

Computer Vision & Image Analysis

2.1. Introduction:

The current state of evaluation in textile fibrous structures depends on different tests and many of these tests require subjective evaluation by trained personnel. Therefore, modeling the fibrous structures in digitized form allows more objective study and evaluation. It also reveals some of the subtle properties that might be tedious for human operators to find systematically without mistakes. There are different systems for acquiring the physical fibrous structures to the digital world and some of these methods depend only on the projections of the structures or reconstruct a digital three-dimensional (3D) model for them. In this chapter, we will present some of the image acquisition techniques relevant to the current work, discuss the digital representation of images, and go over the stages required in image processing and computer vision (CV) to allow understanding the image content, then conclude with the challenges facing the current state of computer vision.

2.2. Object digitization technologies:

The types of images in which we are interested are generated by the combination of an “*illumination*” source and the *reflection* or *absorption* of energy from that source by the elements of the “*scene*” being imaged. We enclose *illumination* and *scene* in quotes to emphasize the fact that they are considerably more general than the familiar situation in which a visible light source illuminates a common everyday 3-D (three-dimensional) scene. For example, the illumination may originate from a source of electromagnetic (EM) energy such as radar, infrared, or X-ray source, or from less traditional sources such as ultrasound or even a computer-generated illumination patterns. Similarly, the scene elements could be familiar objects or being fibers, yarns, and fabric surface & structure, as emphasized in this study. Depending on the nature of the source, illumination energy is reflected from, or transmitted through, the objects. In some applications, such as Scanning Electron Microscopy (SEM), the reflected or transmitted energy is focused onto a photoconverter (*e.g.* a phosphor screen), which converts the energy into visible light.

2.2.1. Sensing elements within the visible range

Electromagnetic waves can be conceptualized as propagating sinusoidal waves of varying wavelengths and energies. If spectral bands are grouped according to their energy, we obtain a spectrum that ranges from gamma rays (highest energy) at one end to radio waves (lowest energy) at the other end. It is very common in many imaging applications to use illumination sources within the visual bands of the electromagnetic spectrum.

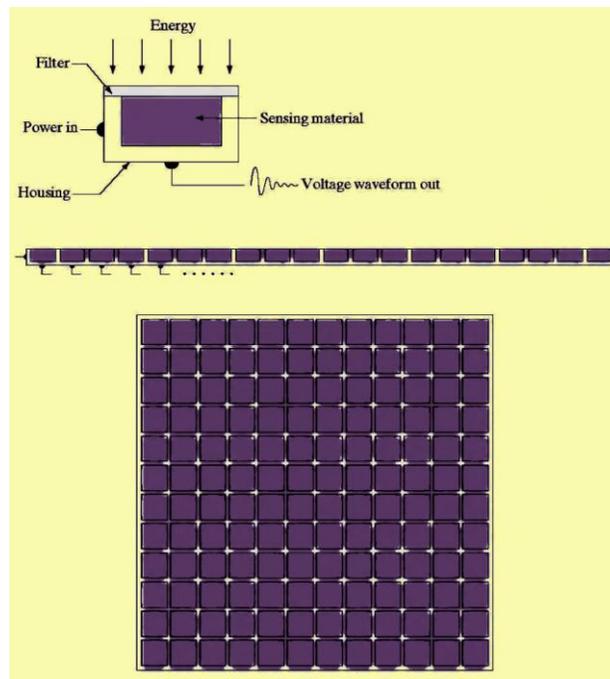


Figure 1. Arrangements of sensors for image acquisition: Single imaging sensor (top), Line sensor (middle), Array sensor (bottom) (Reproduced from Ref. [1])

The three principal sensor arrangements used to transform illumination energy into digital images are shown in Figure 1. The idea for a single sensor is that the incoming energy is transformed into a voltage by the combination of input electrical power and sensor material (e.g. silicon photodiode) that is responsive to the particular type of energy being detected. The output voltage waveform is the response of the sensor(s), and a digital quantity is obtained from each sensor by digitizing its response. The use of a filter in front of the sensor improves its selectivity; for example, a green (pass) filter in front of a light sensor favors light in the green band of the color spectrum results in a stronger output for green light than for other components in the visible spectrum.

The in-line arrangement of sensors in the form of a sensor strip provides imaging elements in one direction. The imaging strip gives one line of an image at a time, and this strip is implemented in two ways. Firstly, the object can move in front of the sensor strip and the obtained signal is accumulated and processed to build the global picture of the image. This technique is implemented in instruments for measuring yarn irregularity and yarn diameter such as the optical module on the recent versions of Uster Evenness Tester, and the Constant Tension Transport (CTT) tester produced by Lawson-Hemphill. Secondly, the object is static and the sensor strip moves perpendicular to its direction to complete the other dimension of the two-dimensional image. This is the type of arrangement used in most flatbed scanners and sensing devices with 4000 or more in-line sensors are possible. Some of our fabric surface images were digitized

using this technique which allows a uniform distribution of lighting on the fabric surface and high resolutions of the fabric's images are possible, an example of such image is shown in Figure 2.



Figure 2. Fabric surface as digitized using an in-line arrangement of sensors in a flatbed scanner

Numerous electromagnetic and some ultrasonic sensing devices are frequently arranged in an array format, such as the 2D array shown in Figure 1. This is also the predominant arrangement found in digital cameras such as the charge-coupled device (CCD) arrays, which can be manufactured with a broad range of sensing properties and can be packaged in rugged arrays of 4000 x 4000 elements or more. Since the sensor array shown in Figure 1 is two dimensional, its key advantage is that a complete image can be obtained by focusing the energy pattern onto the surface of the array. The principal manner in which array sensors are used is shown in Figure 3 where the energy from an illumination source being reflected from a scene object, but the energy also could be transmitted through the object. The first function performed by the imaging system is to collect the incoming energy and focus it onto an image plane. If the illumination is light, the front end of the imaging system is a lens, which projects the viewed scene onto the lens focal plane, as shown in Figure 3. The sensor array, which is coincident with the focal plane, produces outputs proportional to the integral of the light received at each sensor. Digital and analog circuitries sweep these outputs and convert them to a signal, which is then digitized by another section of the imaging system as diagrammatically shown in Figure 3.

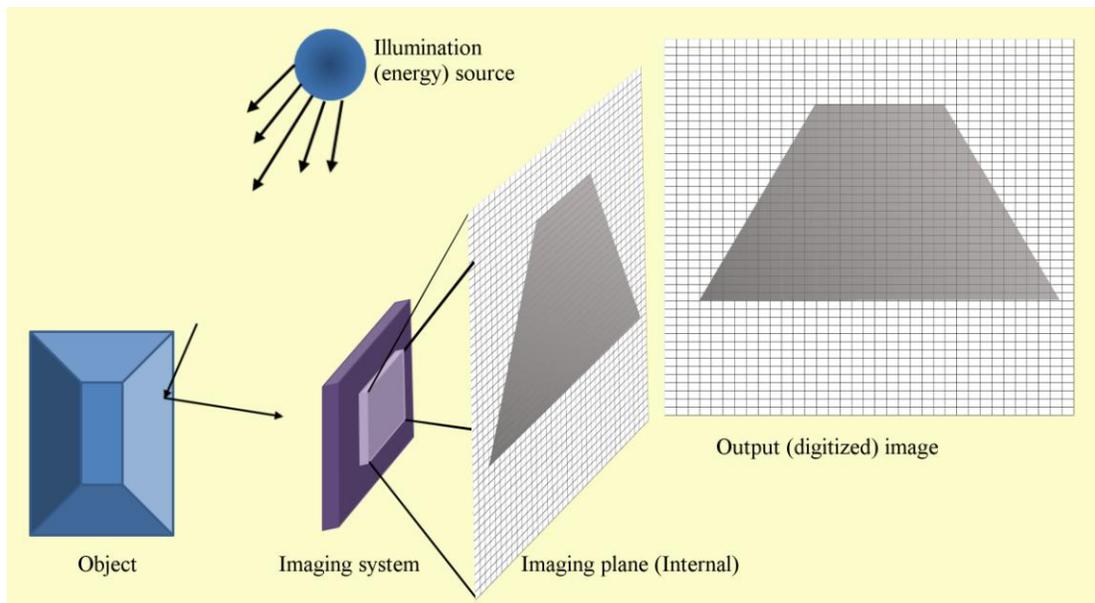


Figure 3. An example of the digital image acquisition process with a 2D array of sensors

In our work, we used this type of image acquisition many times where, for instance, a high speed CCD cameras was used to shoot the images of yarns running at a speed of 100 m/min as shown in Figure 4. In this work, four different instruments were synchronized to test the yarn properties simultaneously. The CCD camera was also installed on a microscope to digitize the microscopic images of fibers and yarns. This type of camera with the 2D arrays of sensors was also used during the evaluation of the fabric structures, fabric faults, and fabric appearance (especially for its pilling), as will be shown in the second part of this work that covers the applications.

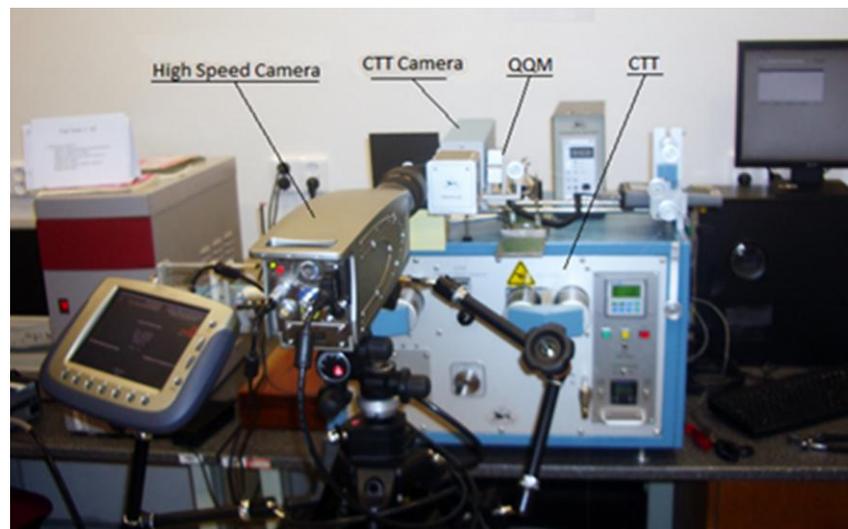


Figure 4. A high speed camera acquiring the images of yarn at 100 m/min

2.2.2. X-ray sensing elements

The previous methods of image acquisition and digitization mainly utilize illumination sources within the visible region of electromagnetic (EM) spectrum. In this section we will switch the focus to a more powerful technique in acquiring images by using the x-rays. X-rays are among the oldest sources of EM radiation used for imaging and best known use of x-rays is medical diagnostics, but they also are used extensively in industry and other areas, like astronomy. In our work, we recently utilize this technology to reconstruct digital models of fibrous structures and study the geometry and the internal structure of yarns with a neither destruction to the studied sample, nor the application of any chemicals that might affect its physical properties.

Computed Tomography (CT) is an imaging method that employs the tomography (from Greek words *tomos* means “slice” and *graphein* means “to write”) where “digital geometry processing” is used to generate a three-dimensional image of the internals of an object from a large series of two-dimensional X-ray images taken around a single axis of rotation. The conception of the CT idea started at the end of the 1960s and the first commercially viable CT scanner was invented in 1972 by Hounsfield who won the 1979 Nobel Prize in medicine for this work. The CT technology was basically implemented in medical imaging and the clinical CT became the radiology’s powerhouse as the first method to non-invasively acquire images of the inside of the human body. The method has evolved rapidly and was implemented in industrial fields by the end of the 1980s as one of the favorite Non Destructive Testing (NDT) techniques. The diversity of CT applications, with objects of different sizes, shifted the interest from large objects (as human bodies) to smaller ones and the need for the “higher spatial resolution” scanners started to emerge. The higher spatial resolution is obtained by either using clinical flat-panel imaging systems that achieve resolutions in the order of 150–200 μm or by using dedicated micro-CT ($\mu\text{-CT}$) scanners, such as the one used in this study, which can usually achieve a spatial resolution less than 0.5 μm .

The principle of CT scanning stems from the fact that the information available from a single projection of an object in engineering drawing is limited and another projection is necessary to obtain the third projection and ultimately reconstructing the 3D perspective of the object. This explanation in engineering drawing applies also to CT scanning, where a single x-ray projection shows a superimposition of all objects in the path of the X-ray and therefore hard to understand the volumetric structure of the object. The information can be increased by taking two (and more) projections, however, increasing the number of projection directions (views) is of little help because the observer is not able to mentally solve the superposition problem and to “reconstruct” the internal information of the object. Fortunately it can be shown that a complete “computed” reconstruction of the object’s interior is mathematically possible as long as a large number of views “*tomos*” have been acquired “*graphein*” over an angular range that covers an angle of at least 180° . This acquisition scheme is implemented in CT scanners by using an X-ray tube together with a detector while the object is rotating within path of the X-ray beams as demonstrated in Figure 5.

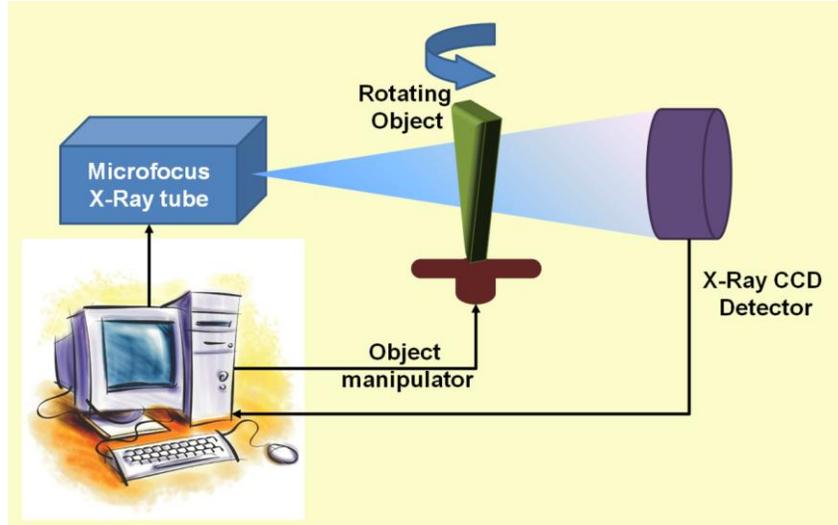


Figure 5. Schematic representation for the principle of CT scanning

Each x-ray projected image is a representation of the object's X-ray absorption along straight lines in a specific direction. For an incident x-ray with initial intensity I_0 , an object of thickness d , and attenuation coefficient μ , the number I of quanta reaching the detector is given by the exponential attenuation law:

$$I = I_0 e^{-\mu d} \quad (1)$$

The negative logarithm $p = -\ln I/I_0$ of each intensity measurement I gives information about the product of the object attenuation μ and its thickness d and proportionate to the tube current I (given a constant incident intensity I_0). For nonhomogeneous objects, the attenuation coefficient is a function of x , y , and z and the projection value p corresponds to the line integral along line L of the object's linear attenuation coefficient distribution $\mu(x, y, z)$:

$$p(L) = -\ln \frac{I(L)}{I_0} = \int_L \mu(x, y, z) dL \quad (2)$$

For flat-panel CT, the line L can be parameterized by the rotation angle α and the detector coordinates (u, v) . We are interested in gaining knowledge of $\mu(x, y, z)$ by reconstructing the acquired data $p(L)$ and the CT "image reconstruction" process is defined as the process of computing the image $f(x, y, z)$ as an accurate approximation to $\mu(x, y, z)$ from the set of measured projection values $p(L)$.

To simplify the mathematics behind the image reconstruction process, assume a single-slice CT scanner whose detector consists of one detector-row only. The raw data would be given by setting the longitudinal detector coordinate v to zero: $p(\alpha, u, 0)$. The easiest way to perform image reconstruction of these mid-plane data is to make a *change of variables* to obtain raw data

in parallel-beam geometry and replace the source position α and the detector position u by other variables: θ that represents the ray angle with respect to the coordinate system, and ξ that represents the rays distance to the center of rotation. This allows the formulation of $\hat{p}(\theta, \xi) = p(\theta, u, 0)$ and the new two variables are related according to the relation

$$x \cdot \cos(\theta) + y \cdot \sin(\theta) = \xi \quad (3)$$

The process of changing the variables to parallel geometry is known as *rebinning* and the parallel beam image reconstruction consists of a filtering of the projection data with the reconstruction kernel followed by a back-projection into image domain. This can be formulated mathematically as:

$$f(x, y) = \int_0^\pi d\theta \hat{p}(\theta, \xi) * k(\xi) |_{\xi=x \cdot \cos(\theta)+y \cdot \sin(\theta)} \quad (4)$$

Where $k(\xi)$ is the reconstruction kernel and there are different convolution kernels available (e.g. smooth, standard, and sharp...,etc) to allow modifying the image sharpness (spatial resolution) and the image noise characteristics. This process is called filtered back-projection (FBP) where the projection data are to be convolved with the reconstruction kernel $k(\xi)$. The filtered data are then back-projected into the image along the original ray direction for all ray angles θ .

The extension to cone-beam data, where $v \neq 0$ in general, is straight forward and known as Feldkamp-type image reconstruction and today's micro-CT image reconstruction algorithms are mainly of this type. In Feldkamp-type reconstruction, the variable v is ignored during the filtering step (as done above) but accounted for the true ray geometry during the back-projection by using a three-dimensional back-projection. The achievable spatial resolution of a given micro-CT scanner is mainly determined by the magnification factor, by the detector pixel size and by the size of the focal spot. A certain spatial resolution can only be achieved when it is not limited by the voxel (the volume element) size. Ideally, the voxel size should be half of the spatial resolution value or less. On the other hand, the reconstruction times can range from some minutes up to hours, depending on the number of voxels used. The time for image reconstruction is in the order of $O(N^4)$ when N projections are back-projected into a volume of size N^3 . Other techniques that cut down this effort to the order $O(N^3 \log N)$ are described in the literature, but still need more improvement to be included in the product implementation for μ -CT cone-beam reconstruction [2].

In one of our recent studies on the utilization of CT [3], an air-jet yarn sample was scanned and the obtained projected images were used to reconstruct the 3D digital model of the yarn as demonstrated in Figure 6. The digital model of the yarn was manipulated in different ways where the yarn was magnified without losing the resolution and details, as shown also in Figure 6. The details of the 3D model depend mainly on the resolution of the CT scanner during the image acquisition stage not just on the resolution of the image that is magnified. The presence of the 3D digital model allows some treatments such as clipping and cutting certain parts of the structure as

well as slicing the yarn structure at any required plane direction. It is, therefore, very useful to use this model and obtain cross-sectional images along the yarn length, as demonstrated in Figure 6, without the need to apply any additional chemicals such as the ones required for hardening the yarn before its physical slicing using sharp edges and the microtome.

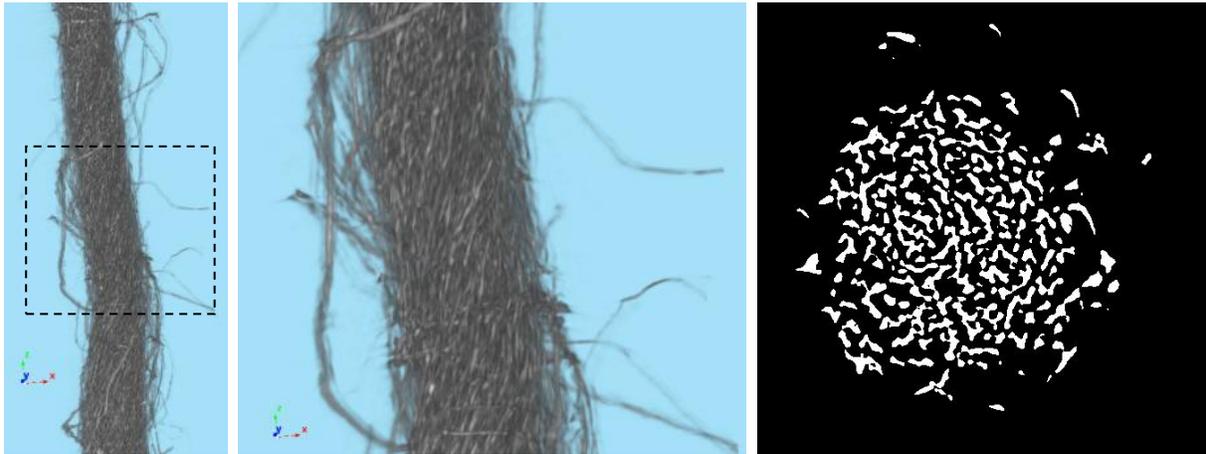


Figure 6. Reconstructed digital model of the yarn (left) that allows its visualization with a higher magnification, as the dotted rectangle (middle), and with cross-sectional slicing (right)

2.3. Digital image representation:

After the acquisition of images, they are represented in a digital form by two-dimensional functions of the form $f(x, y)$. The value or amplitude of f at spatial coordinates (x, y) is a positive scalar quantity whose physical meaning is determined by the source of the image. Most of the images in which we are interested in our studies are monochromatic images, whose values are said to span the gray-scale. Images can be also represented in color-space which “usually” means three channels for the same image that carry information on the red, the green and the blue levels in the image. For grayscale image that is generated from a physical process, its values are proportional to energy radiated by a physical source (e.g., electromagnetic waves). As a consequence, $f(x, y)$ must be nonzero and finite; that is,

$$0 < f(x, y) < \infty \quad (5)$$

The function $f(x, y)$ may be characterized by two components; the amount of source illumination *incident* on the scene being viewed, and the amount of illumination *reflected* by the objects in the scene. Appropriately, these are called the illumination and reflectance components and are

denoted by $i(x, y)$ and $r(x, y)$, respectively. The two functions combine as a product to form $f(x, y)$:

$$f(x, y) = i(x, y)r(x, y) \quad (6)$$

where

$$0 < i(x, y) < \infty \quad (7)$$

and

$$0 < r(x, y) < 1 \quad (8)$$

Equation (8) indicates that reflectance is bounded by 0 (total absorption) and 1 (total reflectance). The nature of $i(x, y)$ is determined by the illumination source, and $r(x, y)$ is determined by the characteristics of the imaged objects. It is noted that these expressions also are applicable to images formed via transmission of the illumination through a medium, such as x-rays. In this case, we would deal with a *transmissivity* instead of a *reflectivity* function, but the limits would be the same as in equation (8), and the image function formed would be modeled as the product in equation (6).

After the image $f(x, y)$ is *sampled* and *quantized* so that the resulting digital image has M rows and N columns and the values of the coordinates (x, y) become discrete quantities. Integer values are used for these discrete coordinates and the values of the coordinates at the origin, for example, are $(x, y) = (0, 0)$. The next coordinate values along the first row of the image are represented as $(x, y) = (0, 1)$. It is important to keep in mind that the notation $(0, 1)$ is used to signify the second sample along the first row and it does not mean that these are the actual values of physical coordinates when the image was sampled. Figure 7 demonstrates this coordinate convention.

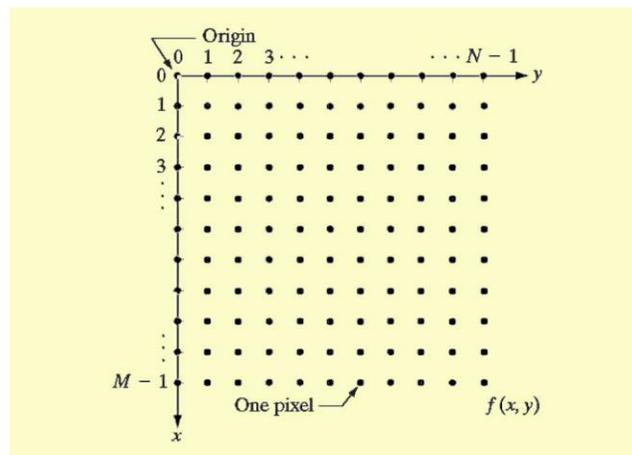


Figure 7. Coordinate convention used in representing digital images (Reproduced from Ref. [1])

The notation introduced in the preceding paragraph allows us to write the complete $M \times N$ digital image in the following compact matrix form:

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0, N-1) \\ f(1,0) & f(1,1) & \dots & f(1, N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1, N-1) \end{bmatrix} \quad (9)$$

The right side of this equation is the digital image and each element of this matrix array is called an image element, picture element, pixel, or pet. In some discussions, it is advantageous to use a more traditional matrix notation to denote a digital image and its elements:

$$A = \begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0,N-1} \\ a_{1,0} & a_{1,1} & \dots & a_{1,N-1} \\ \vdots & \vdots & & \vdots \\ a_{M-1,0} & a_{M-1,1} & \dots & a_{M-1,N-1} \end{bmatrix} \quad (10)$$

It is clear that, $a_{i,j} = f(x = i, y = j) = f(i, j)$, so equations (9) and (10) are identical matrices.

Once the image is represented in a matrix form as expressed in equation (10), most of the mathematical rules and operations of linear algebra can be performed on the image. One of the very attractive features that drove me to work with images analysis is the possibility to make a sense of some mathematical calculations that can be performed in multi-dimensional space. For example, we start learning the rules of calculus in two-dimensional space with a single independent variable then generalize that to multi-variable calculus in multi-dimensional space. The advantage of the 2D space is that it allows us to graphically visualize different definitions such as the meaning of a tangent or the meaning of the integration of the curve...,etc. The generalization of these rules to multi-variables is challenging because of the lack of the graphical interpretation of the mathematical calculations. Digital imaging analysis in its 2D representation presents the intermediate step between the one dimensional and the multi-dimensional data where the rules can be applied and visualized at the same time. An example of this is the Fast Fourier Transform (FFT) which is usually introduced in 1D and it becomes familiar and easy to be understood with a 1D set of data, but 2D FFT is possible to be understood in treating the digital images. Another example is the application of the variational methods and the optimization of multi-variable functions that are necessary in most of the quantum mechanics applications. It is easier to understand these variational methods and their mathematical complexity when you deal with a 2D set of data (in the form of an image) such as our work on Chan-Vese segmentation model [4].

It is also important to note that the digital representation of an image might be deceiving for the human eye and might be hard to understand. Let us look at Figure 8, for example, which shows in its upper part an image of a woven fabric that is very familiar to the human eye and can be recognized easily, however it is more difficult for human to interpret the other two representations for the same image of the woven fabric that are shown at the bottom of Figure 8. The other two images in the bottom of Figure 8 are just two representations of the same fabric image by two different ways; one representation of the brightness (gray-scale) values of the picture and the other as a representation of the fabric image in the frequency domain after the application of the Fast Fourier Transform (FFT). It is clear from Figure 8 that; although all representations contain exactly the same information, it is very difficult for a human observer to find a correspondence between them, and without the prior knowledge of the image presented in the top of the figure, it is unlikely that one would recognize the fabric structure from the other two representations.

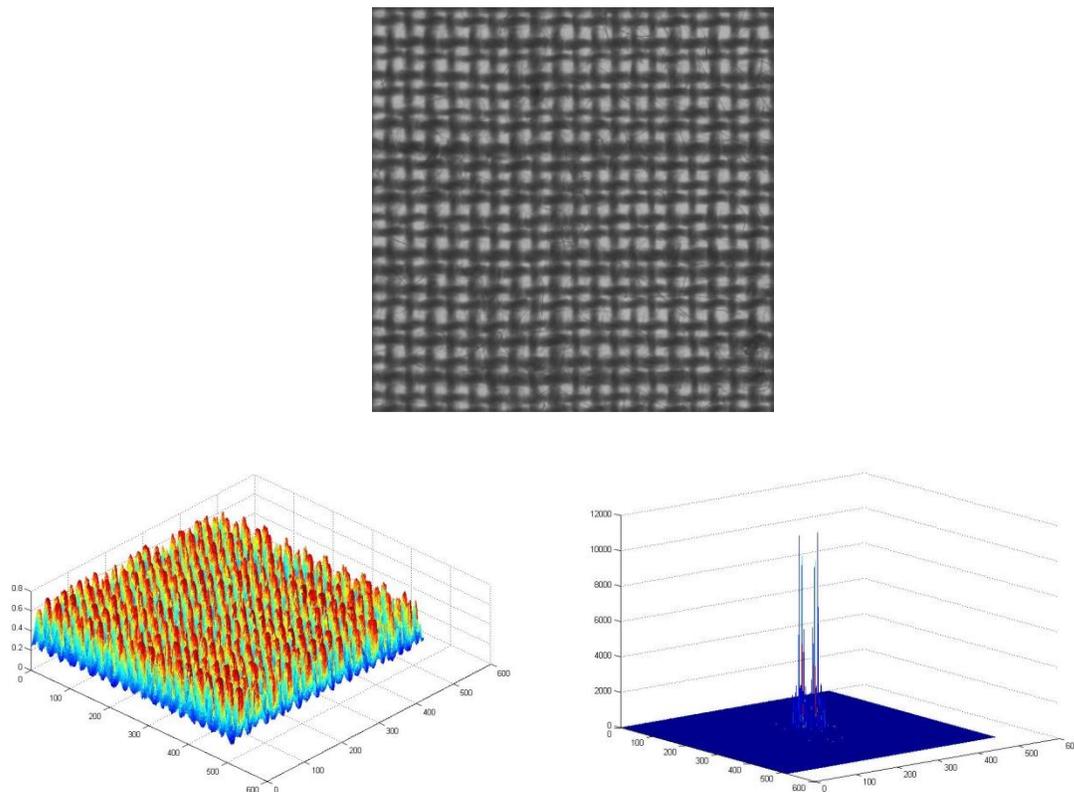


Figure 8. 2D image of a woven fabric (top) with two different representations of the same image as expressed by the gray levels (bottom left) and the Fourier transformed image (bottom right) (Note; the main peak at the center of the FFT image was suppressed to present the other peaks)

2.4. Image understanding:

Humans (and other living animals) use different senses to interact with their environment and one of the most important senses is the vision as it allows humans to *perceive* and *understand* the world surrounding them. Computer vision is the technology that aims at duplicating the effect of human vision by electronically *perceiving* and *understanding* that digital image. Image understanding by a machine can be seen as an attempt to find a relation between input image(s) and previously established models of the observed world. Transition from the input image(s) to the model reduces the information contained in the image to relevant information for the application domain. This process is usually divided into several steps and levels where the bottom layer contains raw image data and the higher levels interpret the data. Computer vision designs the intermediate representations and algorithms serving to establish and maintain relations between entities within and between the layers.

Image representation can be roughly divided according to data organization into four levels, as shown in Figure 9. The boundaries between individual levels are inexact, and more detailed divisions are also proposed in the literature. Figure 9 suggests a bottom-up way of information processing, from signals with almost no abstraction, to the highly abstract description needed for image understanding. Note that the flow of information does not need to be unidirectional; often feedback loops are introduced which allow the modification of algorithms according to intermediate results.

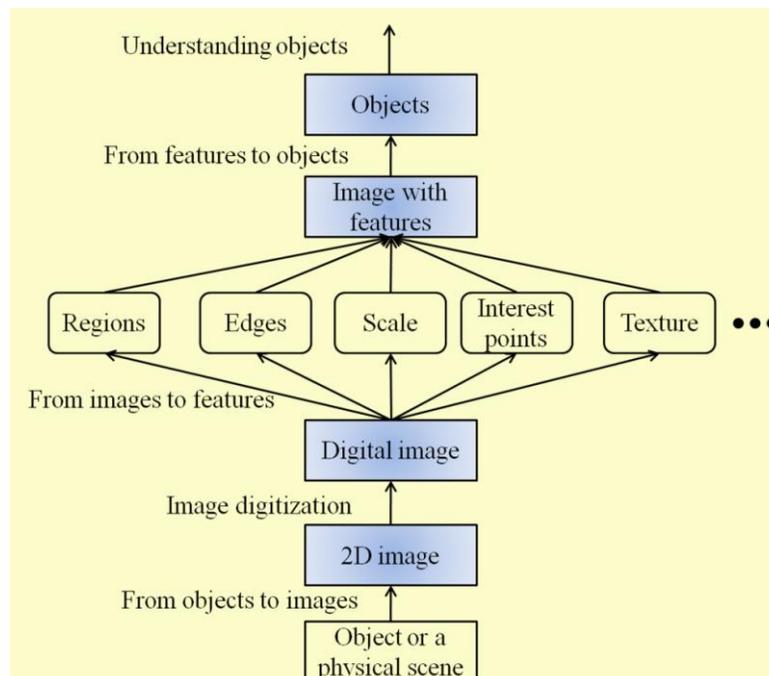


Figure 9. Different image representation (shaded rectangles) suitable for image analysis problems in which objects have to be detected and classified (Reproduced from Ref. [5])

This hierarchy of image representation and related algorithms is frequently categorized in a simpler way with *low*, *medium*, and *high level* image processing and understanding, as shown in Figure 10. Although the borders between these levels are vague (especially the low and medium levels) and some techniques might be categorized in one or another, the low-level processing methods are usually distinguished by using very little knowledge about the content of images. In the case of image components' recognition by a computer, it is usually provided by high-level algorithms or directly by a human who understands the problem domain. The *Low* and *medium* level methods often include image compression, pre-processing methods for noise filtering, edge extraction, and image sharpening. *Low* and *medium* level image processing uses data which resemble the input image; for example, an input image captured by a camera is 2D in nature, being described by an image function $f(x, y)$ whose value, at simplest, is usually brightness depending on two parameters x and y the coordinates of the location in the image.

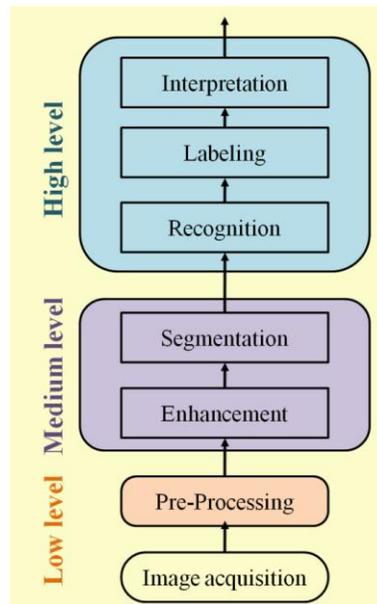


Figure 10. Basic steps in image recognition and interpretation

Low and medium level computer vision techniques overlap almost completely with digital image processing, which has been practiced for decades. The following sequence of processing steps is commonly seen: An image is captured by a sensor (such as a CCD camera) and digitized; then the computer suppresses noise (image pre-processing) and maybe enhances some object features which are relevant to understanding the image. Edge extraction is an example of processing carried out at this stage. Image segmentation is the next step, in which the computer tries to separate objects from the image background and from each other. Total and partial segmentation may be distinguished: total segmentation is possible only for very simple tasks, an example being the recognition of dark non-touching objects from a light background. In more complicated problems (the general case), medium-level image processing techniques handle the partial

segmentation tasks, in which only the cues which will aid further high-level processing are extracted. Often, finding parts of object boundaries is an example of low-level partial segmentation. Object description and classification in a totally segmented image are also understood as part of the medium-level image processing. Other medium-level operations are image compression, and techniques to extract information from (but not understand) moving scenes. An example of the low and medium levels processing of the longitudinal view of a yarn as well as its cross-sectional view is shown in Figure 11.

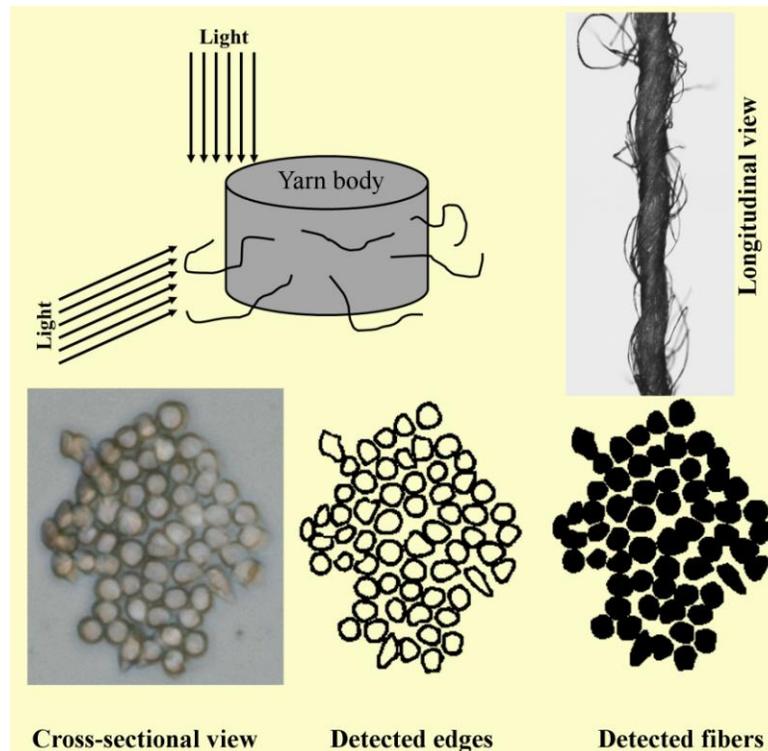


Figure 11. Low and medium level processing of digital yarn image

Most current low-level image processing methods were proposed in the 1970s or earlier. Recent research is trying to find more efficient and more general algorithms and is implementing them on more technologically sophisticated equipment; in particular, parallel machines are being used to ease the enormous computational load of operations conducted on image data sets. The requirement for better and faster algorithms is fuelled by technology delivering larger images (better spatial resolution), and color. A complicated and so far unsolved problem is how to order low-level steps to solve a specific task, and the aim of automating this problem has not yet been achieved. It is usually still a human operator who finds a sequence of relevant operations, and domain-specific knowledge and uncertainty cause much to depend on this operator's intuition and previous experience.

Low-level data are comprised of original images represented by matrices composed of brightness (or similar) values, while *high-level* data originate in images as well, but only those data which are relevant to high-level goals are extracted, reducing the data quantity considerably. High-level data represent knowledge about the image content; for example, object size, shape, and mutual relations between objects in the image. The high-level vision system includes three steps: recognition of the objects from the segmented image, labeling of the image and interpretation of the scene. Most of the artificial intelligence tools and techniques are required in high level vision systems. Recognition of objects from the image can be carried out through a process of pattern classification, which at present is realized by supervised learning algorithms (e.g. artificial neural networks). The interpretation process, on the other hand, requires knowledge-based computation (e.g. fuzzy logic controllers). Therefore, high-level processing is based on knowledge, goals, and plans of how to achieve those goals which tries to imitate human cognition and the ability to make decisions according to the information contained in the image. High-level vision begins with some form of formal model of the world, and then the ‘reality’ perceived in the form of digitized images is compared to the model. A match is attempted, and when differences emerge, partial matches (or sub-goals) are sought that overcome the mismatches; the computer switches to low-level image processing to find information needed to update the model. This process is then repeated iteratively, and ‘understanding’ an image thereby becomes cooperation between top-down and bottom-up processes. A feedback loop is introduced in which high-level partial results create tasks for low-level image processing, and the iterative image understanding process should eventually converge to the global goal.

Although the task of understanding the objects in an image is challenging in computer vision, the digital representation of an image in the form of a numerical matrix might also be challenging for human to understand the image. An example of this is the representation of the fabric image and its other two derivatives shown in Figure 8 and having the same amount of information. Therefore, understanding the digital image requires a lot of a priori knowledge by humans to interpret the images; while, on the other hand, the machine only begins with an array of numbers and will be attempting to make identifications and draw conclusions from data. Internal image representations are not directly understandable, while the computer is able to process local parts of the image, it is difficult for it to locate global knowledge. General knowledge, domain-specific knowledge, and information extracted from the image will be essential in attempting to ‘understand’ these arrays of numbers.

2.5. Closing remarks on CV challenges:

Computer vision involves sequence of operations that are characteristic of image understanding; such as: image capturing, early processing, segmentation, model fitting, motion prediction, qualitative / quantitative conclusions. Giving computers the ability to see is a challenging task and research in the field faces many obstacles. Examples of the computer vision challenges are:

- The *loss of information* where we live in a three-dimensional (3D) world and the available electronic visual sensors (*e.g.* cameras) used to digitize this 3D world usually give two-dimensional (2D) images. This projection to a lower number of dimensions is associated with an enormous loss of information where the projected images (in 2D) are capable of *mapping* points along rays but does not preserve angles and collinearity. Therefore, it is not enough to have a picture of an object to figure-out its dimensions and a *reference scale* is needed to differentiate between large and small objects in the acquired image. This challenge directed us to use some of the advanced techniques such as the computed tomography scanning to reconstruct a reliable model of our fibrous structures.
- The image *interpretation* is another challenge because humans bring their previous knowledge and *experience* to the current observation when they try to understand an image. Human ability to *reason* allows representation of long-gathered knowledge and its use to solve *new* problems. From the mathematical logic and/or linguistics point of view, interpretation of images can be seen as a mapping:

Interpretation: image data → model

The (logical) model means some specific world in which the observed objects make sense and there are several possible interpretations of the same image(s). An example might be: nuclei of cells in a biological sample, rivers in a satellite image, or parts in all industrial process being checked for quality. Introducing interpretation to computer vision allows us to use concepts from mathematical logic, linguistics as *syntax* (rules describing correctly formed expression), and *semantics* (study of meaning). Considering observations (images) as an instance of formal expressions, semantics studies relations between expressions and their meanings. The interpretation of image(s) in computer vision can be understood as an instance of semantics. Artificial intelligence (AI) has invested several decades in attempts to endow computers with the capability to understand observations; while progress has been tremendous, the practical ability of a machine to understand observations remains very limited.

- Another challenge is the *noise* that is inherently present in all measurements in the real world, but its existence in images requires different mathematical tools which are able to filter out the noise and deal with the uncertainty in the acquired images. Complex mathematical tools make the image analysis more challenging compared to the standard (deterministic) methods. It is also useful to use fuzzy logic controllers in such problems which allow some degree of uncertainty during the decision making about the presented data.
- The *redundancy of data* (too much data) is a challenge in computer vision due to the nature of huge data collected from images and video sequences. To get a sense of the data acquired, an example of non-compressed A4 sheet of paper scanned monochromatically at 300 dots per inch (dpi) at 8 bits per pixel will correspond to about 8.5 mega bytes (MB). Non-interlaced RGB 24 bit color video 512 x 768 pixels, 25 frames per second, makes a data stream of 225 Mb per second. If the image and video processing we perform is not very simple, then it is hard to achieve a real-time performance (i. e. to process 25 or 30 images per second) with traditional poor hardware. Therefore, it is very encouraging to apply new methods that reduces the amount required calculations such as our introduced method [6] to analyze the tremendous amount of ictures produced from the high speed camera imaging.
- The physics involved in image formation interferes with the *measured brightness* in images to form another challenge in computer vision. The radiance (brightness, image intensity) depends on the irradiance (light source type, intensity and position), the observer's position, the surface local geometry, and the surface reflectance properties. This is the reason why image capturing physics is often avoided in practical attempts aiming at image understanding, and a direct link between the appearance of objects in scenes and their interpretation is sought.
- The *scope* of the image is another challenge where a big difference can be observed by viewing an image through local window or through global view. Image analysis algorithms commonly analyze a particular storage bin in an operational memory (e.g. a pixel in the image) and its local neighborhood; the computer sees the image through a keyhole. Seeing the world through a keyhole makes it very difficult to understand more global context of the image.
- *Dynamic scenes* such as those to which we are accustomed, with moving objects or a moving camera, are increasingly common and represent another way of making computer vision more challenging.

References:

- [1] R. C. González and R. E. Woods, *Digital image processing*. Prentice Hall, 2002.
- [2] W. Semmler and M. Schwaiger, *Molecular Imaging I*. Springer, 2008.
- [3] M. Eldessouki and S. Ibrahim, “Computed Tomography Application For investigating The Internal Structure Of Air-Jet Yarns,” in *the 20th International Conference: Structure and Structural Mechanics of Textiles*, 2014.
- [4] M. Eldessouki and S. Ibrahim, “Chan-Vese Segmentation Model For Faster And Accurate Evaluation of Yarn Packing Density,” (*In Press. Text. Res. J.*
- [5] T. Svoboda, J. Kybic, and V. Hlavac, *Image Processing, Analysis & and Machine Vision- A MATLAB Companion*. Thomson Learning, 2007.
- [6] M. Eldessouki, S. Ibrahim, and J. Militky, “A Dynamic and Robust Image Processing Based Method for Measuring Yarn Diameter and Its Variation,” *Text. Res. J.*, vol. 84, no. 18, pp. 1948–1960, 2014.