N,M) \end{bmatrix} \quad (2)$$

Digital images can be obtained indirectly, through the scanning process of analog images from a physical medium (photographic film, photographic paper, etc..) or directly using a recording device (digital camera, digital camcorder, etc.) [6].

2.2.2 Specific parameters of digital images

In order to describe, in a quantifiable manner, digital images, it can be used the following parameters: coordinates (abscissa, ordinate) of a pixel, intensity, image size and resolution.

The **coordinates of a pixel** represent the set of values usually associated to the position of a point from the image plan.

Pixel intensity represents the value or set of values associated with different levels of colors emphasis from digital images. In computer programming, the corresponding pixel intensity values are associated with certain data types which can be numeric data (real or integer having predetermined upper and lower limits), of character type (color information for every pixel), or logical type (Table 1).

Table 1. Data types used to describe digital images programmatically

Type	Description	Nr.bytes	Interval
Double	Real numbers, double precision	8	[-10^{308}, 10^{308}]
Uint8	Integers without sign, 8 bytes	1	[0, 255]
Uint16	Integers without sign, 16 bytes	2	[0, 65535]
Uint32	Integers without sign, 32 bytes	4	[0, 4294967295]
Int8	Integers with sign, 8 bytes	1	[-128, 127]
Int16	Integers with sign, 16 bytes	2	[-32768, 32767]
Int32	Integers with sign, 32 bytes	4	[-2147483648, 2147483647]
Single	Real numbers, simple precision	4	[-10^{38}, 10^{38}]
Char	Character, Unicode representation	2	-
Bool	Logic, values are 0 or 1	1	-

Image size is expressed by a pair of pixel numbers on the abscissa (horizontal) and the ordinate (vertical) or the values associated with length units (centimeters or inch) on both directions.

Image resolution is the number of pixels reported to unit length (usually inches) measurable on both directions (horizontal and vertical) dpi (dots per inch). As the resolution increases, the image quality becomes better, but the amount of memory occupied by the associated image file will be larger. There is always an optimal resolution (maximum legibility), depending on how the digital images are represented (printed on paper, on screen display). Also in the category of specific image parameters **contrast** can be categorized, as a measure of the difference between illuminated and dark areas from an image.

2.2.3. Digital image description using RGB color space

RGB descriptor (TrueColor image) is achieved with a three-dimensional matrix M x N x 3 by associating to each pixel a set of three values reprezenting three shades of colors (R - Red), green (G - Green), blue (B - Blue). Thereby, it can be considered that the descriptor is composed of a three bidimensional matrix with the dimension M x N, one for each basic color (Fig. 2).

The number of bits reserved for a pixel is called a pixel color depth and using RGB description of an image, this will have the value 24 (each 8 bits per color basis). The total number of overtones resulting from the combination of 24-bit is 2^{24}, meaning more than 16 million colors (to be more accurate 16,777,216 colors). Therefore the images described in the RGB image space are called "natural" or "real " images. It is possible also to use 32-bit for image represntation. In this case the others 8 complementary bits are used to represent ALPHA component (transparency).

For this reason, this representation is called ARGB representation. Transparency is often used in certain graphic applications and gaming.

RGB space is often displayed as a cube (Fig. 2). Along the main diagonal are found the shades of gray from black to white.

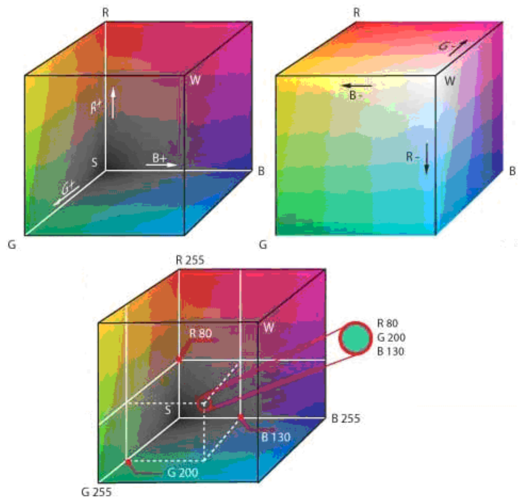

Fig. 2. RGB Space[10]

Starting from the way of how the images are defined, intensity descriptors and binary representations are used to define monochrome images, meanwhile those indexed, respectively the RGB representations are used for color images.

There are other ways to represent images. For example, in terms of the spectrum, the colors are divided into chromatic color (spectral color) and achromatic colors (white, black, gray) having the main characteristics hue, saturation and intensity. The color of an object depends on the spectral composition of light, the pigmentation and the subject which observe it.

2.2.4. Storing images on files

Regardless of how it was obtained, the matrix associated with a digital image is stored on a *specific file*. Some of the most popular graphical file formats are presented in Table 2.

Image processing applications have built in conversion functions to convert the image descriptors with different classes of data. The core inside the system used for recognition and measurement of objects, utilize a complex data structured created by specialist from Intel corporation, called *IplImage*. This way, used to store digital images in computer memory, allows fast access to information associated with a digital image such as pixel values, the number of channels, etc.

Table 2. Main image file formats

Format	Description	Extension file
JPEG	Joint Photographic Experts Group	.jpg, .jpeg
BMP	Windows BitMaP	.bmp
GIF	Graphics Interchange Format	.gif
TIFF	Tagged Inage File Format	tif, .tiff
PNG	Portable Network Graphics	.png

Once the image frame is captured and converted into RGB24 color space using the Sample Grabber component, this will be passed through a series of processing (filtering) aimed at distinguishing the object from the background, its recognition and subsequent measurement of size and 3D model generation. In Fig. 3 are presented the processings applied to an image taken from a video sequence realized through the software application.

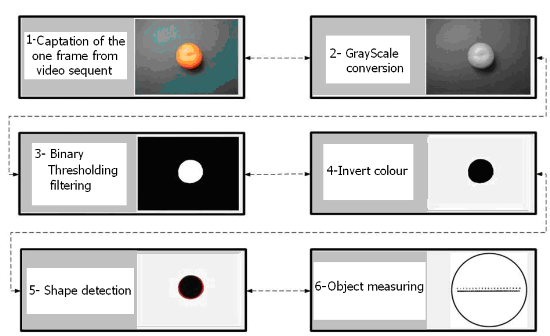

Fig. 3. Video sequence processing at the SampleGraber filter level

The image taken from the video sequence is converted into RGB color space so that to each pixel from the image there are assigned three values in the interval [0, 255] for the three components (Red - Red, Green - Green, Blue - Blue) as it can be seen from Relation 1. YUY2 format can be seen as YUV format.

$$R = \min((9525 * (Y - 16) + 13074 * (V - 128)) >> 13,\ 255)$$
$$G = \min((9525 * (Y - 16) - 6660 * (V - 128) - 3203 * (U - 128)) >> 13,\ 255) \quad (1)$$
$$B = \min((9525 * (Y - 16) + 16351 * (U - 128)) >> 13,\ 255)$$

For processing and analysis, the first operation consists in passing the image from the RGB color space into black and white space (grayscale). The conversion is done at each pixel using the following equation:

$$red=green=blue=0.299*red + 0.587 * green + 0.114 *blue \quad (3)$$

Image conversion to black and white space will have a huge impact on processing speed, shortening the total time by a rate of three. The next step consists in the binarization of image by defining a threshold that will filter the image in two parts, the background (black color) and the object to be grasped (white color). The threshold is calculated by the ratio of light and dark areas in the image. To better distinguish objects from the background it is applied another filter to reverse the color using the following equation:

$$valoare\_pixel=255-valoare\_pixel \quad (4)$$

The next step, object shape recognition is extremely important in the process of determining the preconfiguration of the prehensor and to generate the 3D model. In this step, object can be classified (sphere, parallelepiped, cylinder or cone) and it is possible to calculate its size in pixels.

3. CONCLUSIONS

A grasping operation depends crucially on the correct view of the object which will be grasped. After obtaining the image of that object, the techniques used for image processing and analysis plays a critical role in the process of recognition the object shape and pre-configuration of the gripper.

Image processing and analysis used with the view of grasping can be realized using existing software technologies adapted properly to this purpose and conceived of specific software modules.

REFERENCES

[1] **Belkasim, S. O., et al.**, *Pattern recognition with moment invariants: a comparative study and new results*, Pattern Recogition, nr. 24(12), pag. 1117, 1991.

[2] **Duda, R. O., Hart, P. E.,** *Pattern classification and scene analysis*, New York, Wiley, 1973.

[3] **Haralick,R.M., et al.,** *Textural features for image classification*, IEEE Transactions on Systems, Man and Cybernetics., SMC-nr. 3(6), pag. 610-621, 1973.

[4] **Itu, A.,** *Contribuţii la strategi de prehensiune în medii reale si virtuale folosind un prehensor antropomorf cu trei degete-* teză de doctorat, Universitatea Transilvania din Braşov, 2010.

[5] **Kittler, J.,** *Feature selection and extraction, în Handbook of Pattern Recognition and Image Processing*, Academic Press, New York, pag. 59, 1986.

[6] **Luculescu, M. C.,** *Cercetări asupra structurilor biologice vizuale umane privind diagnosticarea afecţiunilor maculare* – teză de doctorat, Universitatea Transilvania din Brasov, 2008.

[7] **Pratt W.,** *Digital Image Processing*, PIKS Inside, Third Edition, John Wiley & Sons, Inc., 2001.

[8] **Staretu, I.,***Gripping Systems*, Derc Publishing House, USA, 2011.

[9] **Viola, P. and Jones, M. J.,** *Robust real-time face detection*. International Journal of Computer Vision, 57(2):137--154, May 2004.

[10] **www:** http://en.wikipedia.org/wiki/RGB_color_space