Alder IP Pty Ltd
IP Lawyers and Patent Attorneys
ABN: 28157949114
Suite 302, 10 Help Street
Chatswood NSW 2067
Australia
P: +61 02 90079916
Email: info@alderip.com.au

PRIVILEGED AND CONFIDENTIAL

Ion Video Pty Ltd
Level 2, 161 Collins St,
Melbourne 3000, Victoria

Alder IP Ref: 42282

Attention: Anthony Baker

April 01, 2026

Re: Patent "Freedom to Operate" search for ION Video Pty Ltd relating to "A Non-Destructive Virtualisation and Programmable Assembly of Rendered Video"

Dear Anthony,

We have completed our Freedom to Operate patent search relating to **A Non-Destructive Virtualisation and Programmable Assembly of Rendered Video** as described in the provisional specification filed and the documents shared.

A. SCOPE

A Freedom to Operate Search is a search performed to determine whether a specific product, process, can be commercially launched without infringing on existing, in-force patents held by others.

The primary objective is to determine if the production, sale, or use of the invention infringes upon any existing patent claims. If the Applicant (company) has the freedom to operate within its industry without facing any legal consequences.

The FTO search was conducted using different databases, such as Patseer, Patentscope, USPTO, ESPACENET, Free Patents Online, AusPat and Google Patents to extract relevant references.

The search covered the below jurisdictions:

- USA
- Europé
- Australia and
- WIPO

For the US database, the search was for patent applications granted and/or published in the US in the last 25 years; while for the WIPO database, the search was in the last 31 months.

Please note that patents filed outside of the aforementioned jurisdictions would not have been detected. Additionally, only patents filed at least 18 months prior to this search would have been detected by the search as it generally takes 18 months for the patent offices to publish the specifications.

For infringement to occur, the alleged infringer must be shown to have used the claimed invention for a commercial purpose within the protected jurisdiction during the life of the patent. The alleged infringing product must take or use each and every feature of at least one claim of a granted patent to infringe.

Pending patent applications cannot be technically infringed until they have been examined, found valid and then subsequently been granted. Pending applications identified in the report were not extensively reviewed by Alder IP as the claims and scope of pending patent applications is usually restricted during patent examination processes.

Additionally, it is possible, under special circumstances in USA, to request reissuance or re-examination to vary the claims of granted patents and these instances would not have been detected by this FTO Search.

Please note that expired patents are open to public access and freely available to be used without license or approval of the owner.

In USA, the maximum term may be extended wherein the owner has informed the Patent Office of delays caused by clinical trials (these extensions may be up to 5 years) and also US Patent Office awards automatic extensions to maximum term for time lost during the examiner's consideration and evaluations.

We note that whilst all possible care is taken when compiling this report, some patents may not have been taken that were incorrectly coded or entered by the relevant patent offices and we are not responsible for errors in the Patent Offices databases.

B. KEY FEATURES

In conducting the search, we used the essential features of the invention listed below :

A. A computer-implemented method for non-destructive virtualisation and programmable assembly of rendered video, comprising:

1. Separating, by one or more processors, a digital video file compliant with an ISO Base Media File Format into (i) structural metadata describing temporal and spatial organisation of the video and (ii) compressed media samples containing audiovisual data

2. Storing the compressed media samples without modification, duplication, or re-encoding, and maintaining addressable byte-range access to said samples.

3. Generating a virtualised structural representation of the video that references the compressed media samples via byte offsets and time-aligned indices.

4. Creating a machine-readable semantic manifest associated with the virtualised structural representation, the semantic manifest comprising time aligned descriptors, conceptual labels, and mappings between semantic meaning, temporal ranges, and corresponding byte ranges of the compressed media samples.

5. Receiving an input instruction comprising a natural language prompt or programmatic query.

6. Interpreting, by an artificial intelligence system, the input instruction to determine an intended video output.

7. Mapping the intended video output to one or more temporal segments defined within the semantic manifest.

8. Resolving the temporal segments into corresponding byte ranges of the compressed media samples via the virtualised structural representation.

9. Dynamically generating, at runtime and without modifying the compressed media samples, a standards-compliant video stream by constructing container structural metadata that references the resolved byte ranges; and delivering the generated video stream for playback.

10. The video stream is assembled deterministically in response to the input instruction without transcoding, re-rendering, or duplicating the compressed media samples, thereby enabling programmable, queryable, and non-destructive interaction with rendered video content.

C. SUMMARY OF KEY IDENTIFIED PATENTS

Assignee /Applicant	Patent/Publication No.	Title	Expected Expiry Date	Potential Risk of Infringement	Comment
Comcast Cable Communications LLC	US 11,308,159 B2,	Dynamic detection of custom linear video clip boundaries	29-05-2039	Low	This document discloses a method of generating content portions (assembling video segments) by comparing input queries with structural metadata to identify specific temporal boundaries. It maps to almost all taxonomy features except for specific addressable byte-range storage. The linear content metadata is used as a virtualized representation to locate and extract relevant portions. The present invention maintains addressable byte-range access to media samples without modification or duplication. It specifically resolves temporal segments into corresponding byte ranges via a virtualized structural

					representation, a mechanism not absent in this document. Furthermore, the cited document establishes temporal boundaries (time-based), whereas the present invention resolves those temporal segments into corresponding byte ranges via a virtualized structural representation
Google LLC	US 12,149,773 B2	Voice-based scene selection for video content on a computing device	23-05-2043	Low	This cited document identifies video content and accesses scene metadata containing semantic descriptions and timestamps. It interprets spoken utterances to determine if they are scene playback requests and causes a media player to seek to a specific location. The present invention creates a virtualized structural representation referencing media samples The cited document does not disclose the use of byte offsets Th feature of **standards-compliant video stream** at runtime without transcoding or re-rendering is absent in the cited document.
Rishi Kumar	US20250291845 A1	Artificial intelligence assisted streaming video scene selection	Pending	Low	Cited document employs a machine learning model to interpret user queries describing scenes and determines relevant scenes within a content database for display. It delivers video segments for playback based on the query. The present invention separates a digital video file into structural metadata and compressed media samples. The cited document does not disclose this separation.. Furthermore, the present invention creates a semantic manifest that maps conceptual labels directly to byte ranges of the compressed samples.

A more detailed summary of the full list of identified patents is attached as **Annexure A** and this list includes extracts of the independent claims. The above summary is focused on the key patents of interest that were detected and focused on the closest examples of the prior art in our opinion.

We also note that we have assumed that all patent and patent applications identified in this search were valid and that no assessment of novelty or inventiveness of each identified patent has taken place.

D. CONCLUSION

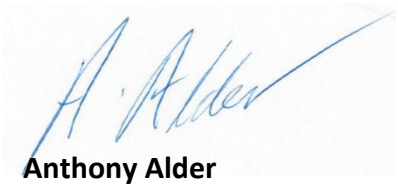
Based on the above analysis, Alder IP believes that ION Video product does not reasonably infringe any of the reviewed patents provided that the risk mitigation advice has been accepted and implemented for the relevant jurisdictions.

Please refer to **Annexure A** for a detailed summary of all patents reviewed.

This report and the analysis within are intended solely for the purpose of providing an opinion regarding the product. This interpretation is based on publicly available information from various sources and should not be construed as a formal legal opinion.

The findings are for the internal use and no part of this report shall be publicly distributed, published, or offered for sale without the explicit permission. This report is intended to serve as a risk assessment tool, it does not constitute a guarantee of definitive legal clearance for commercialization.

Yours faithfully



Anthony Alder
NSW Supreme Court Solicitor/Patent Attorney
Btech (Biotech) LLB MIP FIPTA

Encl. Annexure A – Detailed Summary
Annexure B – FTO Raw Data Report

Annexure A

Assignee/ Applicant	Patent/Publication No.	Title	Expected Expiry Date	Claim discussion	Proposed Risk Mitigation
Comcast Cable Communications LLC	US 11,308,159 B2	Dynamic detection of custom linear video clip boundaries	29-05-2039	1.A method comprising: receiving a query associated with content, the query comprising a first portion and a second portion; determining a first match in content metadata for the first portion; determining, based on the first match, a start boundary preceding a time associated with the first match; determining a second match in the content metadata for the second portion; determining, based on the second match, an end boundary following a time associated with the second match; and generating, based on the start boundary and the end boundary, a portion of the content. 9. A method comprising: receiving a query associated with content, the query comprising a first portion and a second portion; determining a first match for the first portion in linear content metadata; determining, based on the first match, a start boundary preceding a time associated with the first match; determining a second match for the second portion in second linear content metadata received subsequent to the first match; determining, based on the second match, an end boundary following a time associated with the second match; determining a third match for the first portion or the second portion in third linear content metadata received subsequent to the second match; extending, based on the third match, the end boundary to a time associated with the third match; and generating, based on the start boundary and the end boundary, a portion of the content. 16. A method comprising: receiving a query associated with content, the query comprising a first portion and a second portion; determining a first part of content metadata that is analogous to the first	Comcast focuses on detecting boundaries in linear content metadata to define clips. The present invention virtualizes the file structure. It separates structural metadata from media samples to enable addressable byte-range access without re-encoding, which is a specific file-system level innovation not present in Comcast metadata-matching approach. Avoid claiming the calculation of start/end times as the primary invention. Focus on the mapping of those boundaries to addressable byte ranges for non-destructive assembly.

				portion; determining a start boundary associated with the first part of the content metadata, wherein the start boundary precedes a time associated with the first part of the content metadata; determining a second part of the content metadata that is analogous to the second portion; determining an end boundary associated with the second part of the content metadata; and generating, based on the start boundary and the end boundary, a portion of the content.	
Google LLC	US 12,149,773 B2	Voice-based scene selection for video content on a computing device	23-05-2043	1. A method implemented by one or more processors comprising: receiving, from a user and via a computing device, a spoken utterance that includes a query; identifying video content being presented in a vicinity of the user by a media player application when the spoken utterance is received from the user; accessing scene metadata associated with the identified video content, wherein the scene metadata includes, for each of one or more respective scenes in the identified video content, semantic scene description data describing the respective scene and timestamp data identifying one or more locations in the identified video content corresponding to the respective scene; determining, based on the query and the scene metadata associated with the identified video content, whether the query in the spoken utterance is a scene playback request directed to the media player application to play a requested scene in the identified video content; in response to determining that the query in the spoken utterance is a scene playback request, causing a media control command to be issued to the media player application to cause the media player application to seek to a predetermined location in the identified video content corresponding to the requested scene and identified in the timestamp data of the scene metadata for the identified video content; and in response to determining that the query in the spoken utterance is not a scene playback request directed to the media player application, causing a non-scene playback request operation to be executed for the query included in the spoken utterance. 12. A system comprising one or more processors and memory operably coupled with the one or more processors, wherein the memory stores instructions that, in response to execution of the instructions by one or more processors, cause the one or more processors to perform a	Google patent is a UI/UX method for voice-based scene selection. Distinguish based on how the video is delivered. Google seeks within an existing file. The present invention dynamically generates a new standards-compliant stream at runtime by mapping semantic queries to specific byte offsets, without transcoding. Avoid focusing on the spoken utterance or voice input. Rather, focus on the backend resolution where the query is resolved into a virtualized structural representation.

				method that includes: receiving, from a user and via a computing device, a spoken utterance that includes a query; identifying video content being presented in a vicinity of the user by a media player application when the spoken utterance is received from the user; accessing scene metadata associated with the identified video content, wherein the scene metadata includes, for each of one or more respective scenes in the identified video content, semantic scene description data describing the respective scene and timestamp data identifying one or more locations in the identified video content corresponding to the respective scene; determining, based on the query and the scene metadata associated with the identified video content, whether the query in the spoken utterance is a scene playback request directed to the media player application to play a requested scene in the identified video content; in response to determining that the query in the spoken utterance is a scene playback request, causing a media control command to be issued to the media player application to cause the media player application to seek to a predetermined location in the identified video content corresponding to the requested scene and identified in the timestamp data of the scene metadata for the identified video content; and in response to determining that the query in the spoken utterance is not a scene playback request directed to the media player application, causing a non-scene playback request operation to be executed for the query included in the spoken utterance.	
Rishi Kumar	US 20250291845	A1 Artificial intelligence assisted streaming video scene selection	Pending	**1**. An apparatus comprising: at least one memory; and at least one processor coupled to the at least one memory, the at least one processor configured to: receive, from a remote device, a user query describing a scene that is accessible via a streaming video application; search, via a machine learning (ML) model and based on the user query, a content database associated with the streaming video application; determine, via the ML model and the user query, one or more relevant scenes within the content database; and display, via a display device, the one or more relevant scenes. **8**. A computer-implemented method comprising: receiving, from a remote device, a user query describing a scene that is	This pending application covers using ML models to find relevant scene" in a streaming database. Avoid claiming AI-assisted searching or "ML-based scene determination" as standalone nsteps. Position the AI as a mapper rather than a searcher. The AI of the pending

				accessible via a streaming video application; searching, via a machine learning (ML) model and based on the user query, a content database associated with the streaming video application; determining, via the ML model and the user query, one or more relevant scenes within the content database; and displaying, via a display device, the one or more relevant scenes. **15**. A non-transitory computer-readable storage medium comprising at least one instruction for causing a computer or processor to: receive, from a remote device, a user query describing a scene that is accessible via a streaming video application; search, via a machine learning (ML) model and based on the user query, a content database associated with the streaming video application; determine, via the ML model and the user query, one or more relevant scenes within the content database; and display, via a display device, the one or more relevant scenes.	application returns a result, whereas the AI in the present invention drives a file-system operation.
Jason HendersonJulia Guzman-Henderson	US20250047939A1	Machine-Learning Assisted Personalized Real-Time Video Editing and Playback	Pending	**1**. A system for machine-learning assisted personalized real-time video editing and playback, comprising: A user input module for receiving user preferences or search terms; A machine learning engine for analyzing the user's preferences and comparing them to metadata associated with video content; A content filtering and editing engine for editing the video in real time to match the user's preferences; A playback system for delivering the edited video content to the user's device. **4**. A method for machine-learning assisted personalized real-time video editing and playback, comprising: Receiving user preferences via a user input module; Analyzing video content metadata using a machine learning engine; Filtering the video content to remove irrelevant segments based on the user's preferences; Delivering the edited content dynamically in response to real-time user commands.	Jason focuses on real-time video editing and filtering irrelevant segments based on preferences. Jason edits the content, whereas the present invention edits the metadata references (Step 4 & 8) to the underlying samples. Mention that the compressed media samples are stored without modification or duplication (Step 2).
Pelco Inc	AU2011352094 B2	Searching recorded video	29-12-2031	1. A method for searching video data, the method comprising: receiving a search query from a user through a user interface, wherein the search query includes a plurality of query parameters indicative of one or more query characteristics that are characteristics of at least one of a query object or a query event associated with the query object;	Pelco uses distance measures to match search parameters to video segments. Avoid language that suggests the primary goal is to locate a

				calculating a distance measure between the plurality of query parameters and each of a plurality of sets of metadata parameters, each set of the metadata parameters being indicative of at least one of a candidate object or a candidate event, corresponding to what the query parameters are indicative of, in video data; and providing an indication of one or more video segments through the user interface, wherein each of the one or more video segments has a corresponding distance measure less than a threshold value. 7. A non-transitory computer readable medium comprising code configured to cause a processor to: receive a search query from a user through a user interface that includes a set of query parameters indicative of one or more query characteristics that are characteristics of a query object; calculate a distance measure between the set of query parameters and sets of object parameters each associated with an object identified in video frames captured by a camera, the objects being part of the video frames, and the set of query parameters and the sets of object parameters being indicative of similar object characteristics; and provide an indication of one or more video segments through the user interface, wherein each of the one or more video segments includes an object with a corresponding distance measure less than a threshold value.. 12. A video processing system comprising: a user interface configured to receive user input; 21 a video data storage device comprising video metadata associated with video data; and a processor communicatively coupled with the user interface and the video data storage device, wherein the processor is configured to: receive a search query through the user interface that includes a plurality of query values indicative of one or more query characteristics that are characteristics of a query object; and calculate a distance measure between the query values and a set of object values associated with the video metadata stored in the video storage device, the object values being associated with an object in the video data and being indicative of object characteristics similar to the query characteristics.	pre-existing video file or segment based on a similarity score. Do not focus on the mathematical closeness of the search term to the metadata, as this is recited in Pelco. Avoid claims that merely point the user to a list of matching clips. The present invention operates at the bitstream architecture level. It performs a low-level virtualization of the ISO Base Media File Format (ISOBMFF). Pelco treats the video as a searchable black box, whereas the present invention unpacks the box to create addressable byte-range access.
Tyco Fire and Security GmbH	US20250390533A1	Building security system with artificial intelligence video analysis and natural language video searching	Pending	**21**. A method of analyzing video files in a content search system, comprising: applying classifications to video files using an artificial intelligence (AI) model, the AI model trained according to training data comprising images separated into object of interest classes or foreign object classes corresponding to occlusion of an object of interest, the classifications	This document uses AI and NLP to search video files for specific objects in a security context. Avoid claiming NLP extraction of entities.

				comprising one or more objects or events; extracting, using natural language processing, one or more entities from a natural language search query received, in a natural language format, via a user interface, the one or more entities comprising one or more objects or events indicated by the natural language search query; searching the video files using the classifications applied by the AI model and the one or more entities extracted from the natural language search query; and presenting one or more of the video files identified as results of the natural language search query via the user interface. **32**. A system of video file analysis in a content search system, comprising: one or more processing circuits coupled with memory to: apply classifications to video files using an artificial intelligence (AI) model, the AI model trained according to training data comprising images separated into object of interest classes or foreign object classes corresponding to occlusion of an object of interest, the classifications comprising one or more objects or events recognized in the video files by the AI model; extract, using natural language processing one or more entities from a natural language search query received, in a natural language format, via a user interface, the entities comprising one or more objects or events indicated by the natural language search query; search the video files using the classifications applied by the AI model and the one or more entities extracted from the natural language search query; and present one or more of the video files identified as results of the natural language search query via the user interface. **40**. A non-transitory system of video file analysis in a content search system, comprising: one or more processing circuits coupled with memory to store instructions that, when executed by the one or more processors, cause the one or more processors to: apply classifications to video files using an artificial intelligence (AI) model, the AI model trained according to training data comprising images separated into object of interest classes or foreign object classes corresponding to occlusion of an object of interest, the classifications comprising one or more objects or events recognized in the video files by the AI model; extract, using natural language processing one or more entities from a	Focus on the structural resolution. The document presents one or more video files as results. Whereas, the present invention generates a single, continuous, virtualized stream composed of fragments resolved via byte-range indices.

				natural language search query received, in a natural language format, via a user interface, the entities comprising one or more objects or events indicated by the natural language search query; search the video files using the classifications applied by the AI model and the one or more entities extracted from the natural language search query; and present one or more of the video files identified as results of the natural language search query via the user interface.	
Dish Network Technologies India Pvt Ltd	US12225269 B2 |	Methods, systems, and apparatuses to respond to voice requests to play desired video clips in streamed media based on matched close caption and sub-title text	14-02-2040	1. A method, comprising: converting a voice request received at a local device to text; executing, by a server in communication with the local device, a search on a media database to identify media content that matches the text; and playing, on the local device, a video in response to the search identifying the video as matching the text. 8. A search system, comprising: a local device configured to transmit a voice request and to display content returned in response to the voice request; and a server in communication with the local device, wherein the server is configured to receive the voice request from the local device, wherein the server is configured to convert the voice request to text, wherein the server is configured to execute a search on a media database to identify media content matching the text, and wherein the server is configured to return search results comprising the identified media content to the local device. 16. A search server comprising a processor in communication with a non-transitory storage medium configured to store instructions that, when executed by the processor, cause the search server to perform operations, the operations comprising: converting a voice request received at a local device to text; executing a search on a media database to identify search results comprising media content that matches the text; and returning the search results to the local device.	This document converts voice to text and searches media databases/closed captions to play a matching video. Avoid text-to-video matching logic. Emphasize the Mapping to Semantic Manifest. The document relies on external text. On the other hand, the present invention creates a unique time-aligned mapping between semantic meaning and byte ranges, allowing for programmable interaction with the video file.