ON VIDEO ABSTRACTION SYSTEMS' ARCHITECTURES AND MODELLING

Víctor Valdés, José M. Martínez

Victor.Valdes@uam.es, JoseM.Martinez@uam.es

SAMT 2008, 3-5 December 2008, Koblenz (Germany)

Universidad Autónoma de Madrid
E28049 Madrid (SPAIN)

Video Processing and Understanding Lab
Grupo de Tratamiento e Interpretación de Vídeo

Outline

- Introduction
- Simplified Functional Architecture
- Towards a Generic Video Abstraction Architecture
- Abstraction Systems Modelling
- Generic Video Abstraction Architecture
- Conclusions

Outline

- Introduction
- Simplified Functional Architecture
- Towards a Generic Video Abstraction Architecture
- Abstraction Systems Modelling
- Generic Video Abstraction Architecture
- Conclusions

Introduction (I)

- **Video abstraction systems aim to ease the browsing of video repositories reducing the time needed to select the desired video**
  - Reducing the time spent visualizing the video (preview abstract)
  - Reducing the time (and bandwidth) for downloading the video
- **Video abstract: shorter but representative representations (semantic coverage) of the original content**
- **Video abstraction modalities can be grouped in two main groups**
  - Video-skim based summaries: highlights' videos, fwd video, trailers, etc.
  - Key-frame based summaries: story-boards, slide shows, video posters, etc.

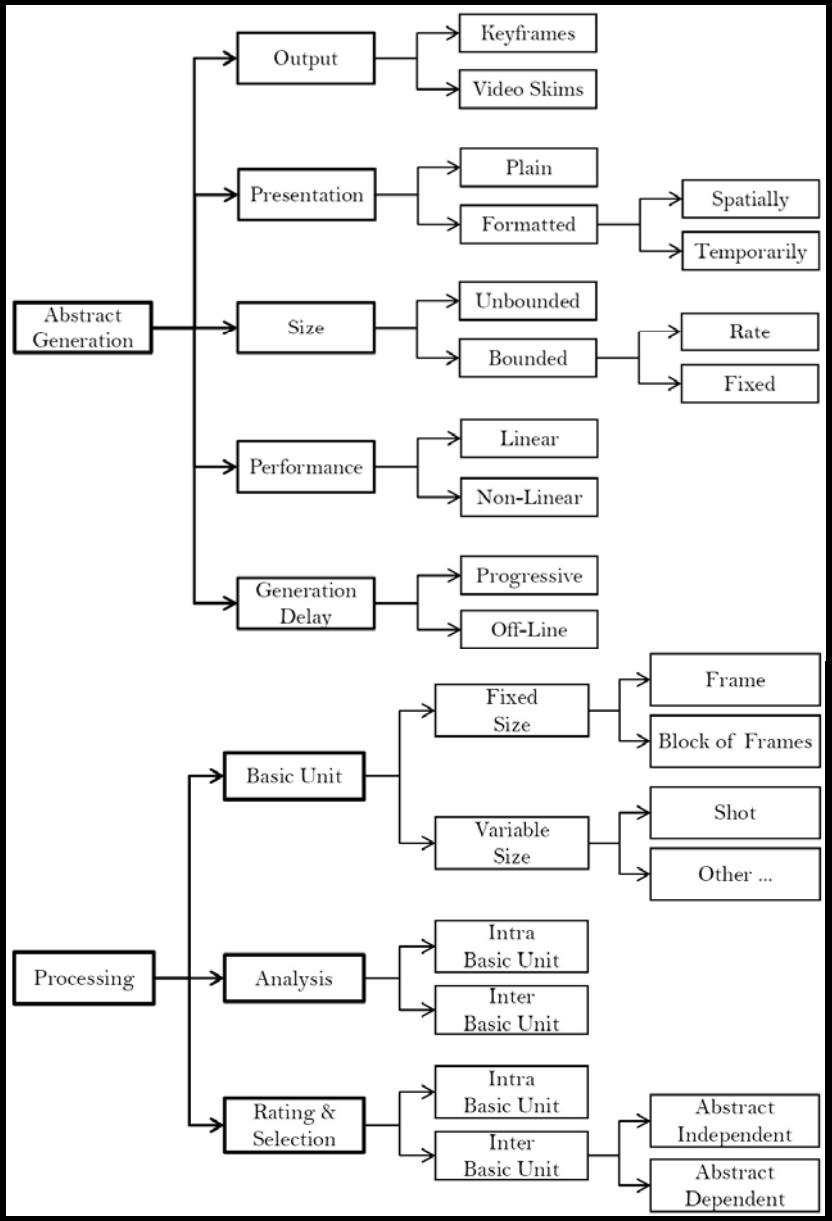
Introduction (II)

- **There exist a high heterogeneity in the different approaches to video abstraction, both at complexity level as well as at the huge amount of algorithms and techniques**
  - Nevertheless, most of these approaches share conceptual stages
  - Therefore it is possible to review and synthesize the different approaches to propose a generic abstraction functional model as well as a generic video abstraction architecture
- **In order to synthesize the different approaches it is good to look for a taxonomy of video abstraction systems from an operational point of view**
  - We have proposed a taxonomy grouped in two levels: external and internal characteristics
    - These characterization allows to group the different approaches in order to further synthesize their proposals in the different models that finally yield a generic architecture

Introduction (III)

- External characteristics specify how the result looks like (abstract modality, presentation, size) and external processing aspects (performance, generation delay).

- Internal characteristics are related to how the algorithms work with respect to BU: size of BU, analysis, scoring and selection in intra- or inter-BU mode



Introduction (IV)

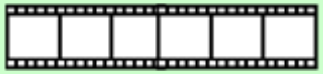
Objectives

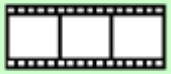
- Definition of a common framework enabling the application and study of abstraction techniques
  - The proposed models will ease the generic study of abstractions mechanisms and the restrictions required for building systems with specific external characteristics from an operational point of view
- Most of the existing literature, tutorials and surveys of video abstraction systems' State-of-Art deal with algorithms' categorization but not so many with architectural aspects
  - and none of them from a generalization point of view
- Our approach is to synthesize existing State-of-Art approaches to generalize them into a unified generic architecture for video abstraction systems
  - We may be somehow biased to create an architecture that accommodates on-line video abstraction (although the final architecture covers also off-line abstraction)

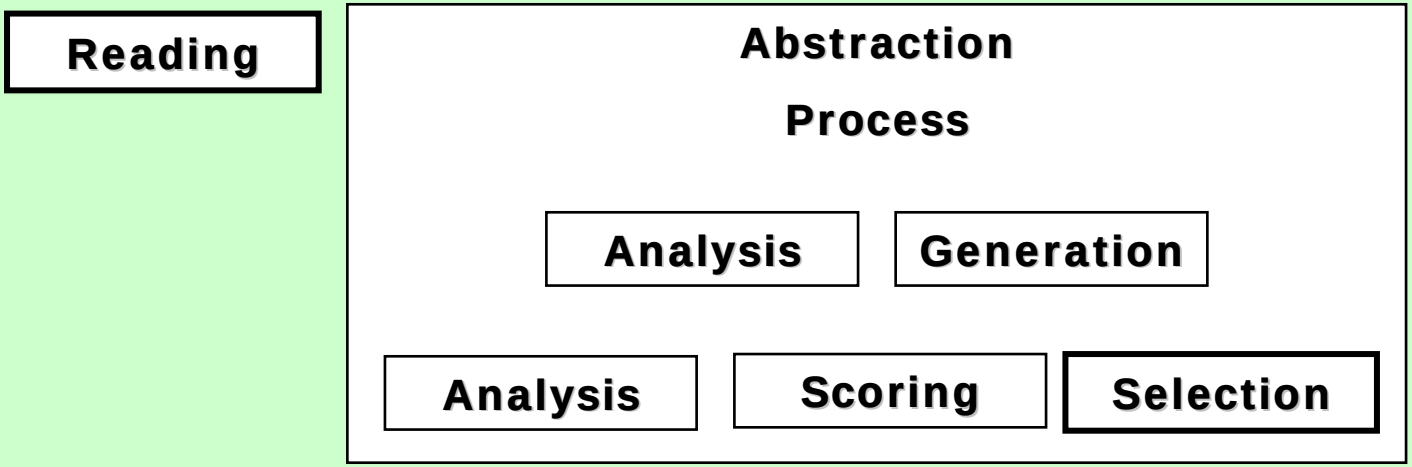
Outline

- Introduction
- Simplified Functional Architecture
- Towards a Generic Video Abstraction Architecture
- Abstraction Systems Modelling
- Generic Video Abstraction Architecture
- Conclusions

Simplified Functional Architecture (I)

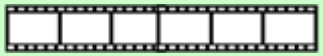






- Whilst this is a complete set, only the reading, selection and writing stages are mandatory (even for the most simple approaches like uniform subsampling or random selection of BUs)
  - Another view of this simplified approach may include always scoring and selection, but this is more “complex” and imposes a restriction in the (naïve) selection stage (a scoring stage with binary output that will be followed by a naïve binary selection) for the simplest subsampling approach.

Simplified Functional Architecture (II)

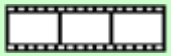


Reading

Abstraction Process

Analysis | Generation

Analysis | Scoring | Selection



Writing

- Scoring and Selection modules can balance the complexity of the generation stage
  - Simple scoring followed by complex selection
  - Complex scoring followed by a simple threshold based selection
- Any abstraction system can fit in this model
  - by putting all the algorithm complexity in the scoring module with a binary output with respect to the inclusion or exclusion of the processed BU (naïve selection)
  - Usually there will be a balance
    - Selection based on quantitative characteristics (e.g., size, continuity) and maximization of the accumulated score based on the individual scoring at the scoring stage (without knowing details of the scoring)
- The functional architecture can be completed with the minimal (but generic) set of repositories and data flows in order to have a Generic Video Abstraction Architecture

Outline

- Introduction
- Simplified Functional Architecture
- Towards a Generic Video Abstraction Architecture
- Abstraction Systems Modelling
- Generic Video Abstraction Architecture
- Conclusions

Towards a Generic Video Abstraction Architecture (I)

- **The objective is to provide a modular, as simpler as possible, architecture were all the abstraction approaches fit.**

- **Besides architectural modularity, there is a modularity with respect to data processing units (Basic Units –BUs-) that are processed one after the other in each module**
  - BUs may range from single frames to the complete video sequence, including, among others, specific frames (e.g., I-frames), GoPs, shots, ...

- **The interface between modules is defined as the information (video content and metadata, as well as information about the parts of the summary already processed –e.g., already rejected or selected-) passed between them each time a BU is processed at each module.**
  - Whilst the processing is BU-by-BU, it may happen that BUs are not delivered from a module until a group has been processed.

Towards a Generic Video Abstraction Architecture (II)

- **The abstraction process is considered as the flow of BUs through the different modules**
  - Each module can accumulate, process, redirect, discard or select BUs
  - Each module can produce metadata of the original BUs (low-level features, semantic classification, ...) as well as metadata of the abstract (what happens to one BU may imply recalculation of the remainder or future BUs in the processing... allowing feedback...)
    - Content Metadata travels associated to the BUs
    - Abstract Metadata is stored in a repository giving the opportunity to be used by previous modules for processing next BUs
  - Each module may use additional contextual metadata for customizing the video abstract
    - User preferences

Towards a Generic Video Abstraction Architecture (III)

Repositories

- Abstract metadata repository with Information about the currently generated abstract
  - Actual length of the abstract
  - BUs already selected and their description
  - …
- User Preferences Repository in order to guide the abstraction process by user defined constraints
  - Target length of abstract
  - Presentation modality and media format
  - Content genre preferences (classification) for filtering during scoring or selection
  - … Features to analyze?

Outline

- Introduction
- Simplified Functional Architecture
- Towards a Generic Video Abstraction Architecture
- **Abstraction Systems Modelling**
- **Generic Video Abstraction Architecture**
- **Conclusions**

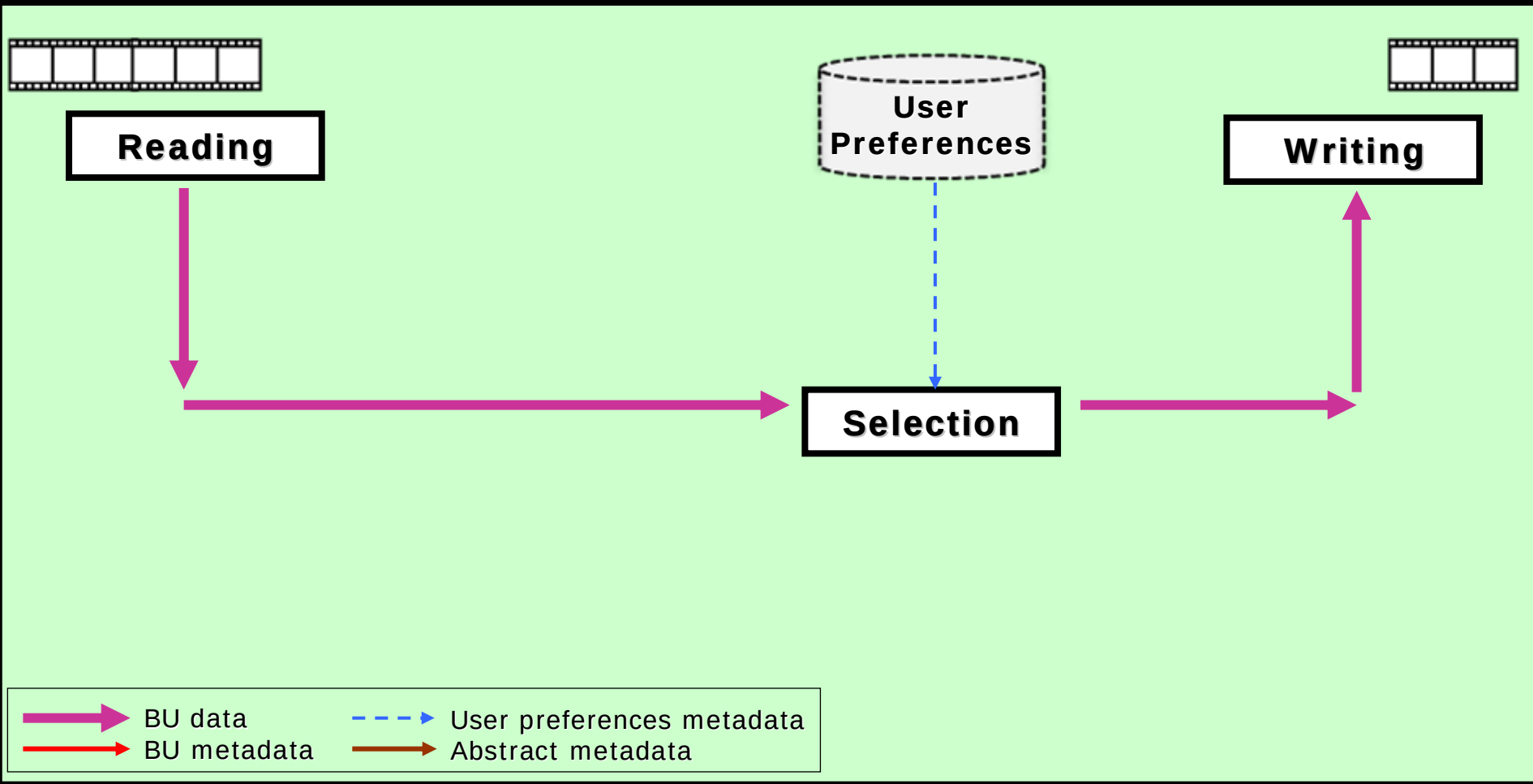
Abstraction Systems Modelling (I): Introduction

- **In order to reach the generic architecture, and starting from the functional modules and additional components already identified, we will progress from simple abstraction approaches to more complex ones (complex models cover and expand the simpler ones)**
  - Non-iterative systems: each BU is processed at most one time per module. Three models are identified:
    - Only selection
    - Analysis, scoring and selection
    - Analysis, scoring and selection with abstract metadata (feedback based on already created abstract)
  - Iterative systems: each BU can be iteratively scored after being processed by the selection stage, even the BUs can be sent to the scoring after other BUs have been processed
    - Analysis, iterative scoring and selection with abstract metadata (feedback based on already created abstract) and re-scoring of “surviving” BUs.

Abstraction Systems Modelling (II): Non-iterative, only selection

- **Most simple system**
  - Only selection is applied to the defined BUs (or a keyframe of each BU)
  - User preferences: abstract rate (defined as rate of BUs)
- **Examples**
  - Subsampling: usually uniform but may be random
    - Size: unbounded if the size of the original video is unknown, the system may adapt the sampling rate to the target rate
    - Delay: negligible and progressive

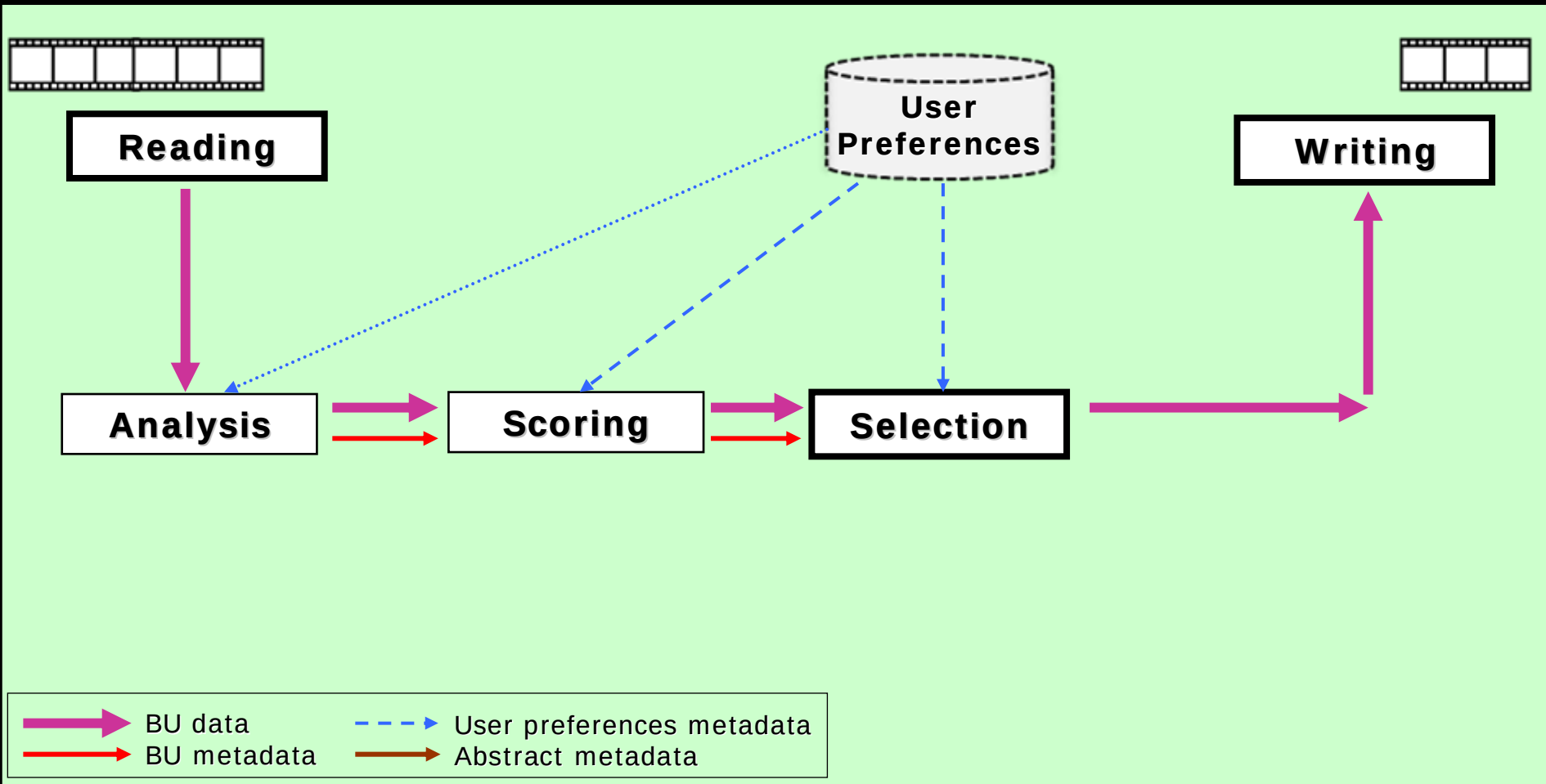
Abstraction Systems Modelling (III): Non-iterative, only selection



Abstraction Systems Modelling (IV): Non-iterative, analysis, scoring and selection

- **Complete non-iterative system without abstract metadata repository**
  - The Analysis module provides the value of different features
  - Scoring depends only on the original BUs (no feedback) creating a relevance value from the output of the analysis module
  - User Preferences: for scoring based on content classification and for selection (based on output length, for example) –for analysis may select relevant features-
- **Examples**
  - Adaptive subsampling systems: based on the relevance value, each BU (or group of BUs) is subsampled with a different rate at the selection stage
  - Relevance curve-based systems: based on the relevance value each BU is selected or discarded if the value is over or below a threshold
  - Clustering based systems (off-line): the clustering is performed in the scoring module based on the relevance value (or the vector of features from the analysis stage), and the score is given based on the distance of the BU to the centroid of its cluster. Selection will select the BUs closer to each cluster centroid. The number of clusters is a priori defined taking into account the size restriction.

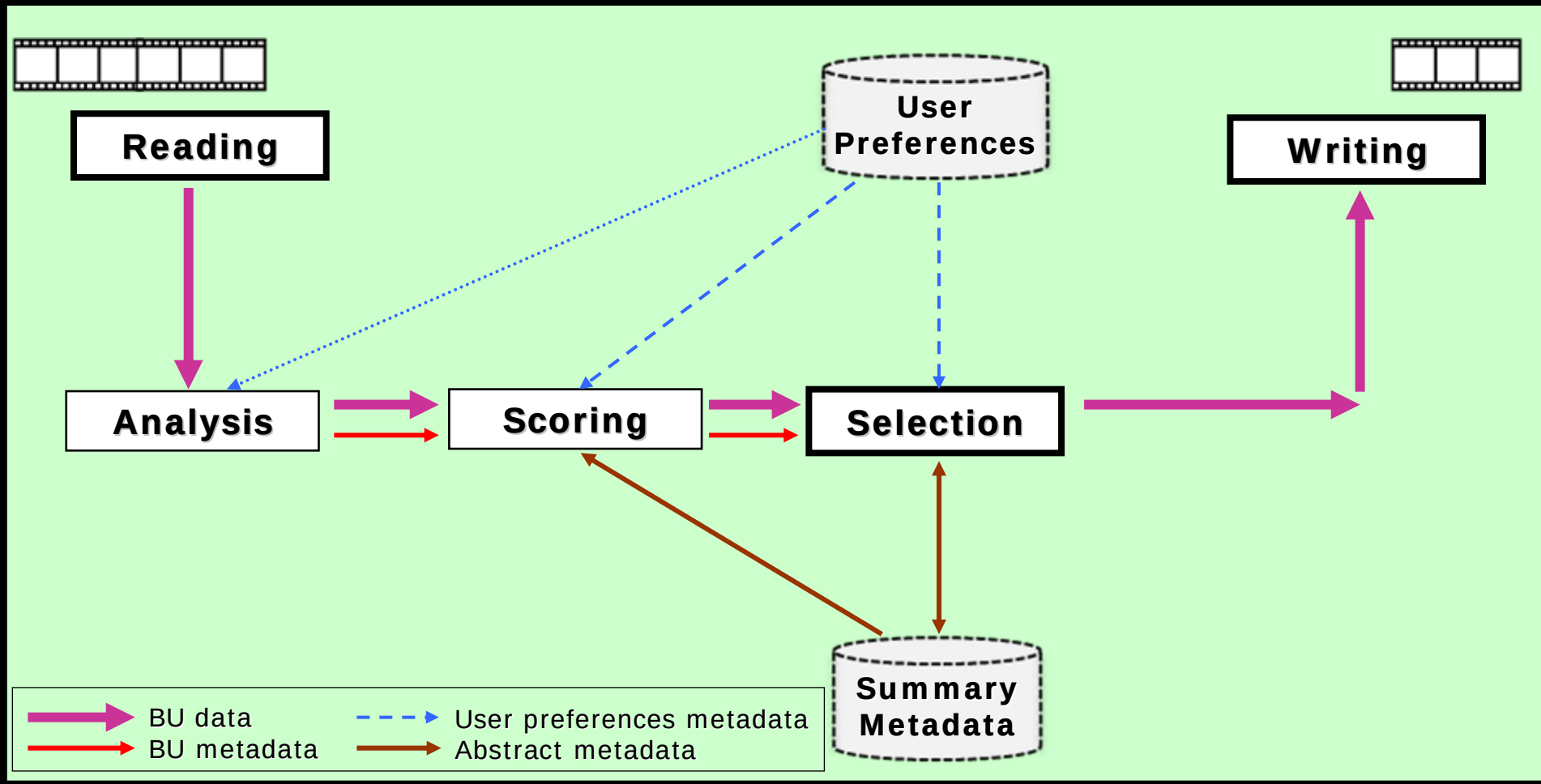
Abstraction Systems Modelling (V): Non-iterative, analysis, scoring and selection



Abstraction Systems Modelling (VI): Non-iterative, analysis, scoring and selection with metadata feedback

- **Complete non-iterative system with abstract metadata repository**
  - Scoring depends on the original BUs and the already selected Bus (e.g., for reducing redundancy and indirectly enhancing semantic coverage with a non-iterative approach).
    - Allows feedback
- **Examples**
  - Filtering by content change: scoring is based on analysis results and penalized (even with a temporal decay in the penalization) if similar content has already been selected (e.g., retake removal(on-line)/selection(off-line) in TRECVID BBC Rushes). The model allows to accommodate content filtering (e.g., junk removal like clapboards in TRECVID BBC Rushes) if the abstract metadata is preloaded with “forbidden” BUs (metadata of them)
    - In the case of the simplest selection only model, the abstract metadata may help to reduce redundancies
  - Adjustable rate depending on content selected (target versus actual rate)

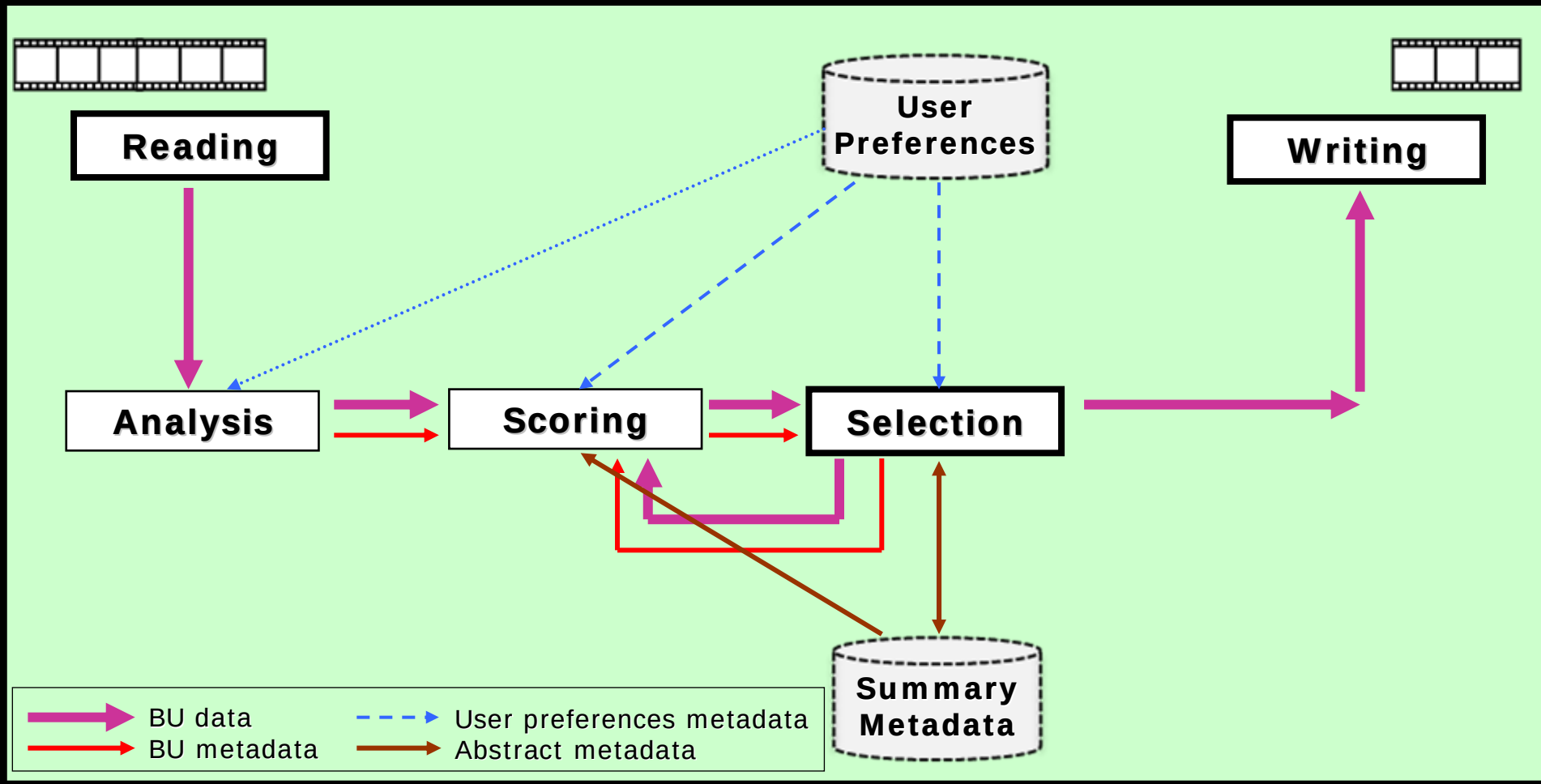
Abstraction Systems Modelling (VII): Non-iterative, analysis, scoring and selection with metadata feedback



Abstraction Systems Modelling (VIII): Analysis, iterative scoring and selection with metadata feedback

- **Complete iterative system with abstract metadata repository**
  - Allows iterative processing of BUs, providing a second feedback loop. After selection or rejection the remainder BUs can be scored again for maximizing the abstract criteria (e.g., semantic coverage)
- **Examples**
  - Maximum frame coverage: after analysis the scoring module calculates the number of BUs similar to the one being processed (e.g., counting the number of BUs with a distance of the feature vector less than a threshold). In the selection module the BU with higher coverage is selected and all the BUs with (another) minimum distance from the one selected are discarded (they are already represented). The remainder of BUs are sent to the scoring module for a new rating
    - Adaptive clustering of subsequences after iterative removal of most representative clusters

Abstraction Systems Modelling (IX): Analysis, iterative scoring and selection with metadata feedback



Outline

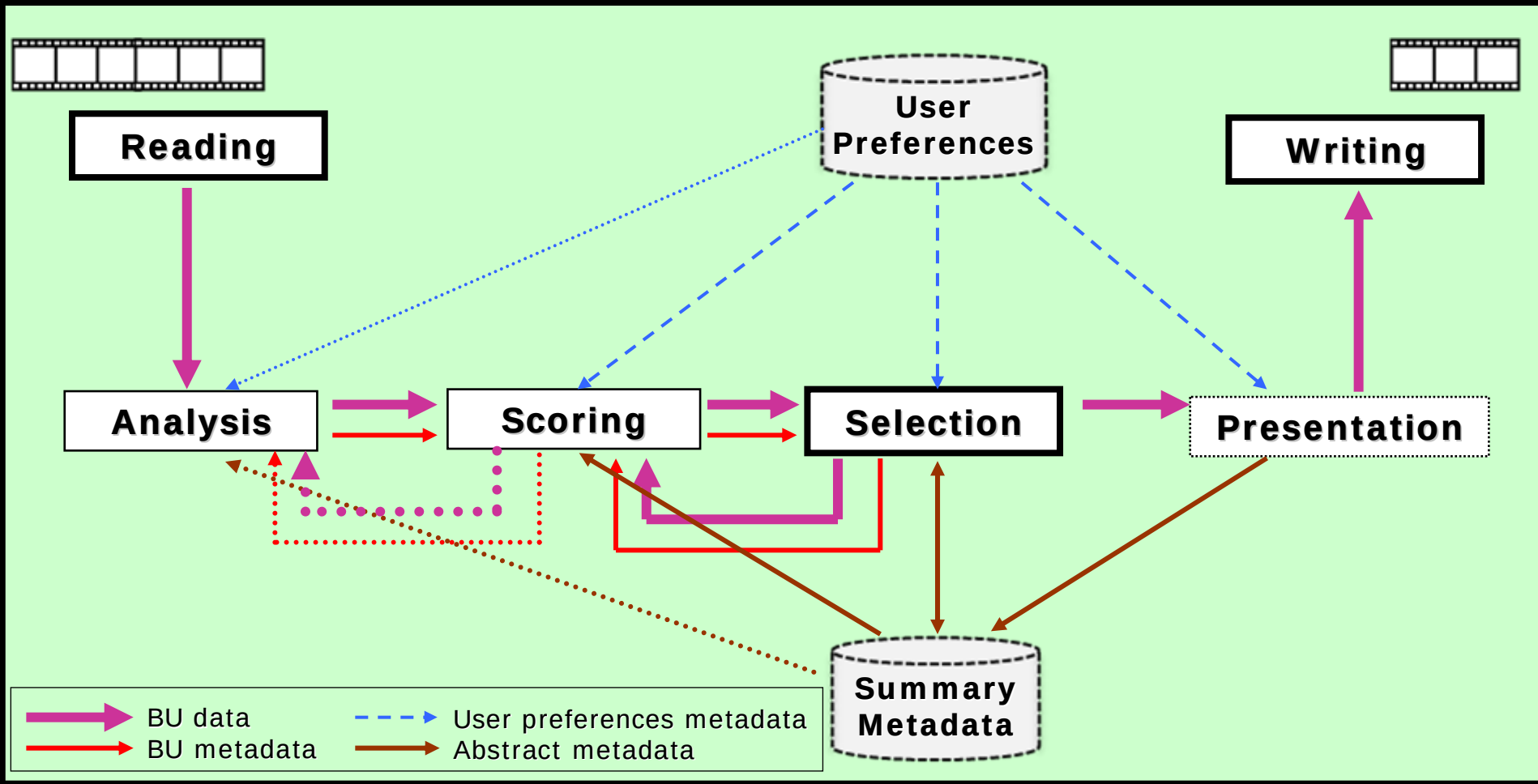
- Introduction
- Simplified Functional Architecture
- Towards a Generic Video Abstraction Architecture
- Abstraction Systems Modelling
- **Generic Video Abstraction Architecture**
- Conclusions

Generic Video Abstraction Architecture (I)

- **As has been seen in the previous progressive modelling each system considered has added additional components to the video abstraction architecture, resulting in a final generic video abstraction architecture**

- **A (secondary) presentation module can be included in order to cover the abstraction approaches that perform some editing or formatting of the video abstract**
  - Video-poster from a set of keyframes, video-in-video, etc.
  - Usually this module has not direct impact in the previous modules, but for generality we propose that they may incorporate user preferences as well as provide metadata to the abstract metadata repository.

Generic Video Abstraction Architecture (II)



Outline

- Introduction
- Simplified Functional Architecture
- Towards a Generic Video Abstraction Architecture
- Abstraction Systems Modelling
- Generic Video Abstraction Architecture
- **Conclusions**

Conclusions (I)

- **The proposed architecture and models allow to categorize existing abstraction systems in order to be able to better understand its pros and contras**
  - Complexity is independent of the classification, as it relies directly in the internal characteristics of the algorithms themselves
  - Categories
    - Not Iterative, Selection
    - Not Iterative, Analysis, Scoring, Selection
    - Not Iterative, Analysis, Scoring, Selection, Metadata feedback
    - Analysis, Iterative Scoring and Selection, Metadata feedback
    - …? Iterative analysis, analysis driven by metadata feedback, …

Conclusions (II)

- **The separation of the abstraction process in "independent stages" allows the generic study of each module and at the same time enables the possibility of developing generic interchangeable modules (once the interfaces are specified) that can be combined in different ways for experimentation.**
  - Divide and conquer for analysis and understanding
  - Modular combination for experimentation and (efficient) new approaches discovery
  - Interfaces to be specified
- **The proposed architecture has allowed to define a set of abstraction system models which can accommodate (almost all of) the existing abstraction approaches in the literature**
  - Additional models may be created for accommodating new future systems starting from the generic architecture
    - The generic architecture may be expanded
      - Backwards compatibility should be assured

ON VIDEO ABSTRACTION SYSTEMS' ARCHITECTURES AND MODELLING

Víctor Valdés, José M. Martínez

Victor.Valdes@uam.es, JoseM.Martinez@uam.es

SAMT 2008, 3-5 December 2008, Koblenz (Germany)

Thanks for your attention!

Universidad Autónoma de Madrid
E28049 Madrid (SPAIN)

Video Processing and Understanding Lab
Grupo de Tratamiento e Interpretación de Vídeo