Grant Agreement no. 640174

PHySIS

Sparse Signal Processing Technologies for HyperSpectral Imaging Systems

**INSTRUMENT: Bottom-up space technologies at low TRL
OBJECTIVE: COMPET-06-2014**

D3.1 Analysis and evaluation of video SSI architectures

Due Date of Deliverable:

31st July 2015

Completion Date of Deliverable:

Start date of project: 1st March 2015   Duration: 24 months

Lead partner for deliverable: **IMEC**   Revision: Final

Project co-funded by the EC within the Horizon 2020 Programme		
Dissemination Level		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including Commission Services)	
CO	Confidential, only for members of the consortium (including Commission Services)	

Document History

Issue Date	Version	Changes Made / Reason for this Issue
20 July 2015	V0.9	Initial draft
29 July 2015	V0.95	Added information regarding data
3 August 2015	V1.0	Added description of link between compressed sensing & detector utilization

Document Main Author(s): Bert Geelen (IMEC), Andy Lambrechts (IMEC), Murali Jayapala (IMEC) , Greg Tsagkatakis (FORTH)

Document signed off by: Panagiotis Tsakalides (FORTH)

Contents

1. Introduction

1.1. Scope

In deliverable D3.1 of WP3, we provide an evaluation overview of various HSI technologies, focusing especially on technologies capable of generating hyperspectral video, i.e., *Snapshot Spectral Imaging (SSI)*. Strengths and weakness of each of the approaches are highlighted.

1.2. Purpose

The purpose of this deliverable is to provide insights to focus our efforts in the remainder of the project on technologies suitable for hyperspectral video, as well as demonstrating the potential of hyperspectral video via a description of a rich collection of data acquired under carefully controlled conditions. This report will provide valuable input to the rest of the project activities (compression, spatial/spectral resolution enhancement) that will result in high performance and compact snapshot video systems.

1.3. Applicable documents

[AD 01]   PHySIS_Proposal-SEP-210155336

1.4. Referenced Documents

[1] Vose, M. and Horton, D., "A heuristic technique for CTIS image reconstruction," Applied Optics, 46(26), 6498-6503 (2007).

[2] Hagen, N. and Dereniak, E., "New grating designs for a CTIS imaging spectrometer," Proc. SPIE, 6565 (2007).

[3] Johnson, W., Wilson, D. and Bearman, G., "Spatial-spectral modulating hyperspectral imager," Applied Optics, 45(9), 1898-1908 (2006).

[4] Gehm, M.E., John, R., Brady, D.J., Willett, R.M. and Schulz, T.J., "Single-shot compressive spectral imaging with a dual-disperser architecture," Optics Express, 15(21), 14013-14027 (2007).

[5] Wagadarikar, A., John, R., Willett, R. and Brady, D., "Single disperser design for coded aperture snapshot spectral imaging," Applied Optics 47(10), B44-B51 (2008).

[6] Gao, L., Kester, R., Hagen, N. and Tkaczyk, T., "Snapshot Image Mapping Spectrometer (IMS) with high sampling density for hyperspectral microscopy," Opt. Express, 18(14), 14330-14344 (2010).

[7] Kester, R., Gao, L. and Tkaczyk, T., "Development of image mappers for hyperspectral biomedical imaging applications," Applied Optics, 49(10), 1886-1899 (2010).

[8] Harvey, A., Fletcher-Holmes, D., Kudesia, S. and Beggan, C., "Imaging spectrometry at visible and infrared wavelengths using image replication," Proc. SPIE 5612, 190-198 (2004).

[9] Hariharan, P., [Basics of Interferometry], Elsevier (2007).

[10] Hirai, A., Inoue, T., Itoh, K. and Ichioka, Y., "Application of multiple-image Fourier transform spectral imaging to measurement of fast phenomena," Opt. Rev. 1(2), 205–207 (1994).

[11] Kudenov, M. and Dereniak, E., "Compact real-time birefringent imaging spectrometer," Opt. Express 20, 17973-17986 (2012).

[12] Tanida, J. and Yamada, K., "TOMBO: thin observation module by bound optics," Lasers and Electro-Optics Society (LEOS), (2002).

[13] Hartley, R. and Zisserman, A., [Multiple View Geometry in Computer Vision], Cambridge University Press (2004).

[14] Ng, R., Levoy, M., Bredif, M., Duval, G., Horowitz, M. and Hanrahan, P., "Light Field Photography with a Hand-Held Plenoptic Camera," Stanford University Computer Science Tech Report CSTR 2005-02, (2005).

[15] Lumsdaine, A. and Georgiev, T., "The Focused Plenoptic Camera", ICCP, (2009).

[16] Geelen, B., Tack, N. and Lambrechts, A., "A snapshot multispectral imager with integrated tiled filters and optical duplication," *Proc. SPIE* 8613, Advanced Fabrication Technologies for Micro/Nano Optics and Photonics VI, (2013).

[17] Geelen, B., Tack, N. and Lambrechts, A., "A compact snapshot multispectral imager with a monolithically integrated, per-pixel filter mosaic," Proc. SPIE 8974, Advanced Fabrication Technologies for Micro/Nano Optics and Photonics VII, (2014).

[18] Hagen, N. and Kudenov, M., "Review of snapshot spectral imaging technologies," Optical Engineering, 52(9), (2013).

[19] Kandylakis Z., Karantzalos K., Doulamis A. and Doulamis N., "Multiple Object Tracking with Background Estimation in Hyperspectral Video Sequences," IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), (2015).

[20] Arnold, T., de Biasio, M. and Leitner, R., "Hyperspectral video endoscopy system for intra-surgery tissue classification," in Sensing Technology (ICST), 2013 Seventh International Conference on, (2013).

[21] E. Honkavaara, H. Saari, J. Kaivosoja, I. Polonen, T. Hakala, P. Litkey, J. Makynen, and L. Pesonen, "Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight uav spectral camera for precision agriculture," Remote Sensing, vol. 5, no. 10, pp. 5006–5039, (2013).

[22] Tochon, G., Chanussot, J., Gilles, J., Dalla Mura, M., Chang, J. and Bertozzi, A. "Gas Plume Detection and Tracking in Hyperspectral Video Sequences using Binary Partition Trees," in IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2014), Lausanne, Switzerland, (2014).

[23] Banerjee, A., Burlina, P. and Broadwater, J., "Hyperspectral video for illumination-invariant tracking," in Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS '09. First Workshop, (2009).

[24] Van Nguyen, H., Banerjee, A. and Chellappa, R.,"Tracking via object reflectance using a hyperspectral video camera," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, (2010).

[25] Wikipedia, "Netpbm format," https://en.wikipedia.org/wiki/Netpbm_format

[26] Exelis Visual Information Solutions, ENVI http://www.exelisvis.com/ProductsServices/ENVIProducts/ENVI.aspx

[27] Veganzones, M., "How to: Load ENVI hyperspectral image data with MATLAB," http://www.ehu.eus/ccwintco/uploads/d/dc/LoadHypercubesMatlab.pdf

[28] Opticks, "ENVI Header Format" http://opticks.org/docs/help/4.3.1/Help/Opticks/Content/Importers_Exporters/ENVI_Header_Format.htm

1.5. Definitions, acronyms and abbreviations

CASSI   Coded Aperture Snapshot Spectral Imager

CTIS   Computed Tomography Imaging Spectrometry

FPA   Focal Plane Array

HWP   Half-Wave Plate

IMS   Image Mapping Spectrometry

IRIS   Image Replicating Imaging Spectrometer

MAFC   Multi Aperture Filtered Camera

NP   Nomarski Prism

SHIFT   Snapshot Hyperspectral Imaging Fourier Transform Spectrometer

SRDA   Spectrally Resolving Detector Arrays

SSI   Snapshot Spectral Imaging

2. Snapshot Spectral Imaging

Snapshot spectral imaging (SSI) is essentially capturing a 3D data cube corresponding to 2 spatial dimensions (x & y) and an additional dimension for spectral information. Snapshot spectral video imaging on the other hand extends the 3D data-cube to a 4D-data cube I(x, y, λ, t), including the time-dimension, as shown in Figure 1.
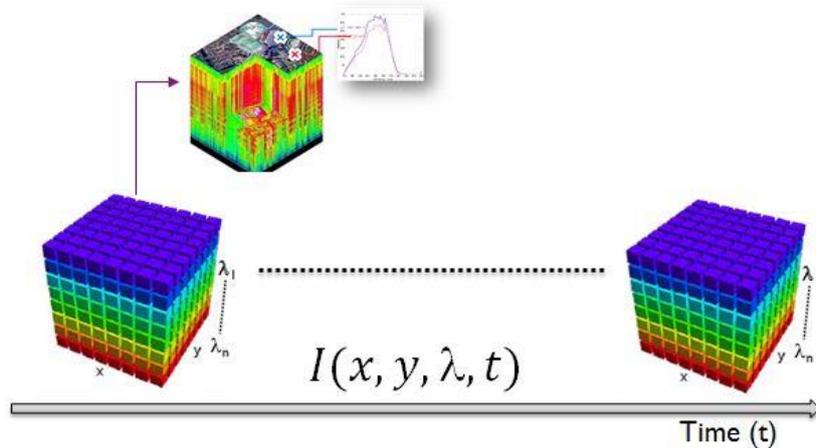
Figure 1: Snapshot video illustration I(x, y, λ, t), where each frame in time is a 3D-spectral data cube

Many different techniques have been developed to enable spectral imaging. Typically a scanning-based solution is used, where at each point in time a subset of the cube is sensed, requiring scanning over the remaining dimensions to fully acquire the spectral cube. This is a major distinction compared to snapshot spectral cameras where the entire cube is acquired at one distinct point in time. As with general scanning spectral imagers, a wide range of approaches may be taken to extract spectral information from the scene and to map pixels of a spectral cube to the 2D sensor plane.

3. Overview of SSI Architectures

3.1. Computed Tomography Imaging Spectrometers (CTIS)

Computed Tomography Imaging Spectrometers (CTIS) use advanced gratings (Computer-Generated Holograms, CGH, see Figure 3) to produce overlapping projections of a spectral cube on a 2D sensor (Figure 2). Then complex algorithms, related to computed tomography (CT)-reconstruction algorithms, are used to extract a spectral cube from this overlapping projected data. Drawbacks of this approach include the complexity of the reconstruction algorithms rendering real-time visualization impossible and the limited resolution of the spectral cube, with pixel counts of unrolled cubes varying from 4.2% of the size of the sensor [1] to 22-64% of the size of the sensor [2]. Moreover, its main limitation as a snapshot instrument is due to the nature of its ill-posed process of acquisition of limited angle projection tomography. Spectral reconstruction accuracy and resolution are thus highly scene dependent and further depend on the iterative algorithm and iteration stopping criteria [2][3].

Figure 2: CTIS tomographic dispersion pattern, showing a complex landscape in panchromatic 0th order in the center, while angular chromatic dispersion creates the surrounding higher orders [3].
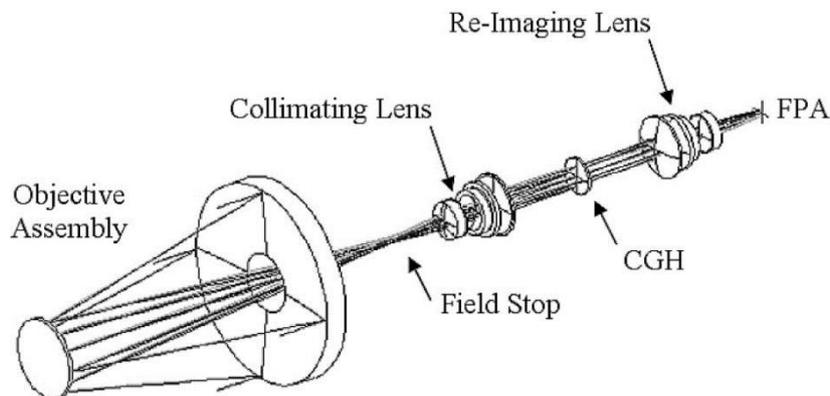
Figure 3: Generalized optical layout of CTIS [3]

3.2. Coded aperture snapshot spectral imagers (CASSI)

Coded aperture snapshot spectral imagers (CASSI) [4][5] attempt to overcome the limits of the spatial versus spectral resolution multiplexing trade-off by undersampling the scene spatially in each band and using compressive sensing to reconstruct the full spatial resolution. This reconstruction again relies on complex algorithms, rendering real-time visualization impossible. Moreover, the compressive sensing concept replaces the spatial versus spectral resolution trade-off with a signal-dependent spatial resolution versus image quality trade-off, which makes the resulting quality unpredictable and typically introduces spatial and spectral reconstruction artefacts. Consequently, the benefit of the gained spatial resolution is reduced.
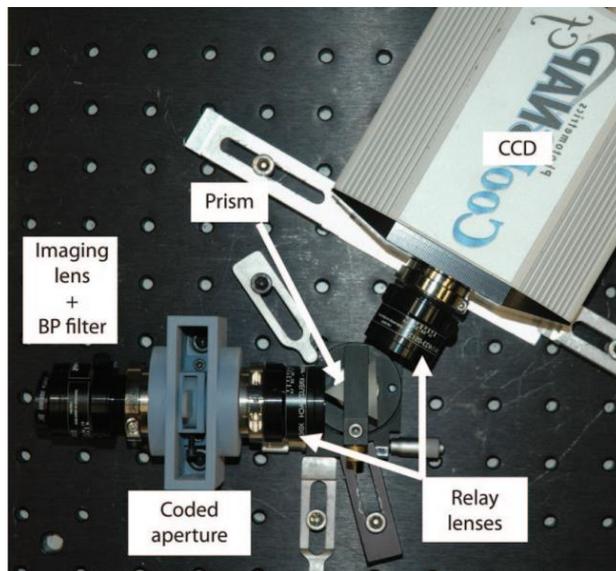
Figure 4: Experimental prototype of the SD (Single Disperser) CASSI [5].

3.3. Image Mapping Spectrometer (IMS)

Image Mapping Spectrometers (IMS) [6] position a micromachined remapping mirror on the image plane of an objective lens (see Figure 5). This mirror is comparable to image slicers used in astronomy and consists of a large number of very small striped facets oriented in N different directions. Each facet redirects one column of pixels to one of N subimages in which the next N-1 columns are empty. Subsequently a prism is used to decompose each pixel spectrally in each column, filling up the remaining N-1 empty columns in each subview with spectral info. Finally, collecting optics ensure the light is refocused again onto the sensor. Drawbacks of this approach include the need for very high accuracy micromachined mirrors with manufacturing defects resulting in crosstalk and reduced light throughput, a limit on the spectral resolution based on the amount of tilt angles (limited by mirror accuracy) and the collecting lens NA (numerical aperture),

the complexity of the complete optical architecture and reconstruction artifacts stitching the columns back together [7].
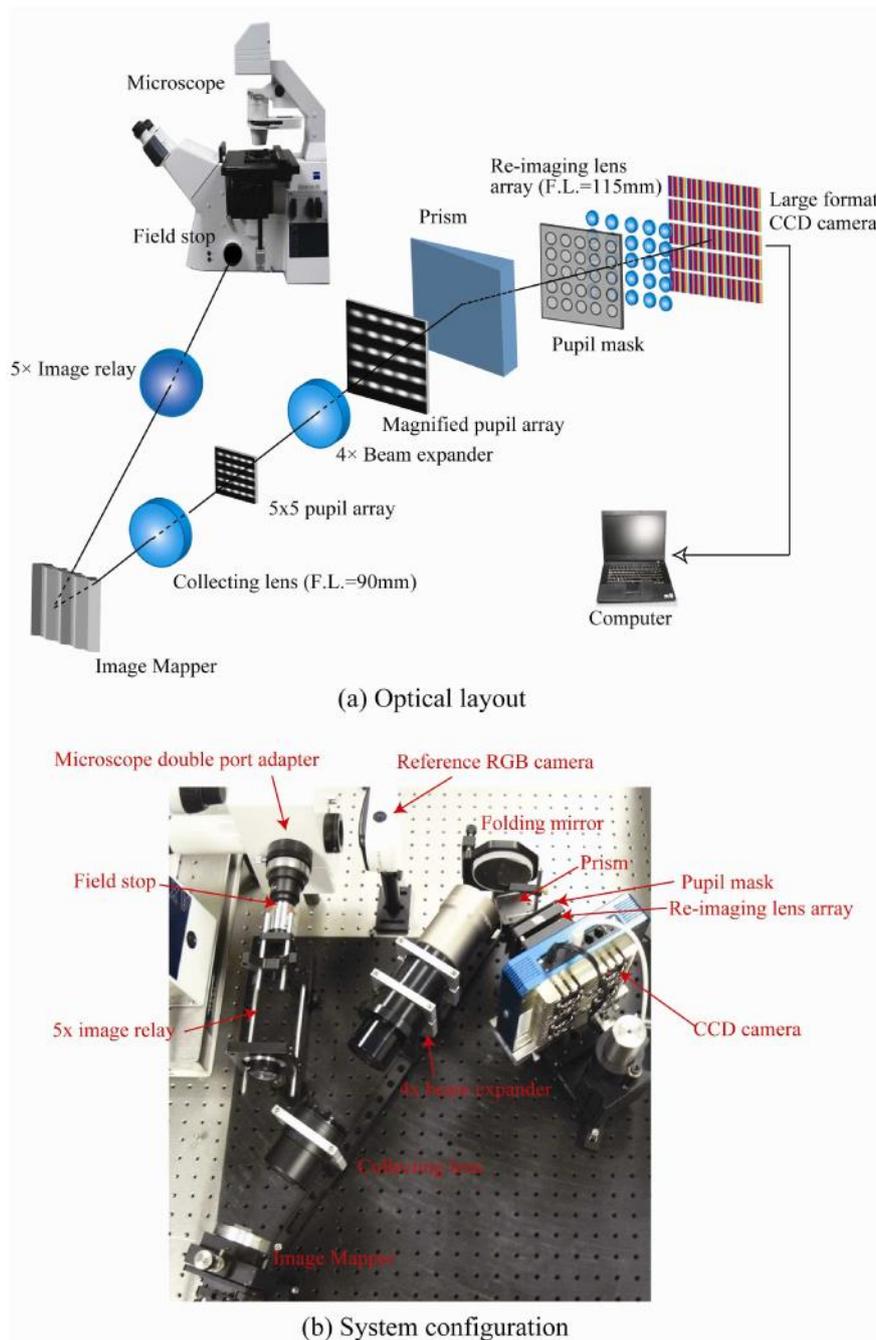
Figure 5: IMS Optical layout and system prototype configuration [6].

3.4. Image-replicating Imaging Spectrometers (IRIS)

Image-replicating Imaging Spectrometers (IRIS) [8] are based on 'Lyot filters' consisting of an assembly of multiple polariser and waveplate combinations, where the consecutive polarisers implement the spectral filtering of the light (see Figure 6). In IRIS the

polarisers are actually 'Wollaston prism polarising beam splitters', which not only filter the light, but also ensure the image duplication. Drawbacks of this system include low light throughput due to polarizing filters and the limited number of bands due to the use of Wollaston prisms to redirect the filtered subimages.
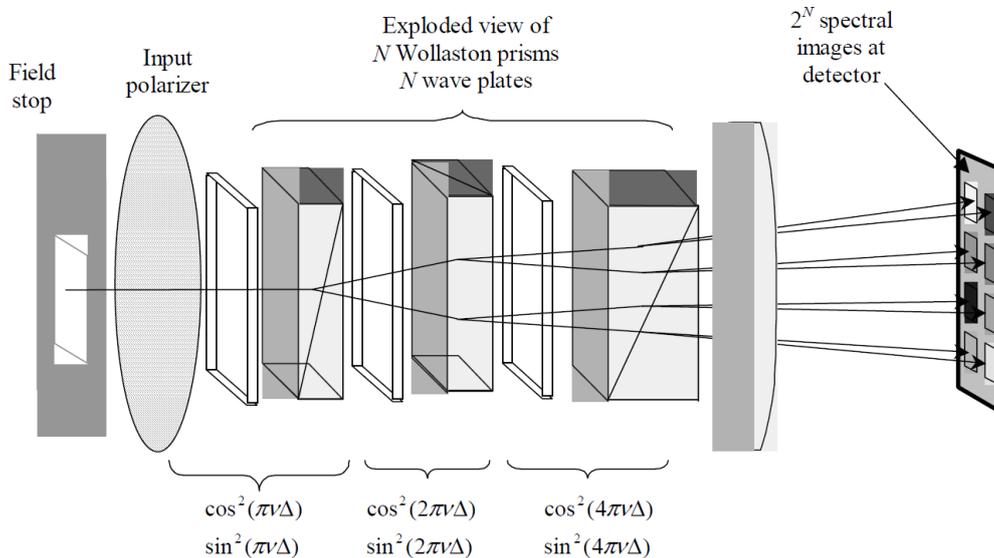
Figure 6: Depiction of principle of operation of Image replication imaging spectrometer (IRIS, [8]).

3.5. Fourier Transform spectroscopes (SHIFT)

Standard Fourier Transform spectroscopes such as the Michelson interferometer operate by sequentially multiplexing different linear combinations of signals using some form of interferometric scanning [9]. The original spectrum may then be extracted at very high spectral resolutions from the resulting signal (interferogram) using inverse Fourier transforms. One of the major benefits of Fourier Transform spectroscopy is the Fellgett advantage, producing a high signal-to-noise ratio in low-light conditions, when measurement noise is dominated by detector readout noise. This is because detector readout noise is independent of the power of incident radiation, whereas the multiplexed signal power rises linearly with the number of multiplexed signals. When operating under a shot noise limited regime, where the noise is proportional to the square root of the incident power, this multiplexing benefit cancels out. Different derived systems have been proposed which forego the need for scanning by simultaneously acquiring different multiplexing combinations, e.g. the 'Multiple-image Fourier Transform Spectrometer' [10], based on tilted mirrors to vary optical path differences and lenslet arrays to duplicate images, and the 'Snapshot Hyperspectral Imaging Fourier Transform spectrometer' or SHIFT [11], based on birefringent prisms and lenslet arrays (see Figure 7). The snapshot Fourier Transform spectrometers are based on very complicated optical architectures and can suffer from reconstruction artifacts due to missalignments.

Drawbacks of all these devices include their complicated optical paths using non-standard and heterogeneously integrated custom components. This leads to a high system cost and typically leads to a loss of optical throughput and alignment and manufacturing problems producing reconstruction artifacts and finally wasted sensor area and smaller spectral cubes.
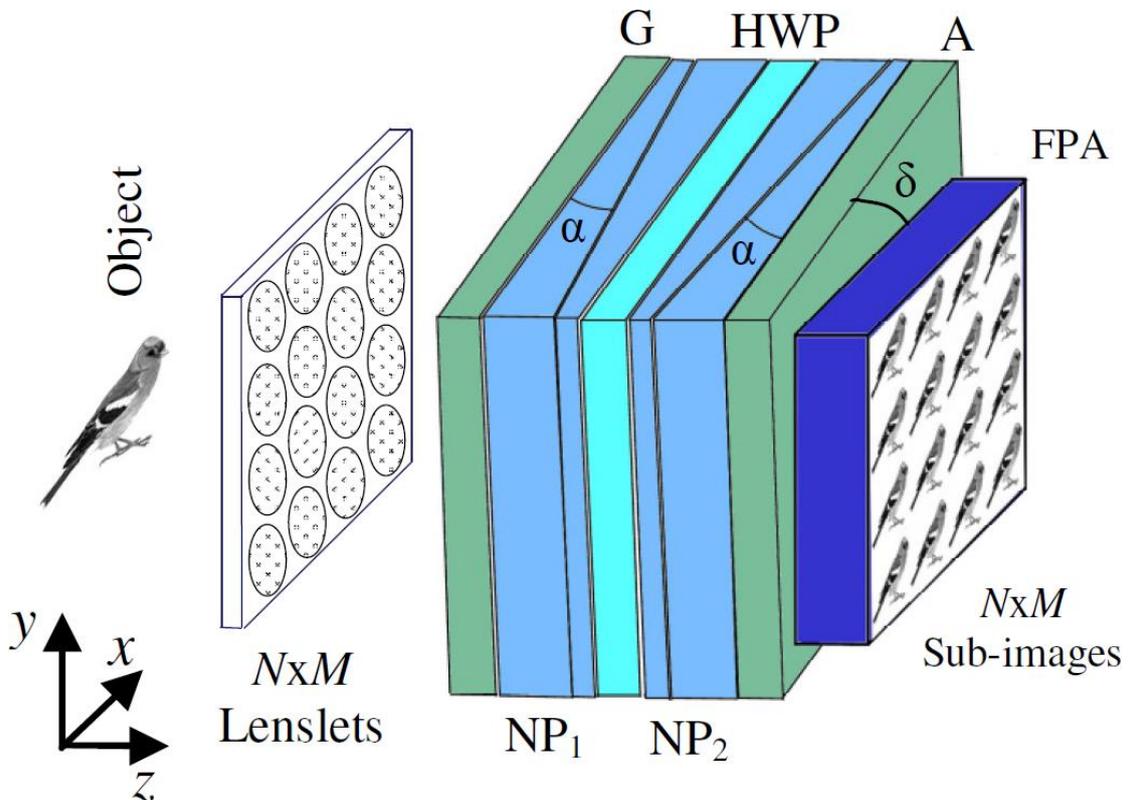
Figure 7: Snapshot Hyperspectral Imaging Fourier Transform Spectrometer or SHIFT, including a lenslet array, G(enerating) and A(nalyzing) Polarizers, two birefringent Nomarski Prisms (NP) and a Half-Wave Plate (HWP) in front pof a focal plane array (FPA) [11]

3.6. IMEC's Multi-Aperture Filtered Camera (MAFC)

The solutions above all use dispersive optics to extract spectral information. Such a dispersive spectral decomposition may be seen as a duplication of the same spatial point over varying wavelengths on the sensor. Imec's snapshot spectral camera uses multiple Fabry-Pérot filters to extract spectral information. The generation of a full spectral cube in one frame period requires duplicating each spatial point onto each of these filters using duplication optics, based upon monolithic lens arrays [16]. The individual lenses in these arrays form different apertures, such that the entire system may be seen as a form of Multi-Aperture Filtered Camera (MAFC). The system is based on spectral filters monolithically integrated directly on top of the imaging sensor, thereby significantly avoiding the optical complexity of systems built around dispersive optics.

D3.1 Analysis & Evaluation of Video SSI Architectures

This optical duplication approach is an extension of the TOMBO (Thin Observation
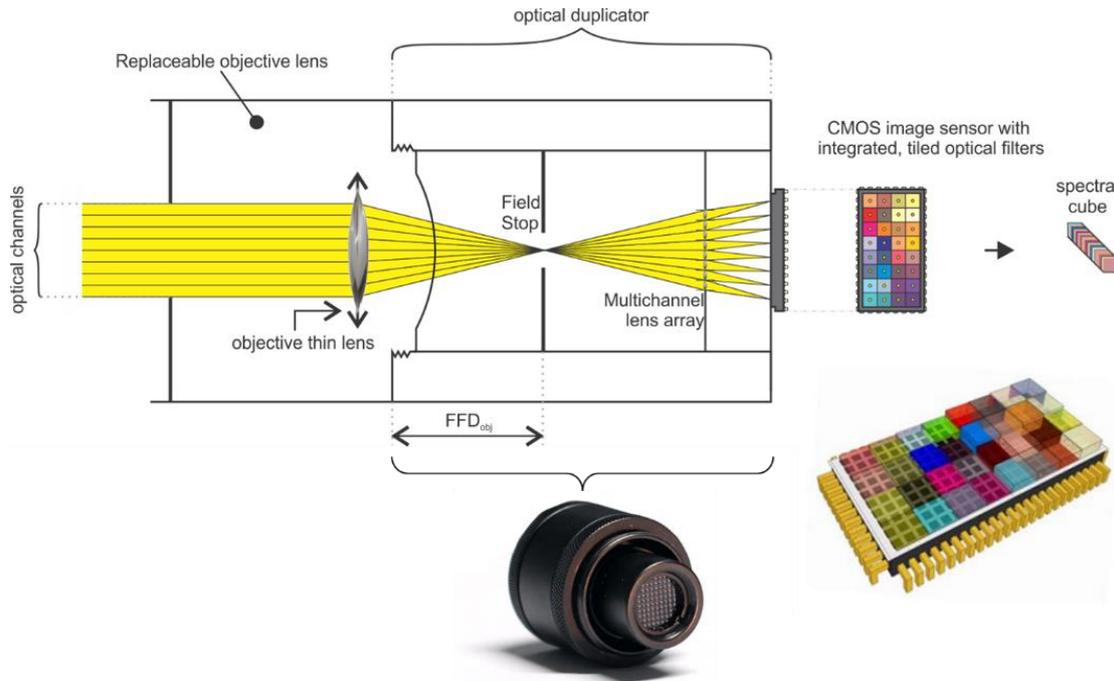
Figure 8: imec snapshot tiled, where scene simultaneously and fully duplicated on each filter tile using optical duplication [16].

Module by Bound Optics) [12] multichannel optics concept, aimed at reducing system thickness. Given the design goal of a small system size, TOMBO does not use an objective lens. Generating a full resolution output image requires the use of super-resolution algorithms, to construct one large image out of many nearly identical smaller tiles. This is due to the lack of an objective lens to regulate the positioning of the scene relative to the lens array.

Multichannel lens arrays are also used as the core around which to build plenoptic or light-field cameras. These cameras enable a richer sampling of the angular characteristics of light and the 4D light-field in order to enable applications such as depth estimation and software refocusing. This suggests that the use of a multichannel lens array is problematic for our snapshot spectral camera, because perspective effects between the different channels could introduce depth-based disparities [13], resulting in a missalignment between the different spectral channels, depending on the 3D geometry of the scene. However, the architecture of plenoptic cameras is fundamentally different compared to that of our optical duplicator to enable light-field sampling: in Stanford (and Lytro)'s plenoptic camera [14] the multichannel lenses are placed on the objective lens image plane, whereas in our optical duplicator they are configured as relay lenses at a relatively high distance behind the objective image plane. In object space the plenoptic multichannel entrance pupils are moved to the object plane. Therefore each channel sees a different part of the scene and the mapping to

optical channels is very depth dependent. As such there is no full scene duplication. Focused plenoptic cameras [15] are hybrids between the basic plenoptic camera and our optical duplicator, but the basic characteristics remain: the multichannel lenses are closer to the objective image plane so that the entrance pupils are closer to the image plane than for our optical duplicator, resulting in optical channels observing different parts of the scene and no full optical duplication.

3.7. IMEC's Spectrally Resolving Detector Arrays (SRDA)

Monolithic integration enables lithographic alignment, improved glare performance, a compact, low-weight and robust system and compatibility with low cost and high volume manufacturing. It also enables a variant of Imec's snapshot spectral camera relying on pixel-level mosaics instead of multi-aperture optics and tiled filters. Compared to our tiled snapshot system [16] the proposed per-pixel system [17] does not require an optical duplicator, further reducing system weight and volume, and avoids potential pupil sharing problems due to non-lambertian light distribution and disparities, in return for a higher spatial Nyquist limit per band and increased deposition alignment requirements. It may be seen as a multispectral extension of the traditional Bayer color filtering concept (see Fig. 3), enabled by lithographic manufacturing of pixel-level interference filters. Compared to these Spectrally Resolving Detector Arrays (SRDA), traditional SSI systems typically rely on complicated optics, increasing system size, weight and cost. Complex reconstruction algorithms are frequently used, which limit response latency.
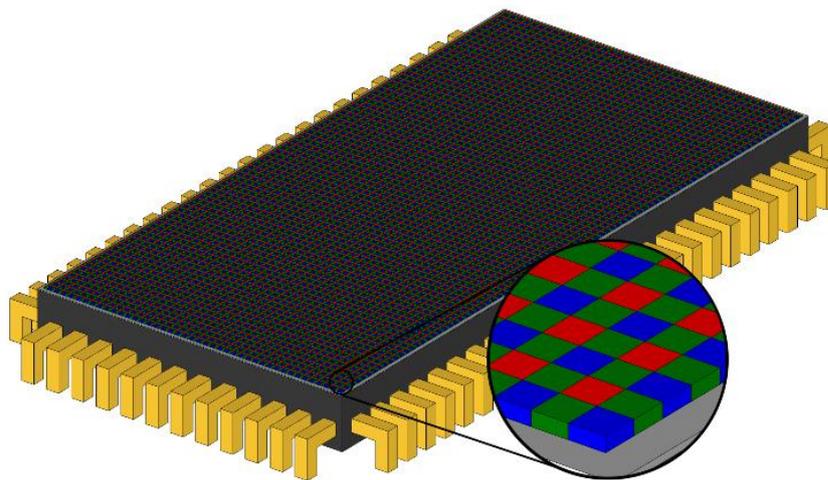
Figure 9: Bayer filter mosaic layout
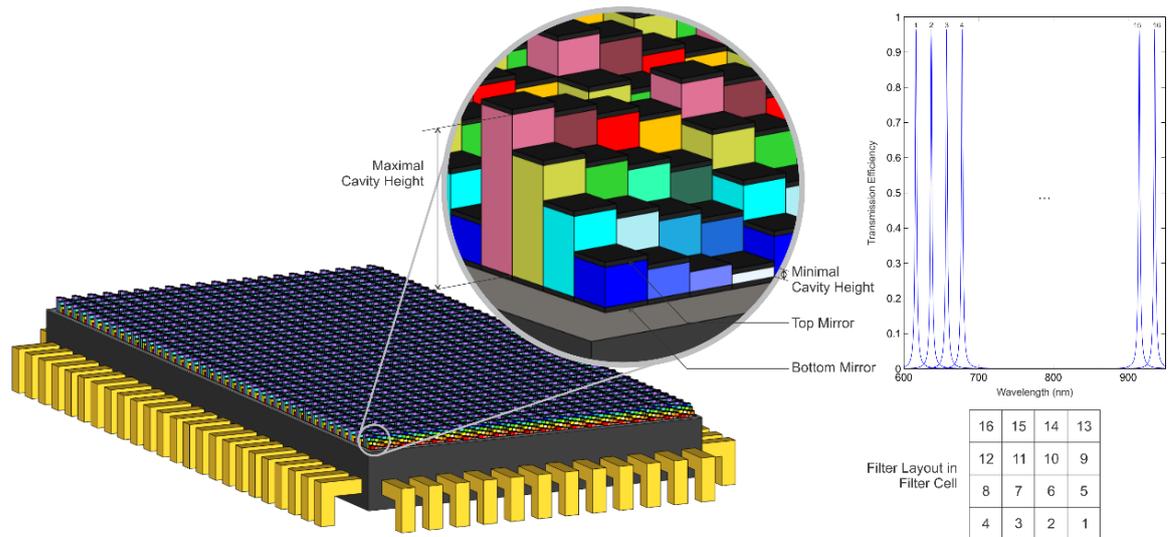
Figure 10: A schematic representation of an integrated implementation of the Fabry-Pérot optical filter on top of standard CMOS imager in a pixel-level mosaic layout (dimensions not to scale, left) and alignment of filters to pixels (right) [17].

4. Review of SSI architectures

A review of the above-mentioned SSI systems and some approaches towards snapshot spectra imaging can be found in [18]. The following section present some important conclusions related to the analysis of these systems.

There are many ways to compare various snapshot implementations, such as compactness, speed, manufacturability, ease of use, light efficiency, and cost. And while these are all important, different system designers have different opinions about each of these factors, so that any discussion can quickly devolve into an argument. In an attempt to avoid explicitly taking sides, this section options to compare the various technologies on a more fundamental level—the efficiency with which they make use of their detector elements. Snapshot spectral imagers generally make use of large detector arrays and can push the limits of existing detector technology, so that their efficiency in using detectors correlates closely with other important issues such as compactness, speed, and cost. This is represented by a metric called the specific information density Q: the product of the optical efficiency η (i.e., average optical transmission times the detector quantum efficiency) with what can be called the detector utilization ζ. The utilization is the number of Nyquist-resolved elements R in the imaging spectrometer datacube divided by the number of voxel samples N required to achieve R. Here $R=R_xR_yR_w$, where R_x, R_y, R_w denote the datacube resolution elements in the x, y, and λ directions. Thus, for a Nyquist-sampled system, the ideal value under this definition is Q = 1. Letting M_u, M_v denote the 2-D detector sampling elements, we have

$$Q = \eta \frac{N_x N_y N_w}{M_u M_v}$$

for optical efficiency η. In order to show that the value for Q among technologies stems from more fundamental considerations than system design parameters such as the aperture diameter and system magnification, we assume ideal conditions for each instrument type and derive the relevant efficiency from the required margins at the focal plane needed to prevent significant crosstalk among elements of the datacube. Here crosstalk is defined as the condition where multiple voxels within the measured datacube each collect a significant amount of signal from the same voxel in the true object datacube and where these two voxels are not physically adjacent to one another in the datacube. For voxels satisfying this condition but that are physically adjacent, we can call the effect blur rather than crosstalk.

For optical efficiency estimates η for each technology, we assume ideal components, so that lenses, mirrors, prisms, and gratings are assumed to have no losses (100% transmission or reflectivity), and that all detectors have an external quantum efficiency of 1.

One of the reasons why we choose the *detector utilization ζ* to define a metric for comparing technologies is that it is in many ways a proxy for other important measures such as manufacturability and system size. The connection arises because, in various ways, all of the snapshot techniques encode the spectral information by expanding the system étendue. If all things are held constant except for the wavelength- dimension of the cube, then, in every instance, increasing N_w requires increasing étendue. And this quickly runs into difficult design constraints—for high-performance more expensive optics (i.e., larger-diameter optical elements that can also handle a wide range of angles). Thus, snapshot systems with lower ζ will generally reach this design ceiling before the higher ζ systems will, and either system size or the angular acceptance of the optics must compensate for the difference in ζ.

The basic premise from which we derive the detector utilization ζ for each technology is that each technique requires a margin around each subsection of the datacube, without which blurring will cause significant crosstalk. For some technologies, smaller margins are easier to achieve than for others, but this factor is ignored here. Those technologies that minimize the number of marginal pixels make the most efficient use (have the highest utilization ζ) of a given detector array, but the actual value of ζ depends on the aspect ratios of the datacube dimensions. For example, from Figure 11 we can see that the MAFC, IRIS and SHIFT technologies use a similar format of projecting elements of the datacube onto a 2-D detector array:
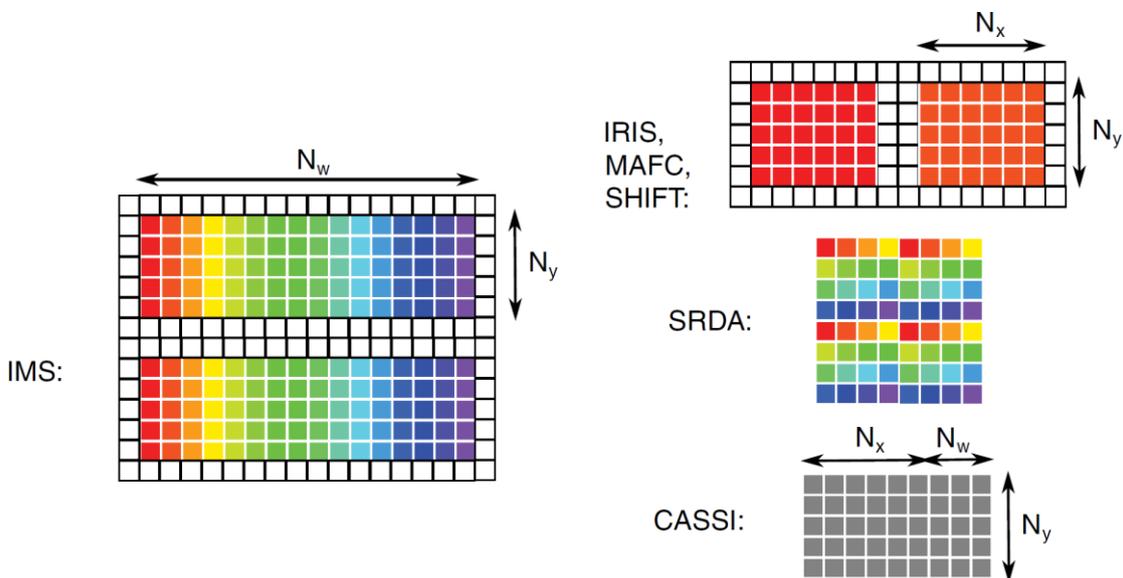
Figure 11: Diagrams showing how the detector utilization formulas are calculated for each architecture, given the basic layout of how the datacube projected onto the two-dimensional detector array. Each square shown here represents a single pixel on the detector array [18]..

For the IRIS, SHIFT, and MAFC technologies, each single-channel slice of the datacube is measured as a contiguous region, so that each wavelength requires a rectangular space of $(N_x + 2s)(N_y + 2s)$, and the total number of pixels needed is

$M_{IRIS} = (N_x + 2s)(N_y + 2s)N_w$.

The value for $\zeta \equiv N/M$ follows directly from this equation as

$$\zeta_{\text{IFS-F}} = \frac{N_x N_y N_w}{(N_x + 2s)(N_y + 2s)N_w} = \frac{N_x N_y}{(N_x + 2s)(N_y + 2s)}.$$

For the IMS technology, an $N_y \times N_w$ swath is measured in a contiguous region on the detector array, so that each swath requires a rectangular space of $(N_w + 2s)(N_y + 2s)$. Multiplying by the total number of x-resolution elements in the datacube gives

$$M_{\text{IFS-M}} = N_x(N_y + 2s)(N_w + 2s).$$

For the filter-array implementation of SRDA, each pixel samples an individual voxel, so that the utilization is inherently equal to 1.

In the case of CASSI, we find that $M_{CASSI} = (N_x + N_w - 1)N_y < N$—that is, the utilization is >1. In fact, the greater the number of wavelengths in the datacube, the greater the utilization for CASSI. Note that, due to problems achieving the micron-scale imaging required to map code elements 1:1 to detector pixel elements, existing CASSI instruments map code elements 1:2, so that they use about four times as many detector pixels as the theoretical value given here, i.e., $M_{CASSI}^{(practical)} = 4(N_x + N_w - 1)N_y$.

Table 1 summarizes the η and M values used to calculate Q for each technology. In the table, note that for the computational sensors (CTIS and CASSI), the number of datacube voxels is related to the number of resolution elements N not through the Nyquist sampling limit but through more complex criteria. When calibrating these computational sensors, M is technically an arbitrary value, but in practice one finds little value in allowing M to exceed the values shown in the table. In addition, for the SRDA row in

Table 1 it is assumed that the implementation uses the filter-array camera. From

Table 1 we can see that the MAFC technologies offer the highest Q for high spatial/low spectral resolution datacubes (squat-shaped cubes), whereas the IMS options offer the highest Q for low spatial/high spectral resolution datacubes (tall cubes). The latter do especially well when the spatial dimensions of the datacube are rectangular Nx ≠ Ny, because the image is resliced along (extremely rectangular) lines. Both have a good detector utilization under those conditions, combined with an optical efficiency η of 1. As indicated in

Table 1 the IRIS approach behaves exactly as the MAFC technologies, but loses a factor of two due to its need to work with polarized input. Assuming well-aligned, pixel-level

filters, the SRDA combines an optical efficiency η of 1with a sensor utilization of 1, and thus also offers the highest Q.

Each of the technologies listed in

Table 1 is also classified according to the method used to divide the light into voxel elements. The majority of technologies use division of field [F] (also called division of focal plane), in which the light is either filtered or divided into separate beams according to its placement within the image. Division of amplitude [A] is the next most common method in which the light is divided into separate beams by allocating a portion of light into each beam, as a simple cube beamsplitter does. Only two other methods exist: division of pupil [P] (also called division of aperture) and compressive sensing [X].

Technology	Class	η	M (pixels used)
CTIS	A	1/3	$\sim N$
CASSI	X	½	$N_y(N_x+N_w-1)$
IMS	F	1	$N_x(N_y+2s)(N_w+2s)$
IRIS	A	1/2	$(N_x+2s)(N_y+2s)N_w$
SHIFT	P	¼	$(N_x+2s)(N_y+2s)N_w$
MAFC	P	1	$(N_x+2s)(N_y+2s)N_w$
SRDA	F	1	$N_xN_yN_w$

Table 1: The Classification type of each technology, and ideal values for the optical efficiency η and the number of detector pixels used.
Finally, it should be emphasized that the analysis summarized in

Table 1 provides a theoretical benchmarking of these different classes of SSI cameras, but not of their concrete instantiations or implementations: the analysis assumes the cameras are constructed using ideal optics, cameras, SLM's etc. but in reality these component choices will be trade-offs not just between price and 'high-level performance', but also between metrics such as volume, weight, power-consumption, reliability, and detailed performance metrics such as SNR, MTF, spectral and temporal resolutions, distortion, ... In practice, the component selection trade-offs will often be simpler for devices based on simple system architectures. Actual devices could also be additionally benchmarked according to a selection of these different quality criteria. Finally, as mentioned above, detector utilization ζ represents how efficiently the different classes of SSI cameras manage to multiplex a 3D cube of spectral data on to a 2D imaging sensor while avoiding crosstalk. Theoretically though, computational

imaging devices need not avoid crosstalk, but might actually exploit it while still being able to decode the original input cube, thus realizing detector utilizations > 1. It is exactly the goal of the PHYSIS project to develop alternatives to CTIS and CASSI which are able to push the detector utilization beyond 1, while simultaneously keeping an eye on other quality metrics such as SNR.

5. Hyperspectral video

5.1. Snapshot mosaic video

Hyperspectral video sequences with high temporal resolution combine the advantages of both video and hyperspectral imagery. Despite the high spatial resolution of modern RGB or monochromatic video cameras, their low spectral resolution limits their ability to discriminate, classify or identify objects robustly. In contrast, multispectral and hyperspectral sensors offer comprehensive datasets with enhanced discrimination capabilities, useful for the characterization of subtle spectral features and important chemical and physical properties of the observed terrain features/objects. Combining the spectral information with the temporal information offered by video rate imaging will therefore enable novel analysis techniques in a wide spread of applications: have already been employed for several applications in medical imaging [20], precision agriculture [21], gas plume detection [22] and moving object detection and tracking [23][24].

In snapshot spectral acquisition, an entire multispectral data cube is sensed at one discrete point in time. Such snapshot acquisition requires mapping each point in a WxHxNB-sized spectral cube to a pixel on the sensor (see Figure 12). Whereas in scanning spectral imaging the data in a spectral cube is multiplexed over multiple, consecutive frames, **in snapshot spectral imaging this data is multiplexed on to a single frame of the sensor**. Due to the limited pixel real estate on the sensor, there will be a trade-off between spatial and spectral resolution. This again highlights the need for a good sensor utilization, as was covered in the previous section and it confirms the potential of the MAFC systems for high spatial resolution, IMS for high spectral resolution, and SRDA for high spatial or spectral resolutions.
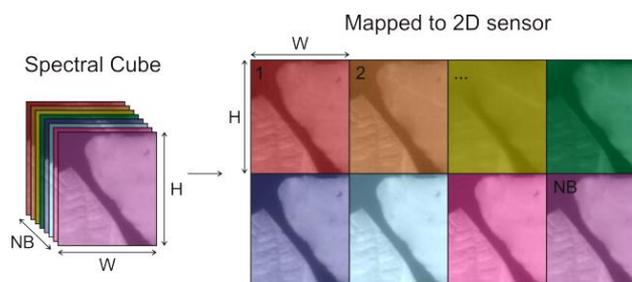
Figure 12: Spectral cube with NB bands and spatial resolution of WxH mapped to NB tiles on 2D sensor.

D3.1 Analysis & Evaluation of Video SSI Architectures

In addition to this, SSI systems based on monolithically integrated filters, i.e. imec's tiled MAFC system and imec's mosaic SRDA system offer a lot of flexibility, both in terms of optics selection and optical throughput, and in terms of high-speed camera selection. Unlike systems based on dispersive optics, such as IMS, for these systems the spectral unit is in effect integrated on the sensor, such that no additional collimating or focusing optics are needed. Combining this with the freedom to select a miniature camera (Ximea xiQ, Figure 13), permits the construction of a portable spectral camera, operated via a laptop USB3 connection and easily deployable in the field (Figure 14).
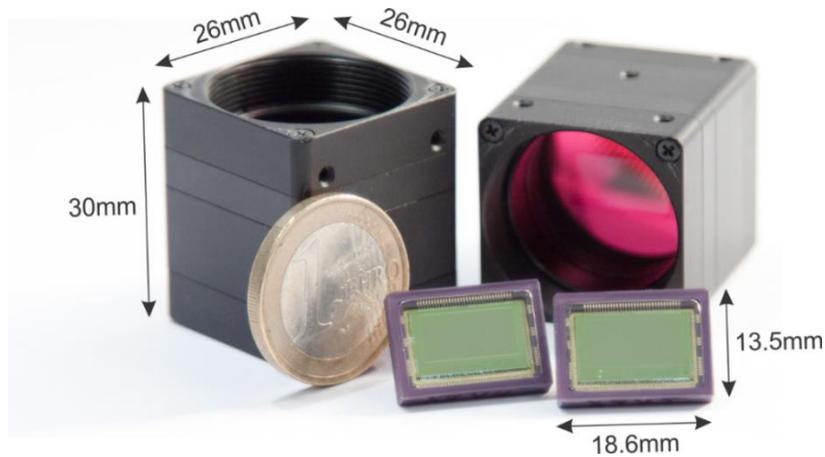
Figure 13: Two packaged mosaic spectral snapshot sensors, alongside two tiny spectral cameras and a euro coin for scale.

Figure 14: The tiny size of the xiQ camera enables flexible outdoor spectral data acquisition.

A demonstration of the potential of SSI video for security applications is given in [19], where the spectral and temporal information is used to do multiple objects detection and tracking. The goal was to design a generic approach able to detect moving objects

like people and crowds in interior and exterior public areas for any security purposes like surveillance or evacuation. The results are illustrated in Figure 15, showing multiple frames of an SSI sequence representing activity on a parking lot, with passing cars and pedestrians, which have successfully been identified and tracked.
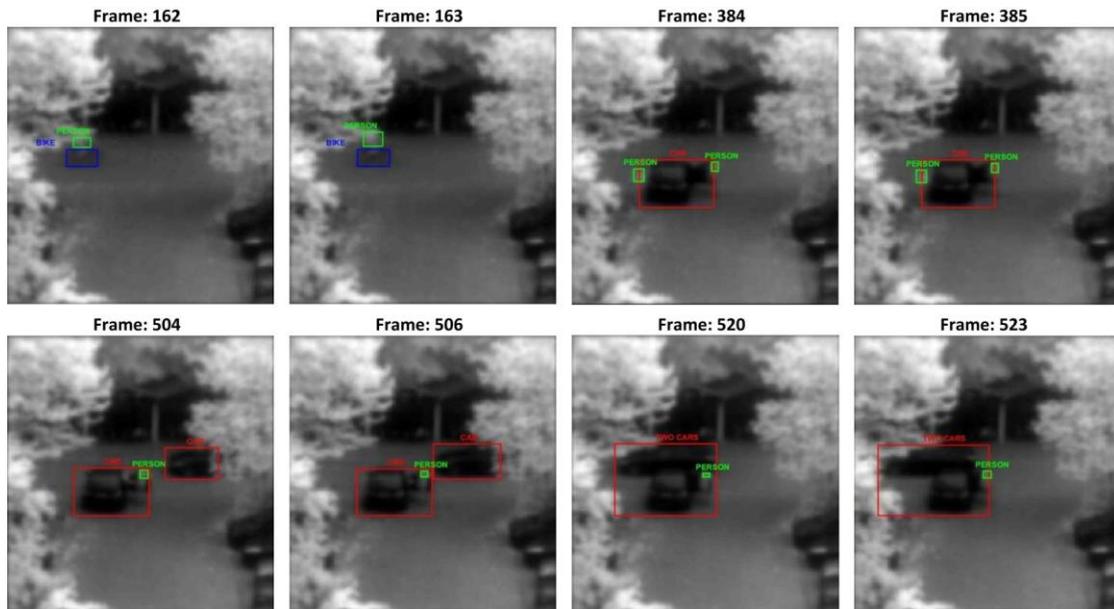
Figure 15: SSI video enabled multiple object detection and tracking [19]

Although Hyperspectral video has great potential in multiple applications, it also introduces extremely demanding data processing and storage rates. More specifically, considering the IMEC Mosaic sensor introduced into the Ximea camera system, the system must be able to process 2Mpixel x 340 measurements/sec = 680 Mpixel/sec, which translates to 680 Mpixel/sec x 8 bits/sample = 5440 Mbits/sec or *680 Mbytes/sec*, assuming data in the mosaic format. If one considers that these data must be super-resolved to full hypercubes, this translates to *17 Gbytes/sec* for the 5x5 binning mode. Even if one reduces the sampling rate by a factor of 10 to 34 measurements/sec, the full output is still above 1 Gbyte/sec, imposing extremely hard constraints on the processing platform.

5.2. Accessing the demonstrator data set

In order to get acquainted with the characteristics of spectral data, imec has provided some example data to the consortium, acquired using different spectral cameras. Some of this data represents raw, unprocessed measurements directly from the camera, whereas other data has undergone a number of different corrections. These low-level corrections will be described in "D3.3 - Report on workflow for implementation of extension of prototypical HSI to improve resolution". The raw frames are stored in a

binary PGM format [25], which may be opened using Irfanview or simply directly loaded in Matlab.

The raw frames represent the 2D frame data directly acquired by our spectral cameras. After low-level processing and reorganization of data according to their spatial and spectral content, the data is reorganized into a 3D spectral cube, with two spatial and one spectral dimension. The provided cubes are stored in an ENVI compatible format. The file format we use is BIL, and each voxel (pixel in 3D hypercube space) is stored as 2 bytes unsigned integers. The cube data itself is stored in a '.raw' file, accompanied by a '.hdr' header file containing the metadata on cube structure and contents. The cube data may be loaded via dedicated software such as ENVI itself [26]. There is also a Matlab command to read in such a cube ('multibandread'). A tutorial on how to do this (including visualization) may be found at [27]. As an example, one of the provided cubes may be loaded as:

```
c=multibandread('.\meat_nir_sc_FD1_FN0.raw',[    216,    409,
25],'uint16',0,'bil','ieee-le');
```

Note that the number of lines and rows (216, 409) does not coincide for every cube, you can find this info on the .hdr files. 25 is the number of bands in this case.

Note also that with the multibandread command you can choose to load only a subset of the rows, columns and bands in the following way:

```
 c=multibandread('.\meat_nir_sc_FD1_FN0.raw',[    216,    409,
25],'uint16',0,'bil','ieee-le',{'Band','Range',[2 10]});
```

All information of the raw data file is stored in the header file. More information on the header file can be found in [28].

5.3. Simulating spaceborne imagery

While the acquired data discussed in the previous subsection will be instrumental in evaluating the performance of the project's various processing modules such as enhancement, compression, and restoration, the data is not sufficient for simulating measurements acquired by a spaceborne platform. To that end, we have considered possible data inputs from the CHRIS sensor and the Hyperion instruments aboard the ESA controlled PROBA-1 and the USGS controlled EO-1 satellites, respectively. We considered there two instruments due to the high spatial and spectral resolution of the acquired data and the similarity between the orbit of these satellites and the low Earth

orbit envisaged for the simulation of the PHySIS platform data. Detailed information is given in Table 2.

Table 2: Characteristics of available HSI sensor and data

Parameter	CHRIS	Hyperion
Spectral bands	19 spectral bands (fully programmable) in the VNIR range (400 - 1050 nm)	220 spectral bands (from 0.4 to 2.5 µm)
Spatial coverage	13x13 km	7.5 km by 100 km
GSD	17 m	30 m
Orbit	550-670 km	690-700 km

Data from these dataset will be modified in order to serve as simulation data for the signal processing and analysis models that will be developed in PHySIS. The detailed description of these data will be included in D3.2 "Report on the evaluation of the prototype HSI system with 3 IMEC HSI sensors".

Furthermore, real HSI video data will be collected by the IMEC and FORTH teams by placing the camera system on the roof of the FORTH building in Crete and imaging the nearby landscape. The maximum distance with line-of-sight that is easily accessible can accommodate measurements up in the order of 70km. These acquired measurements will be reduced in dimension in order to account for much larger distances of a low Earth orbiting satellite, while additional controlled degradation will be introduced in order to reach the TRL4 data required by PHySIS.

6. Summary

In this deliverable, we have provided an overview of evaluation of various HSI technologies, focusing especially on technologies capable of generating hyperspectral video, i.e., Snapshot Spectral Imaging (SSI). Strengths and weakness of each of the approaches have been highlighted, and a quantified comparison based on detector utilization ζ and specific information density Q has been given for each of the SSI technologies, assuming ideal optical components. Detector utilization is a suitable metric for comparing technologies because it is in many ways a proxy for other important measures such as manufacturability and system size.

Based on this analysis, MAFC technologies offer the highest Q for high spatial/low spectral resolution datacubes (squat-shaped cubes), whereas the IMS options offer the highest Q for low spatial/high spectral resolution datacubes (tall cubes). The latter do especially well when the spatial dimensions of the datacube are rectangular $N_x \neq N_y$. Both have a good detector utilization under those conditions, combined with an optical efficiency η of 1. Assuming well-aligned, pixel-level filters, the SRDA combines an optical efficiency η of 1 with a sensor utilization of 1, and thus also offers the highest Q.