# Blur spot limitations in distal endoscope sensors

Avi Yaron [a], Mark Shechterman[b] and Nadav Horesh[b]

[a]Visionsense Inc. Orangeburg, NY.
[b]Visionsense Ltd. Petah-Tiqva, Israel.

## ABSTRACT

In years past, the picture quality of electronic video systems was limited by the image sensor. In the present, the resolution of miniature image sensors, as in medical endoscopy, is typically superior to the resolution of the optical system. This "excess resolution" is utilized by Visionsense to create stereoscopic vision. Visionsense has developed a single chip stereoscopic camera that multiplexes the horizontal dimension of the image sensor into two (left and right) images, compensates the blur phenomena, and provides additional depth resolution without sacrificing planar resolution. The camera is based on a dual-pupil imaging objective and an image sensor coated by an array of microlenses (a plenoptic camera). The camera has the advantage of being compact, providing simultaneous acquisition of left and right images, and offering resolution comparable to a dual chip stereoscopic camera with low to medium resolution imaging lenses. A stereoscopic vision system provides an improved 3-dimensional perspective of intra-operative sites that is crucial for advanced minimally invasive surgery and contributes to surgeon performance.. An additional advantage of single chip stereo sensors is improvement of tolerance to electronic signal noise.

**Keywords:** plenoptic camera, stereoscopy, lenticular array, image resolution, miniature camera

## 1. INTRODUCTION

Two physical factors determine the resolution of a digital camera: the camera objective lens and the pixel-size of the image sensor. In general, the optical resolution of well-designed lenses (in the absence of size constraints, such as a digital camera) is superior to the resolution of the most advanced image-sensors, which therefore requires the use of anti-aliasing filters. When size and costs are constraints, the image sensor cost can be reduced by reducing the chip-area. On small format sensors, it is necessary to reduce the pixel size in order to maintain acceptable image resolution. For instance, PAL/NTSC signals are provided by $\leq 1/6$" format chips with pixel size of $\leq 3\mu$.

In comparison, there are applications where the resolution of the image sensor is, in many cases, superior to the resolution of the optical system. These are usually low cost / low quality optics in hand-held devices such as cell-phone cameras, or high depth-of-field applications such as medical endoscopy. This "excess resolution" is used to great advantage by Visionsense technology to create stereoscopic vision in surgical endoscopes that meet the needs of the medical market. As the human brain is highly sensitive to depth information the image understanding and analysis abilities are notably enhanced by a stereoscopic vision.

### 1.1. Medical endoscopy

Minimally invasive surgery (MIS) began when surgeons classically trained in "open" surgical procedures started to use long, thin imaging devices and surgical instruments to perform operations through small incisions. Early on, many technical factors were down played or overlooked because endoscopic surgical procedures seemed, at the time, intuitively similar to open procedures. Experience, however, has shown that the use of endoscopic instruments requires unique eye-hand coordination. In particular, the use of video cameras and monitors greatly affects the perception of physical reality and therefore performance.

Vision systems in currently available MIS instruments provide 2 dimensional (2D) images that lack depth perception, which restrict the surgeon's perspective and ability to perform complex manipulations. Recently published clinical papers[1,2] have documented that severe errors made during laparoscopic procedures are due to a critical misinterpretation of the video image, not simply errors in surgical technique. A review of surgical injury to the common bile duct found that the damage was caused by misinterpretation of the endoscopic image and incorrect decisions based on false perceptual information. In defense of the surgeons, many of us assume that our eyes are a reliable tool to interpret reality, and we overlook the crucial fact that video images have many limitations that can create a false sense of genuineness.

The flatness of conventional 2D imaging does nothing to augment a surgeon's performance, whereas 3 dimension (3D) stereoscopic vision can enhance image understanding and improve performance in laparoscopic surgery. The Visionsense sensor technology enables natural vision without discomfort, and it affords better results and enhanced confidence among less-experienced surgeons. In addition, stereoscopic vision paves the way for the development of new MIS procedures that use sophisticated articulating surgical instruments.

### 1.2. Stereoscopic plenoptic camera

The idea of a stereoscopic camera based on a lenticular array ("integral fly-eye") was first proposed by the French physicist G. Lippmann.[3] The Visionsense adaptation of this stereoscopic plenoptic camera is shown in Figure 1: The imaging objective is represented by a single lens (L) with a two pupils openings at the front focal plane (P). This setup generates a telecentric objective, in which all the rays passing the center of each pupil emerge as a parallel beam behind the lens. The CCD chip is covered by a lenticular array (LA) — an array of a cylindrical microlenses with zero power axis perpendicular to the paper plane. Each lenticule covers exactly two pixel-columns. Rays that pass through a point at the left aperture ($l$) are emitted as a parallel beam (dashed lines in the drawing) after the imaging lens. These rays are focused by the lenticular-array on the pixels on the right side under the lenslets (designated by a dark rectangles). Similarly, rays that pass through the right aperture ($r$) (dashed-dotted lines) are focused by the lenslets on the left ("dark") pixels. Thus a point O on the object is imaged twice: Once trough the upper aperture generating an image on pixel $o_1$, and once through the lower pupil generating an image on pixel $o_2$. The pixels $o_1$ and $o_2$ are upper and lower views (in the real world — left and right views) of the point O on the object. The distance between a pixel of the left view to the that of the right view (disparity) is a function of the distance of the corresponding point from the camera. The drawing in figure 2 emphasizes the relative alignment of the pupils, LA, and the image sensor pixels.

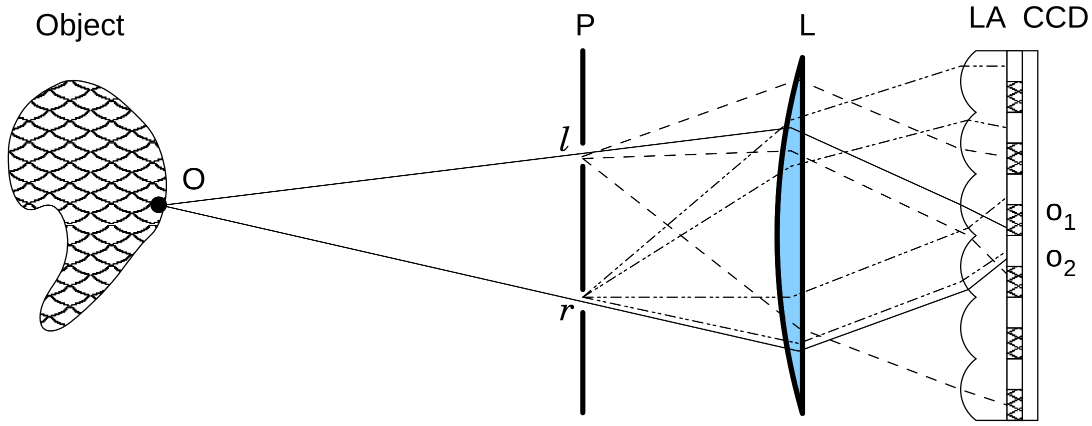
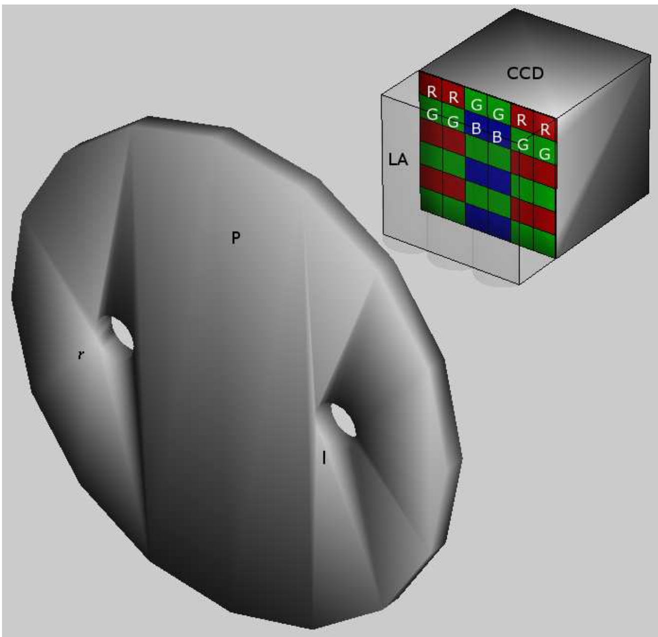



**Figure 1.** Visionsense stereoscopic camera scheme: The camera is composed of a mask with two openings (pupils) in the front focal plane P, and imaging objective (L) and a CCD with a cylindrical microlens array (LA) where each lenslet is 2 pixels wide. Ray from the object that pass through the left pupil ($l$) generate a left view on the "white" pixel columns, and rays that pass through the right pupil ($r$) generate a right view image on the "white" pixel columns. The telecentricity of the imaging objective is not mandatory, but it simplifies the optics by forcing an exact rather then variable registration of the LA to the CCD.

## 2. PLENOPTIC CAMERA FOR MEDICAL ENDOSCOPY

### 2.1. Adaptation of the optical parameters to the environmental conditions

Medical endoscopy have several parameters that are not common in general photography:



[bhp]

**Figure 2.** A 3D illustration of the camera in the scheme in figure 1 without the lens. It illustrates the relative alignment of the image sensor pixels and lenticular array (represented by a small portion of them) and the pupils of the objective lens. Note the arrangement of the colour filters, which is a modified version of the Bayer pattern. After the separation of the odd and even columns (left and right images), each image acquires a Bayer pattern.

1. Illumination is not usually a problem: A Bright Xenon arc-lamp illuminates the scene via a fiber-optics bundle.
2. High depth-of field (DOF) is essential: Objects must be in focus within a range of few millimeters to several centimeters.
3. Very wide field-of-view (FOV) from 70° to 140°.
4. Smallest possible camera diameter: To minimize the incision for camera insertion and reduce the disturbance to the surgical tools.
5. Video images are utilized as a part of a feedback loop: The surgeon uses the information he sees in real-time. The most important information are anatomic and pathologic markers as well the position of the surgical instruments relative to the tissue.

Items 1 and 2 leads naturally to a high f/# camera. Items 3 (especially from FOV $> 100°$) and 4 can be best fulfilled with a simple distal (mounted at the front end of the endoscope) camera. The last item leads to a preference of a stereoscopic vision as noted above.

For these reasons Visionsense developed a single-chip stereoscopic camera, with an optical system similar in size and shape to a conventional 2D system.

### 2.2. Single-chip stereoscopic image-capture

Two principal ways of dividing the resources of a single chip to create stereo images are by splitting either time or chip area:

**Time domain:** The field-sequential format, for example, requires a device that opens one pupil (left or right) every even frame and the other pupil every odd frame. The common devices for this task are either miniature LC shutter or MEMS shutters. We found both techniques unable to cope with the demands for a very small diameter ($< 4$ mm), large FOV and fast response time ($< 100\,\mu$sec).

**Chip area division:** These methods dedicates half of the chip area to an image from the left pupil, and the other half to an image from the right pupil. Although one could assume that this type of multiplexing provides half the image resolution in each eye, because every image contains half the pixels of the image sensor, under the conditions of medical endoscopy this is not necessarily the case.

## 3. IMAGE RESOLUTION OF PLENOPTIC CAMERA

A plenoptic camera can take advantage of the small pixel sizes enabled by the new VLSI design rules to encode a stereoscopic image on the image processor. Thus the surplus sampling frequency can be converted to an essential depth indication for the surgeon.

### 3.1. Factors affecting the resolution of a digital camera

The image resolution of a digital camera is determined mainly by 2 physical factors:*

1. The MTF of the imaging objective
2. The number and size of pixels

The point-spread-function (PSF) of a diffraction-limited lens is given by[4]

$$\mathrm{PSF}(x, y) = \frac{\mathrm{J}_1(\rho)}{\rho}, \tag{1}$$

where $\mathrm{J}_1$ is a first-kind Bessel function, and $\rho$ is given by

$$\rho = \frac{\pi r}{\lambda f}. \tag{2}$$

Here $r$ is the polar-coordinates radius ($r \equiv \sqrt{x^2 + y^2}$), $\lambda$ is the wavelength and $f$ is the lens f/#. The 1D PSF is given by

$$\mathrm{PSF}(x) = \int_{-\infty}^{\infty} \mathrm{d}y \, \mathrm{PSF}(x, y). \tag{3}$$

The 1D MTF is just the Fourier-transform of the 1D PSF:

$$\mathrm{MTF}(X) = \mathcal{F}\{\mathrm{PSF}(x)\}. \tag{4}$$

In the plenoptic camera there are two imaging systems:

1. The imaging objective which has a high f/# (in the range of 8–20), and therefore having a large (several pixels wide) PSF.
2. The lenticular array with a low f/# — on the order of 1.5, which enables a focusing of the light into one pixel.

The pixel density of the imager determines the image sampling rate. In an ideal cases the value of the MTF at the Nyquist-frequency and above should be zero. But the MTF of a diffraction-limited lens (and of nearly all imaging objectives) decreases gradually and never decays down to 0. In practice the MTF at the Nyquist frequency should be within 5–10% of its peak value. Nowadays small-formats image sensors contain pixel size in the range of 1.8–3 $\mu$. Pixels of this range of size offer surplus Nyquist frequency as shown in figure 4.

*There are several other factors such as cross-talk between adjacent pixels (which is not significant in high-quality imaging chips) and the pixels fill-factor figure, which is a minor contributor in most cases. These factors are out of this scope of this article.

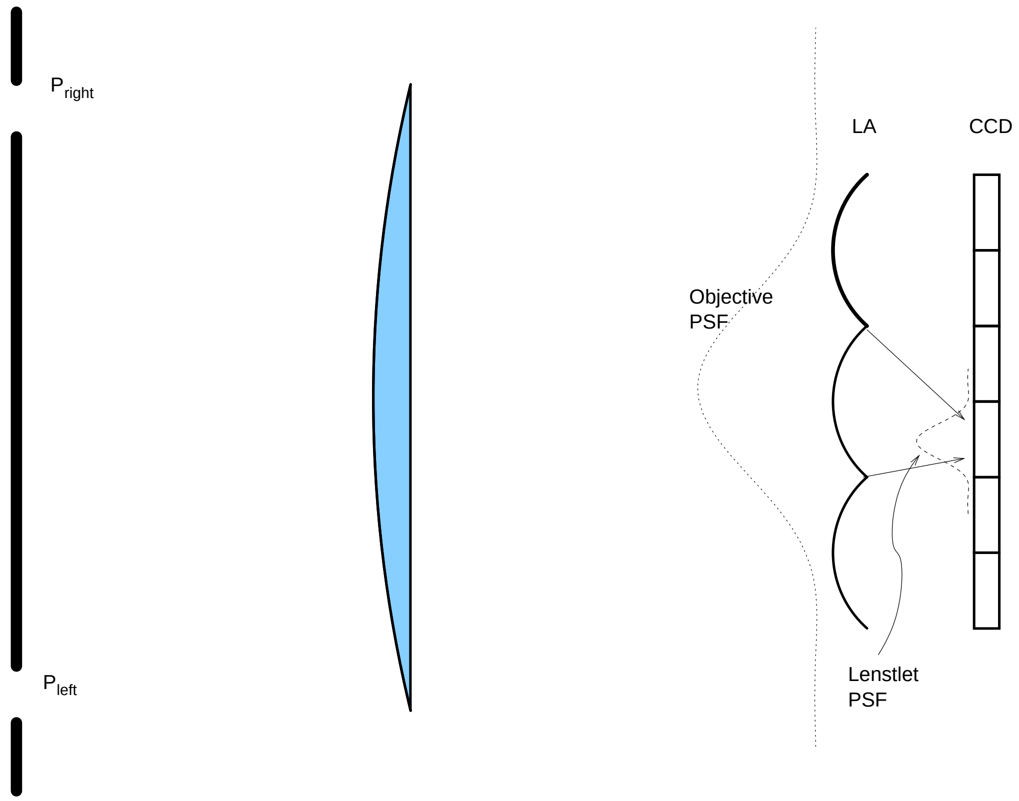



**Figure 3.** point spread functions in Visionsense stereoscopic camera. The light entering from the pupils ($\mathsf{P}_{\mathsf{left}}$ and $\mathsf{P}_{\mathsf{right}}$) is focused by the imaging objective. The objective's PSF is much wider then the lenslet diameter. Each lenslets generate an image of the pupils at their focal plane (in a telecentric objective the pupils are imaged at infinity). The lenslets have a low f/# — in the order of 1.5, therefor the pupils image falls within the width of one pixel.

## 3.2. Plenoptic camera resolution

This section provides a closer look at the Visionsense's camera operation. Consider figure 4: The light penetrates through the pupils ($\mathsf{P}_{\mathsf{left}}$ and $\mathsf{P}_{\mathsf{right}}$) and focused by the imaging objective. The imaging objective PSF (dotted line) is significantly wider then a lenslet. In comparison, the focus of a lenslet (which is in fact the image of a pupil), is smaller then a pixel width. Thus a portion of the image, which is projected across a lenticule (of width of two pixels) is refocused into a single pixels column. This results in a 2:1 horizontal compression of the image.

Consider a *not* state of the art image sensor, with a pixel dimension of $2.5\,\mu$: Its Nyquist frequency — $0.2\,\mu^{-1}$ is far beyond required for f/# of 10 or higher. Now, if we would apply a LA upon the sensor, it would cut its horizontal Nyquist frequency into half (there are two pixels per one lenticule), so the Nyquist frequency would become $0.1\,\mu^{-1}$, of which a f/#=16 lens MTF would generate no aliasing. If that image sensor would be used with an $f/\# = 14^{\dagger}$ objective, it would achieve the same image resolution of 100 lp/mm in either a plain camera, or a stereoscopic plenoptic camera.

The image resolution $R$ can be estimates by

$$R = \min(\mathrm{MTF}_{0.05}, \mathrm{Nyq}) \tag{5}$$

where $\mathrm{MTF}_{0.05}$ is the frequency where the MTF drops to 5% of its peak value, and Nyq is the Nyquist frequency of the image sensor. Figure 5 shows the image resolution due to the pixel size and the diffraction-limited MTF for several f/#.

---

[†]This issue is elaborated in figure 5.

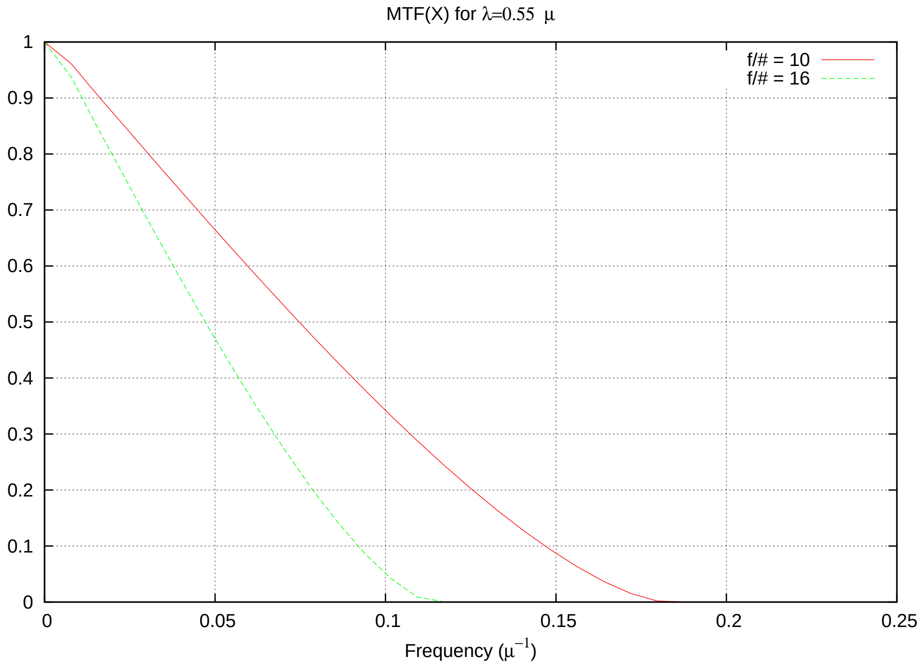



**Figure 4.** MTF of a diffraction limited optics for two f/# (10 and 16) for green light ($\lambda = 0.55\,\mu$). For an image sensor with pixel a dimension $d$ across, the Nyquist frequency would be $1/2d$ on the $x$ axis. For instance a sensor with pixel size of $2.5\,\mu$ would have a Nyquist frequency of $0.2\,\mu^{-1}$.

### 3.3. Improving depth resolution

There are two simple ways to transform a conventional objective into a telecentric one:

1. Place the aperture on the objective front focal plane as shown in Figure 1
2. Add a field-lens in the proximity to the lenticular array.

Visionsense has chosen the first scheme because it facilitates an additional improvement. Future studies will develop additional features that take advantage of the front-mounted dual aperture mask used in Visionsense systems. For example, depth resolution and stereo visualization can be further enhanced by increasing the distance between the dual aperture (the inter-pupillary distance of the device) using periscopic prisms. The periscope overcomes size limitations imposed by the physical geometry of the miniature camera.

## 4. CONCLUSIONS

The drive for visualization systems with the greatest number of pixels (more megapixels!) has not necessarily led to improved image quality. In fact, simply increasing the number and decreasing the size of pixels, in miniature cameras, will result in diminished improvement of image quality. In a pixel size that is much smaller than the blur spot, image quality and pixel count are not synonymous.

It is possible, however, to apply unique technological innovations to these miniature camera systems to further improve resolution and provide additional useful visual information.

The dual aperture effectively create two slightly offset images in a single device. Light from each aperture passes through a single objective lens, which has a large point spread function. Light next passes through a lenticular array of micro lenses that each focus sharply onto 2 pixels of the sensor. The visual data from the sensor is then processed by software that reconstructs the stereoscopic image.

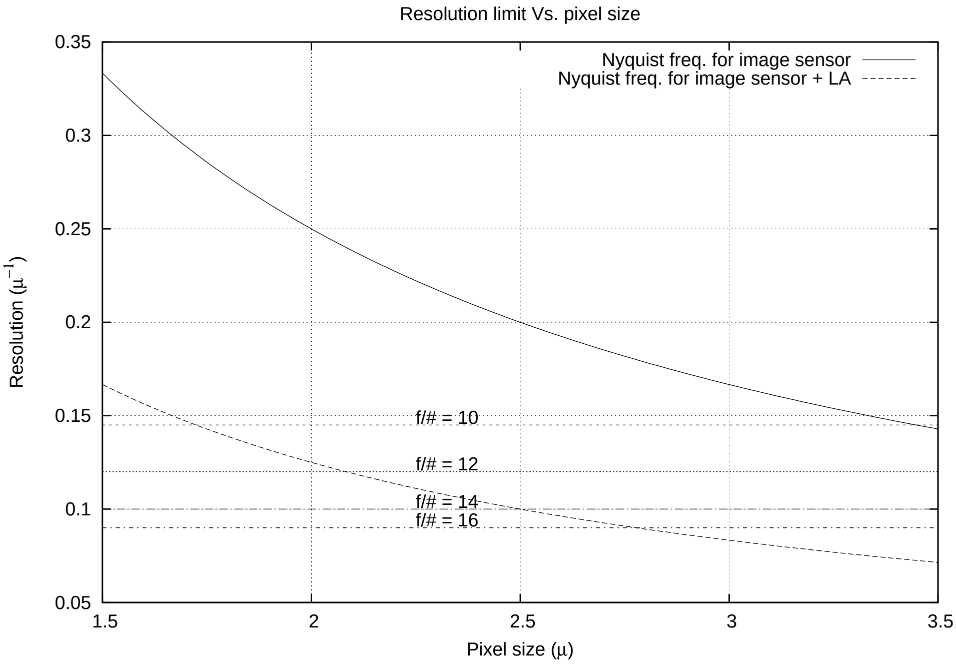



**Figure 5.** Resolution limit Vs. pixel size for several objective f/#: The curves are the Nyquist frequencies as a function of the pixel size. The upper curve plots the Nyquist frequency of "plain" camera, while the lower curve describes a plenoptic camera. The horizontal lines are the 5% cut-off of the MTF for the designated label f/#. The pixel-size range where the LA adds stereo without resolution lost is where the horizontal lines lies below the curves. For example the line of $f/\# = 14$ cuts the lower curve at pixel-size of $2.5\,\mu$, and the upper at $5\,\mu$ (out of the plot's range). This means that at pixel size $\leq 2.5\,\mu$ plenoptic camera would provide stereoscopic image without a resolution loss, at size range 2.5–5 $\mu$ "plain" camera would not suffer resolution loss, while plenoptic camera would suffer from aliasing.

In conclusion, the Visionsense plenoptic visualization system provides high resolution stereo images from a compact device. The small, single camera dual aperture device provides true stereo vision, with an efficient usage of image-sensor area. This technology can be used to great advantage in applications that demand high quality images from miniature cameras.

## REFERENCES


1. K. B. Buckingham, "When laparoscopic gallbladder surgery goes wrong," *Trial* **35**, pp. 20–27, 1999.
2. J. Clavete, L. Sabater, B. Camps, A. verdú, A. Gomez-Portilla, J. Martín, M. A. T. B. Flor, and S. Lledó, "Bile duct injury during laparoscopic cholecystectomy, myth of reality of the learning curve?," *Surg. Endosc.* **14**, pp. 608–611, 2000.
3. G. M. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys.* **7**, pp. 821–825, 1908.
4. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley & sons, 1978.