White Paper: Dimension, Inc. Technology Overview

Introduction

Dimension Inc is well known for development of advanced image processing technologies pertinent to technically demanding video applications such as IoT, Surveillance, Analytics, Video Conferencing, VR/AR, and Streaming Video Platforms. In particular, Dimension has focused upon image and video *reconstruction*, *denoise*, and *compression* (CODEC) as those processing components most critical to video system performance optimization. From a technical perspective, this work has been successful in creation of algorithms and systems of extraordinary power and sophistication. Most notably, combination of *superresolution reconstruction*, *spatiotemporal denoise*, *Nyquist refinement*, and *superresolution-based video compression* have together enabled processing gains at levels heretofore unavailable in the video technology marketplace.

An improved image and video performance capability notwithstanding, what isn't so apparent is Dimension's flexible instancing of technology block-components, performed in such manner development of more complex systems is greatly facilitated. More pointedly, fundamental processing blocks are integrated hierarchically in assembly of more complex systems. This simple 'LEGO-esque' gambit is not as trivially simple as some might imagine, and in fact must be rooted in a comprehensive strategic plan for which efficient system integration accrues as an explicit objective. In Dimension's case, software reuse is critical to conservation of development resources. This has in turn led to design of the aforementioned video applications as integrated products for which a significantly lowered life-cycle cost (LCC) is also indicated.

Accordingly, in what follows, Dimension technology is presented as a hierarchy of building-block resources, ranging from those most fundamental in terms of range of applicability, to those most complex in terms of block integration and functional specialization. This overview then terminates with presentation of complete video applications incorporating Dimension technology as system components. As described above, Dimension technology components are generally categorized as *basic* or *non-basic* with non-basic components hierarchically integrating basic components within some higher level construct:

Component	Basic	Notes
ASVD	✓	Adaptive Spatio-Temporal Video Denoise
PMA	✓	Pattern Manifold Assembly
SREC		Superresolution Enabled CODEC
SREU		Superresolution Enabled Upscaler
SRED		Superresolution Enabled Downscaler
SRNR		Superresolution Nyquist Refinement
SRRS		Superresolution Video Rescaler
VPC		Video Preconditioner
VTT	✓	Video to Text

Table-1: Dimension Technology Components

Hierarchical relationships among Dimension technology components displayed in table-1 are summarized in table-2 below:

Component	ASVD	PMA	SREC	SREU	SRED	SRNR	SRRS	VPC
SREC							✓	
SREU		✓						
SRED		✓						
SRNR		✓						
SRRS				✓	✓			✓
VPC	✓[1]					✓		
								[1]Optional

Table-2: Dimension Technology Component Hierarchical Integration

The technology components cited in tables-1,2 are then flexibly instanced as design building-blocks within a number of Dimension applications and productized APIs:

Component	SREC	SRRS	VTT	Notes
BDA[1]	✓	✓	✓	BigData Assembly
BDE[1]	✓	✓	✓	BigData Exploration
BDD[1]	✓	✓	✓	BigData Discovery
MFVA		✓	✓	Multi-Frame Video Analytics
SRAU		✓		Superresolution Archiver Utility
SRRU		✓		Superresolution Restoration Utility
				[1]BD-IDE Application Component

Table-3: Dimension Application/API Instancing of Technology Components

In what follows, a short description is provided for each cited technology component and application.

ASVD

In most general terms, the spatiotemporal formalism is favored in video noise filtering applications due to an inherent encapsulation and averaging of noise sample ensembles, enabling both a direct noise floor reduction and improved estimation accuracy. In recent work, the spatiotemporal estimation concept has been extended to noise filtering of OCT diagnostic image cubes in which a characteristic multiplicative 'speckle' noise process predominates. MMSE optimal noise transfer is obtained via a multidimensional Weiner filter formulation whereby multiplicative noise is approximated as a non-stationary additive process exhibiting significant noise/signal cross-correlation. In benchmark tests, the Spatiotemporal Video Denoise (SVD) filter structure exhibits an effective noise attenuation with minimal loss of image sharpness or detail. In subsequent experimentation this result was extended to generic content in which both additive and multiplicative noise sources are present.

We then focus upon the noise sampling component, whereby calculation of a given filter transfer function hinges upon accuracy of component PSD estimator terms. It then follows, a reduced filter performance may be expected where PSD estimators do not accurately reflect actual noise process statistics. In particular, noise samples must not be corrupted by residual signal information in generation of statistical bias. This requirement forms the basic selection criterion for a candidate noise sample space; to the extent PSD estimators exhibit bias, a suboptimal noise transfer may accrue.

As originally conceived, SVD noise estimation hinges upon prespecification of contiguous image sample regions, noise-like and persistent over successive frames. However, despite achieving an effective filter realization, it became apparent sample regions meeting these requirements are not generally available for all content of interest. Further, the prescan and analysis steps implicit to prespecification are sequential/blocking on an ASVD process schedule and thus incompatible with Amdahl acceleration as a concurrent streaming process. We address these problems via a schema in which per-pixel noise sample ensembles are extracted directly from spatiotemporal estimation buffers. In this manner, image sample regions are adaptively rendered as complete frames obviating any requirement for special content structure. This fully *adaptive* SVD (ASVD) is thus rendered as a parallelizable streaming process on arbitrary signal-plus-noise mixtures.

As displayed in figure-2 below, ASVD is optionally incorporated as a hierarchical building-block component within VPC, whereby VPC is itself a hierarchical component of SRRS.

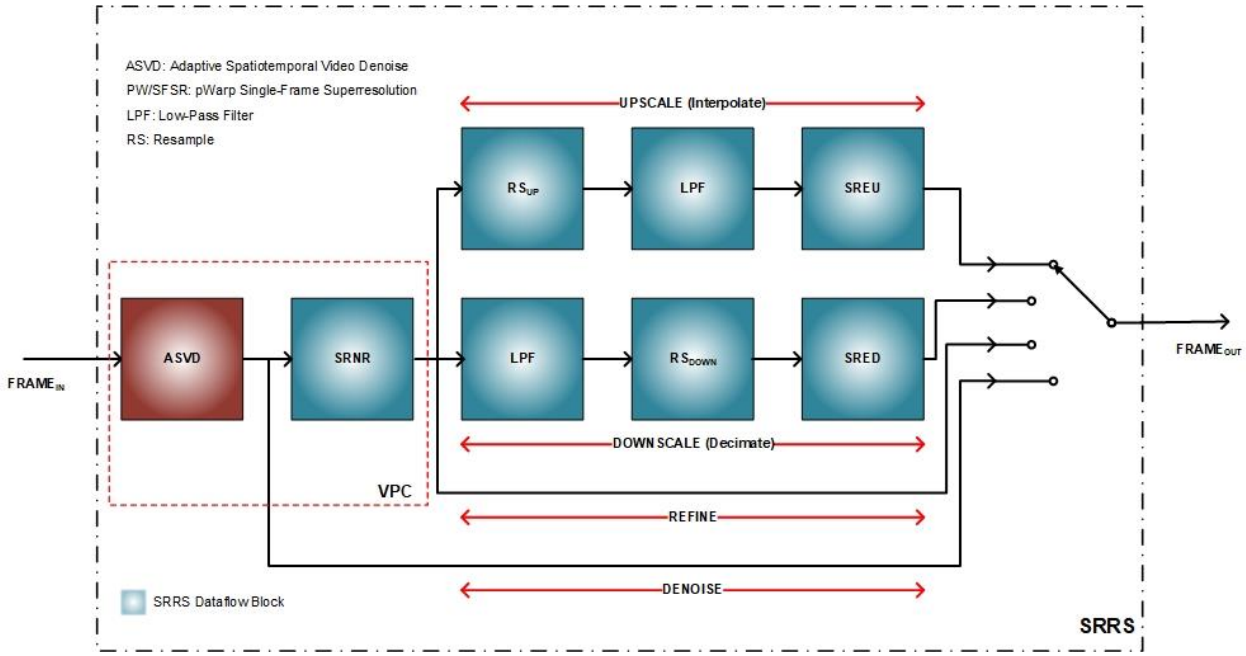

Figure-1: SRRS Architectural Form Incorporating VPC/ASVD

PMA

Photometric Warp (PWRP) superresolution is defined as an image processing technique in which luminance of any image pixel is rendered dependent upon an auxiliary function of some form. Accordingly, PWRP superresolution is obtained where the referenced auxiliary function is also a reconstruction operator. Further, where the reconstruction operator domain is a single frame, we refer to the resulting superresolution construct as single-frame superresolution (SFSR).

As displayed in figure-2, the particular PWRP/SFSR used in this work employs an edge-luminance model both symmetric and scale invariant. In this instance, model symmetry is motivated by an assumed symmetry associated with the fundamental edge-blurring process. Scale-invariance is then assured via a minimal pixel support to which a set of gradient operators are applied for sub-pixel accurate edge-structure detection. In essence, luminance of a detected edge is modified so as to exactly match the luminance model. As an example, support for the edge prototype displayed in figure-1 is reduced from 4-pixels pre-Warp to 2-pixels post-Warp in generation of a 300% increased slope (sharpness) post-superresolution. In what follows, PWRP/SFSR extends this edge-reconstruction model to arbitrary curvilinear structures via directional correlation along edge-detection contours.

As described, PWRP critically depends upon localization of curvilinear edge structures at subpixel resolution. For this purpose, we employ a tailored nonlinear filter that provides this localization at response surface extrema. Once the locus and orientation of an edge-structure at all constituent points is determined, a set of reconstruction filters is instanced at all associated image-space coordinates. The logical sum of all such instances then constitute *pattern manifold assembly* (PMA) as a necessary component of PWRP/SFSR and hence is also functionally implicit to each of SREU, SRED, and SRNR processing blocks.

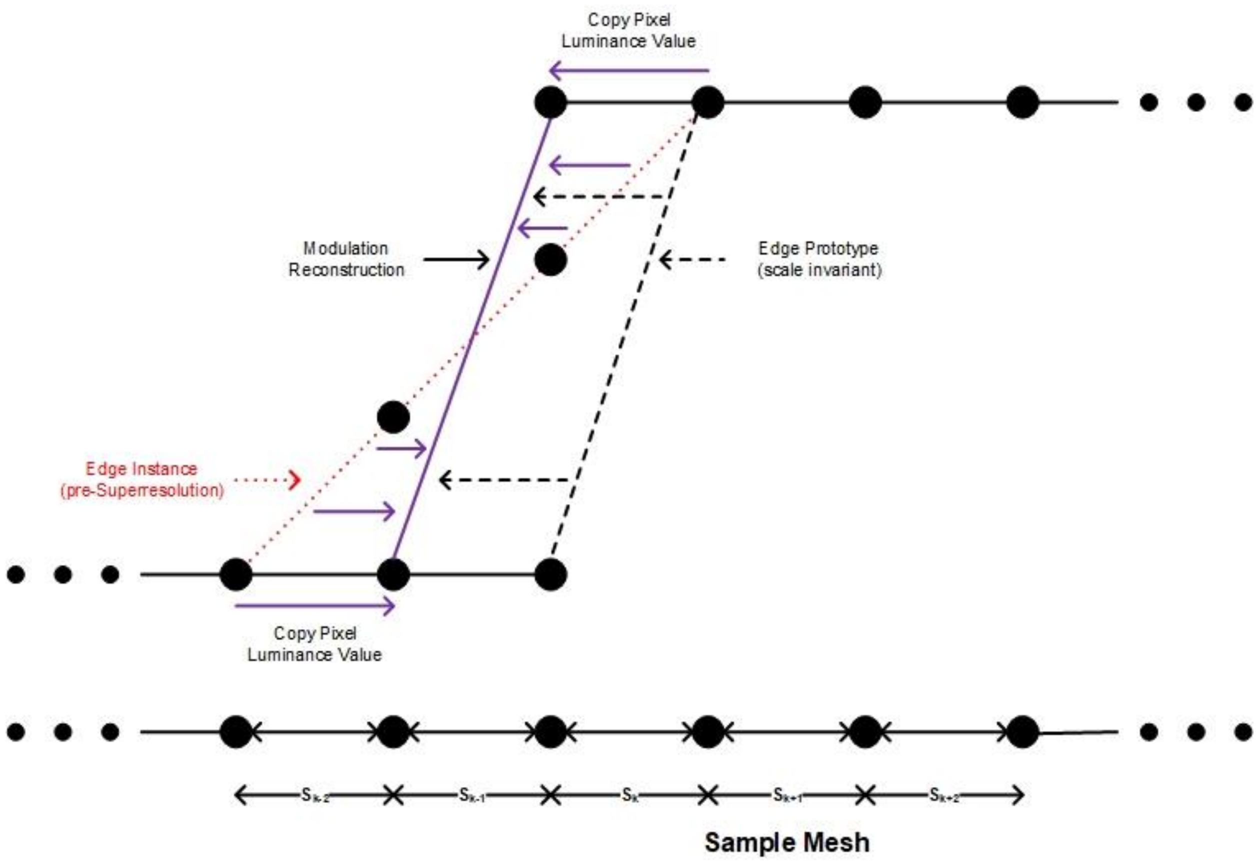

Figure-2: Exemplary Photometric Warp Edge Reconstruction

In table-2 above, PMA accrues as a building-block component of SRED, SREU, and SRNR blocks, in turn instanced as hierarchical building-block components of SRRS, (re: figure-2).

SREU/SRED/SRNR

Each of the superresolution blocks SREU (upscaler), SRED (downscaler), and SRNR (Nyquist refinement) employ photometric warp (PWRP) for purposes of sub-pixel localization of points along curvilinear (1D/line-like) structures appearing within a frame. Resampling and anti-alias filtering are then added so as to render any indicated rescaling process formally valid (re: SRRS).

Here we note SRNR as something of an exception in that sampling rate remains invariant. Thus, the associated Nyquist limit remains the same and no resampling is performed, (i.e., with no attendant requirement for anti-alias filtering). It is then significant the sole purpose of SRNR is image refinement via what is essentially a sharpening process that renders a given frame at maximum resolution, subject to bandwidth limits imposed by Nyquist sampling theory. The major advantage of this particular sharpening process is we avoid noise preemphasis as might result from use of a linear high-pass filter (HPF). Furthermore, inasmuch as each of SREU, SRED, and SRNR are themselves linear, there is no noise upconversion as might result from application of a nonlinear filter of one form or another.

For purposes of software reuse, architectural simplification, and highest possible processing efficiency, SREU, SRED, and SRNR are encapsulated within the SRRS construct. As displayed in tables-2,3, SRRS is instanced as a component block in SREC, BDA,BDE, BDD, MFVA, SRAU, and SRRU. Here, we note not all SRRS internal processing pathways are required in each case. Accordingly, in interests of processing resource conservation, unused SRRS pathways are eliminated via conditional compilation according to an associated design specification. In this manner, any one of SRRS/SREU, SRRS/SRED, or SRRS/SRNR many be isolated according to need. For example, where we consider SRRS as a block-component of the superresolution enabled video CODEC (SREC) construct displayed in figure-2, the optional ASVD is deleted for both ENCODE and DECODE, SREU is deleted for ENCODE, and SRED is deleted for DECODE.

SREC

A number of lossy video CODECs such as MPEG-4 have found widespread acceptance as a video compression technology solution in today's marketplace. In principle, a maximum 200:1 compression ratio is possible with this CODEC. However, the fact MPEG-4 employs a lossy, block-based differential encoding scheme limits useful compression to a more moderate 20:1-40:1 range. From an information theoretic perspective, this reduction may be regarded as manifestation of the fact, at the scale of a given blocksize, only so much redundancy is present in any given image sequence. One consequence of this is overly aggressive redundancy-based encoding tends to create excessive artifacts and noise in an output image.

While block-based differential video compression has proven very successful, the inherent performance limitations of this approach also serves to hinder evolution of streaming video and video surveillance system applications for which an increased level of compression performance is critical. In the superresolution-enabled video codec (SREC) considered here, we adopt a complementary approach whereby the previously discussed photometric warp (PWRP) superresolution is applied to realization of an increased total compression ratio capability. Use of PWRP in this context is motivated by the fact reconstruction filtering of the type being considered is super-Nyquist. That is to say, in upscaling of a given image, reconstruction filtering necessarily synthesizes spectral content at frequencies above an initial Nyquist limit characteristic of a lower resolution. Thus, the internal representation that governs instancing of reconstruction filters may be leveraged as a codebook for an optimal encoding of corresponding structure in the target image. With use of such an encoding, processing gain realization hinges upon two criteria; (i) the encoded edge-contour structure representation remains more compact than the corresponding spectral representation and (ii) object and edge-contour spectra remain separable at a given Nyquist rate boundary. In the specific case of photometric warp, these criteria are satisfied via; (i) an implicit 2D-to-1D dimensional reduction rooted in processing of curvilinear *edge-contours bordering objects*, (i.e. as opposed to *objects*), and (ii) partitioning of edge-contour spectra are resolved to an extent consistent with accurate edge-contour reconstruction. Accordingly, video transmission payload is reduced to a sum consisting of; (i) edge-contour encoding plus (ii) source downsampled to the defined super-Nyquist boundary. Noting the downsampled source may itself be encoded by some CODEC operating on an orthogonal compression principle, (i.e. encode distinct structure) total compression is then given by a product of compression ratios generated via what amounts to a two-stage succession of CODEC transformations. In what follows, this seminal idea forms the basis of a Superresolution-Enabled [video] CODEC (SREC) that employs a lossy, differential/block encoded CODEC such as MPEG-4 as a second layer. In a comprehensive series of experimental trials, this construct is then shown capable of an 'x4', 'x16' increased compression ratio capability relative to MPEG-4 alone. For example, if stage 1,2 compression ratios are 4:1 and 30:1 respectively, the total compression is 4x30:1 = 120:1, adjusted downward so as to account for stage-1 reconstruction filter bank instance ENCODE.

It is noteworthy, a variety of SREC architectural forms are possible, depending upon specific blocks chosen for layer-1,2 processing. However, the source-encoded form displayed in figure-3 serves to illustrate the basic operational principle. In streaming transport mode, as a single-frame superresolution (SFSR) process, SREC accepts and processes video frame-by-frame, according to an ENCODE-transport-DECODE schema:

ENCODE

(1) Pattern Manifold Assembly (PMA)
(2) ENCODE PWRP reconstruction filter bank instances (layer-1)
(3) DOWNSAMPLE frame
(4) ENCODE downsampled frame (layer-2)
(5) INTERLEAVE PWRP/Downsampled frame data
(6) TRANSPORT Interleaved data stream (TX)

DECODE

(1) TRANSPORT interleaved data stream (RX)
(2) DEINTERLEAVE PWRP/Downsampled frame-data
(3) DECODE downsampled frame (layer-2)
(4) UPSAMPLE frame
(5) DECODE PWRP reconstruction filter bank instances (layer-1)
(6) RECONSTRUCT frame at original resolution

As might be expected, DECODE forms an inverted processing sequence relative to ENCODE. It is also significant, given PMA sub-pixel resolution, DECODE reconstruction filters are applied at the exact image coordinates at which filter instance ENCODE was performed. In this manner, edge dislocation noise in the reconstructed image is effectively reduced to zero.

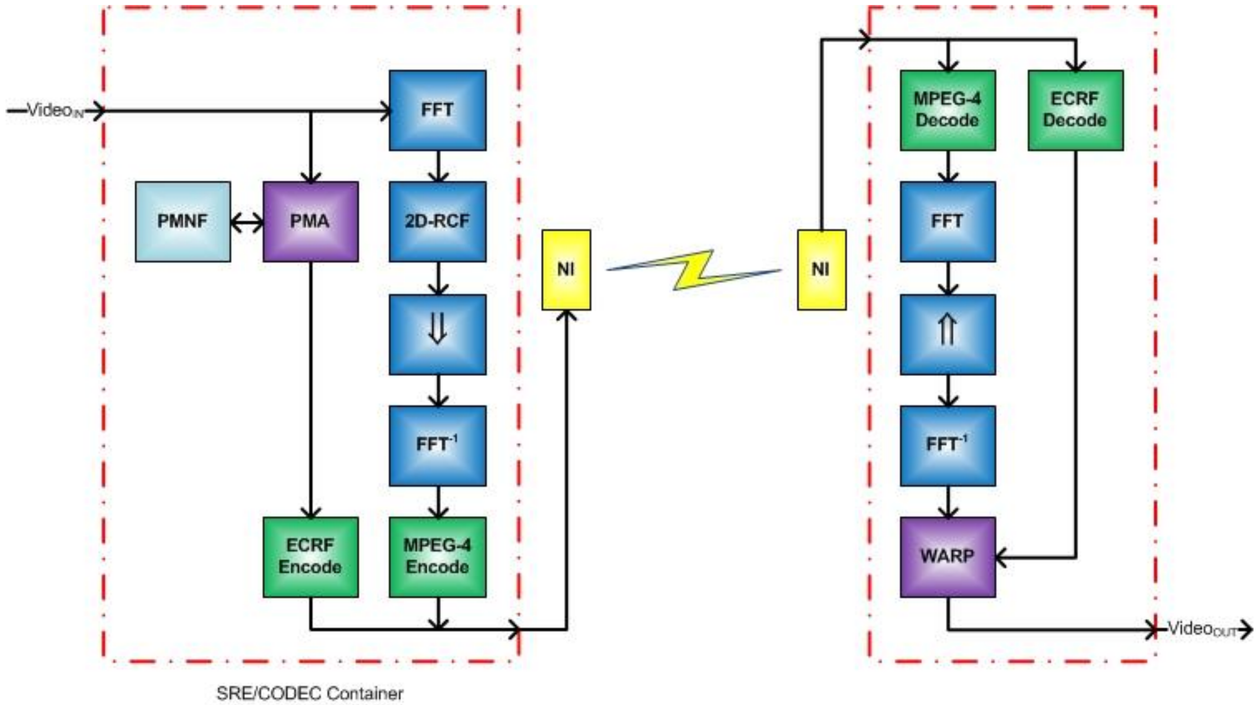

Figure-3: SREC ENCODE/DECODE Block Architecture

In a second SREC realization previously discussed, selected constituent blocks are replaced with SRRS instances, as described above:

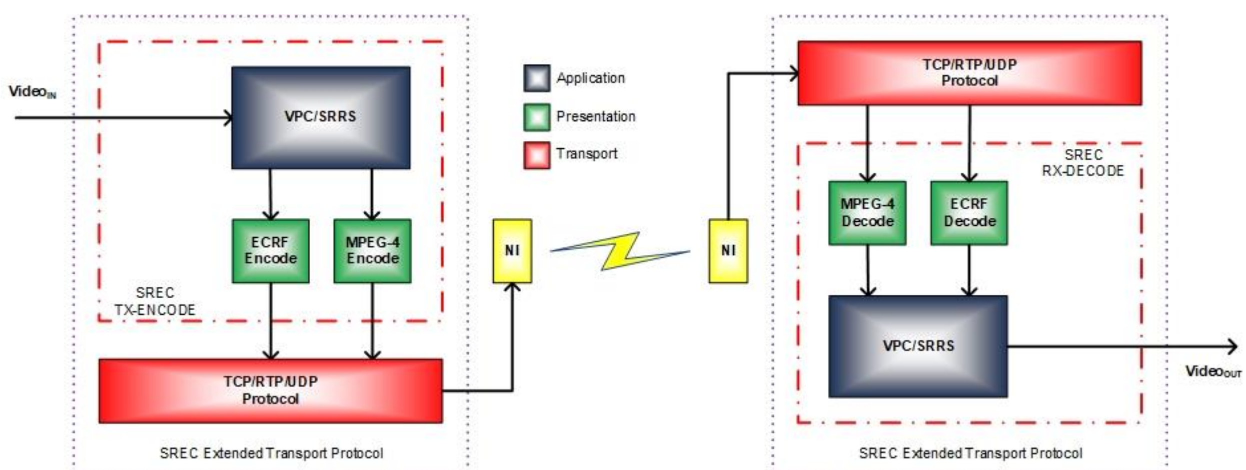

Figure-4: VPC/SRRS-based SREC Architectural Form

Video Analytics

Dimension Inc is well known for development of advanced image processing technologies pertinent to high-performance video applications such as IoT, Surveillance, Analytics, Video Conferencing, VR/AR, and Streaming Video Platforms. In particular, Dimension has focused upon image and video *reconstruction*, *denoise*, and *compression* (CODEC) as those processing components most critical to video system performance optimization. From a technical perspective, this work has been successful in creation of new algorithms of extraordinary power. Most notably, combination of *superresolution reconstruction*, *spatiotemporal denoise*, *Nyquist refinement*, and *superresolution-based video compression* have together enabled video processing gain at levels heretofore unavailable in the video technology marketplace.

As a matter of strategic planning, Dimension R&D has challenged itself with identification of new and emerging market sectors for which Dimension's technology may prove beneficial. One such is the *video analytics* ('VA') domain. From a systems perspective, video analytics applications are quite simple. As shown in figure-5, source video is applied to a bank of machine learning ('ML') classifiers that serve to detect and report objects appearing within a video sequence. In experimentation with such systems, Dimension has discovered an essential relation between video analytics *quality of result* ('QoR') and input video quality. To wit, *the error rate of any machine learning classifier one might employ for purposes of video analytics is highly dependent upon input video resolution and signal-to-noise ratio* ('SNR'). In retrospect, this result may seem obvious. However, it is one thing to guess at a relationship and yet another to prove its existence and then quantify it. Succinctly stated, Dimension has done exactly the latter, the most notable result of which is the *video preconditioning* ('VPC') subsystem cited in tables-1,2 above and discussed within context of *superresolution-enabled rescaling* (SRRS).

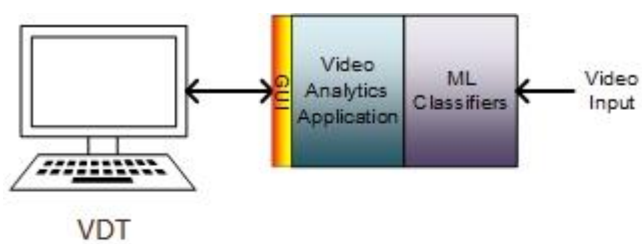

Figure-6: Exemplary Video Analytics System

It is significant VPC incorporates two image/video processing blocks critical to video QoR optimization; (i) adaptive spatiotemporal denoise (ASVD) and (ii) superresolution-enabled Nyquist refinement (SRNR). These blocks are each Dimension innovations and generic in applicability. Thus, it is perhaps obvious VPC would prove of benefit in applications other than the previously discussed superresolution-enabled video CODEC (SREC). In a recent expansion of technical focus, Dimension Inc has identified Video Analytics (VA) and BigData (BD) processing as two such application categories for which VPC will prove beneficial. In particular, Dimension has combined those previously mentioned image processing blocks in creation of a *video preconditioning pipeline* technology that serves to optimize performance of *neural-network* (NN) based *machine learning* classifiers we employ in VA and BD applications. More precisely, VPC generates a composite processing gain that minimizes classifier error rates across a broad swath of input video quality. For example, use of VPC as a VA front-end can be expected to improve overall quality-of-result (QoR). It should be noted where input video is noisy or exhibits reduced bandwidth, this improvement can be dramatic.

BigData-IDE

It should be emphasized, , the VPC advantage articulated here applies to *all* video analytics systems. Arguably, this alone constitutes a highly significant result. However, it soon became apparent Dimension Inc could go much further. With the proven success of VPC, it was a small conceptual step to consider just how VPC might be applied within context of *any ML-based information processing on video content*. It was at this juncture, Dimension shifted its attention to the problem of *BigData*. In this consideration, two facts became apparent; (i) while the term BigData is currently prominent in the media, a formal definition of the term seems absent and (ii) whatever we might take to constitute BigData, more than 50% accrues in form of video. This is in fact Dimension's entry point to the BigData arena. Here, we will adopt a working-definition of BigData in terms of the *sum total of all our electronics communications*. Now, with this conceptual stake-in-the-ground, it seems reasonable to conclude the cited 'more than 50% video' represents a significant strategic opportunity for application of VPC technology within context of BigData exploration, or for that matter any extraction of video-based information from the BigData DATAVERSE.

The bottom-line is, at the current level of technological evolution, we might consider BigData DATAVERSE as existing only in form of a conceptual icon. That is to say, beyond unstructured web-queries, there exists no machine compatible representation for which BigData accrues as an efficiently searchable information resource. This of course represents an impediment to any expanded application of Dimension technology because the BigData market sector is in effect non-existent. However, as a matter of strategic consideration, it is also obvious this BigData we envision will evolve as the next-big-thing within the video analytics milieu. It then follows BigData itself must be invented and, as the immortal John Lennon once advocated, we must think differently in doing so. Why is this? In terms of formal complexity, the BigData DATAVERSE is understood as *exascale*. More simply put, 'BigData' is indeed very big! It then follows, the essential problem of BigData devolves to one of exascale processing. It is this rationale that forms the basis for Dimension's consideration of processing infrastructure sufficient to development of video-based information processing applications on BigData DATAVERSE. This would of course include classic video analytics applications as we have come to understand them, but now with addition of more advanced information processing capabilities.

BDA/BDE/BDD

From what we now construe as the BigData application domain, Dimension has resolved the problem of BigData processing into three component tasks; (i) *assembly*, (ii) *exploration*, and (iii) *discovery*. In simplest terms, BigData Assembly (BDA) is the rendering of an unstructured DATAVERSE into an efficiently searchable form. This task is comprised of an essentially autonomous process of building an entity-attribute graph (EAG) representation to which existing and readily available graph-search techniques may be applied. BigData Exploration (BDE) is then extraction of statistics on *keywords*, *declaratives*, and *relations* (KDR) pertinent to truth-value of some logical-conjecture-on-data (LCOD) posed within context of the analytics we wish to perform. BigData Discovery (BDD) is then the identification of *new* KDR elements, with subsequent reapplication of BDA, BDE, and possible restatement of any LCOD we apply within context of BDE. In effect, BDD will initiate invocation of BDA and BDE as subprocesses within context of an overarching problem of evaluating LCOD truth-values. Dimension's strategic vision is BDA, BDE, and BDD together form a complete processing infrastructure for development of video analytics platforms on BigData DATAVERSE. Based upon this notion, an integrated processing environment for BigData is proposed in figure-7, each component of which will employ VPC technology. From a larger perspective, we see Dimension is thrust into a position of not only developing new video analytics technology but also contributing to an entirely new BigData market sector. It goes without saying, this has proven an exciting prospect for Dimension's further evolution as a technology leader.

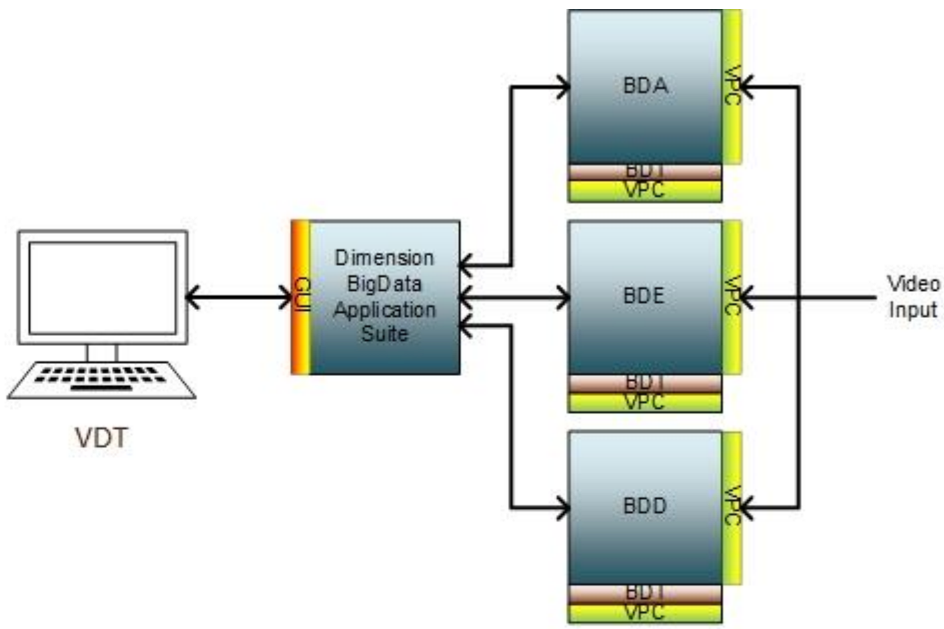

Figure-7: Dimension BigData Application Suite

It is also true each of the aforementioned BDA, BDE, and BDD applications incur significant nondeterminism. That is to say, these tasks are appropriately cast in terms of goal-directed, nondeterministic solution-search we understand as characteristic of *artificial intelligence* (AI) applications. Accordingly, AI architectural forms are indicated for each of these component processes. A highly simplified architectural form by which all component processes may be invoked is displayed in figure-8 for which we note the aforementioned BDD-initiated invocation of BDA and BDE AIs is implemented via a top-level supervisory process that is also AI. In this manner, the problem of BigData DATAVERSE processing further devolves to knowledge-based application of BDA, BDE, and BDD, the result of which we represent as a solution-state trajectory comprised of EAG node-visitations. In this particular architectural variant, we also note use of Dimension's *superresolution-enabled video compression* (SREC) based network video transport (BDT) along with a cloud-based BDA implementation. As displayed in figure-4 above, SREC implementation is further simplified and unified with use of VPC as an architectural building-block. It should be noted *the resulting hierarchical AI architectural-form is representative of an entire class of cutting-edge innovations*, most generally for AI technology and specifically for exascale processing on BigData DATAVERSE.

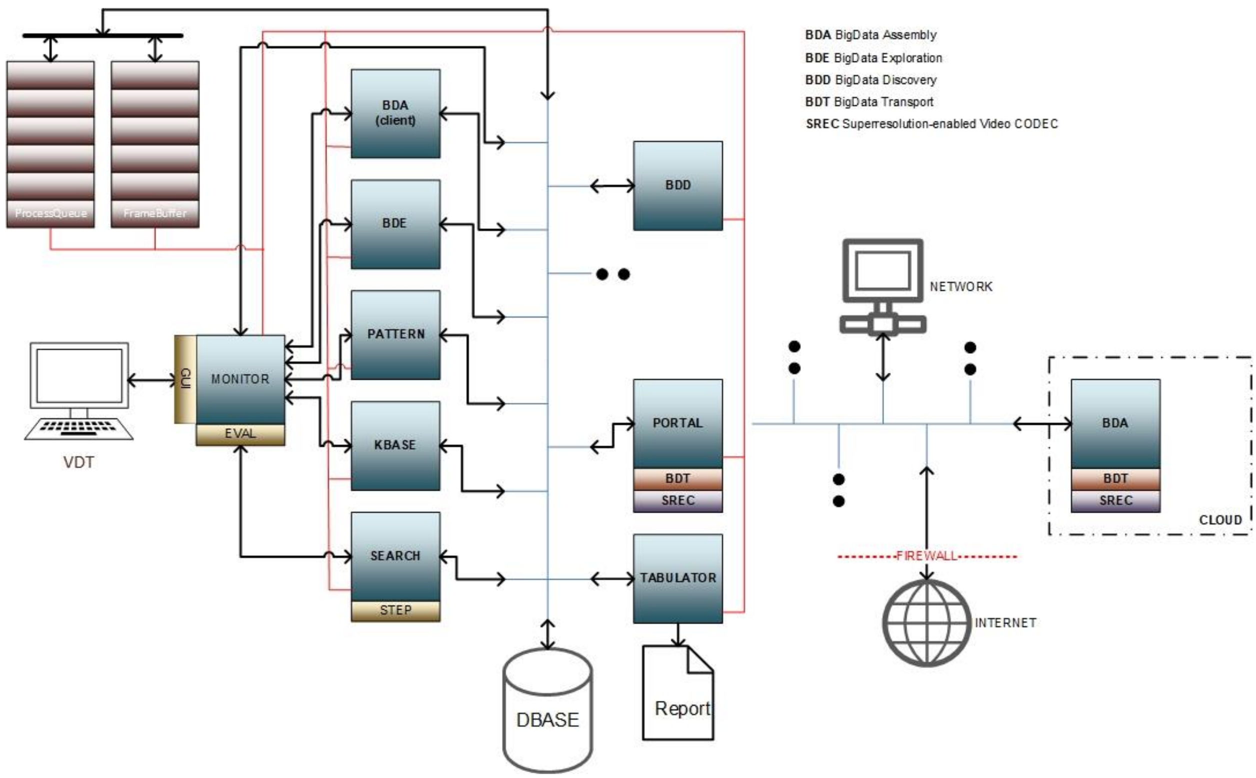

Figure-8: Cloud-Distributed BigData AI Architecture

SRAU

Pending

SRRU

Pending