

5. ビデオデータベース

(マルチメディアデータベース序論, 全6回)

<https://www.kkaneko.jp/de/multimediatdb/index.html>

金子邦彦



ビデオの基本操作

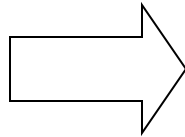


- 再生
- 早送り／巻き戻し
 - 早送り／巻き戻ししながら再生
 - 任意の時点へ移動
- 停止
 - 再生の終了
 - 再生の一時停止

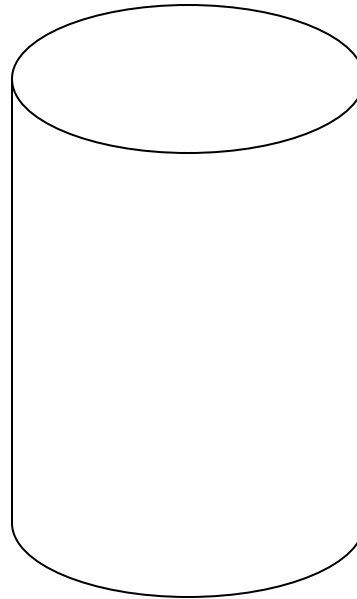
ビデオデータベース



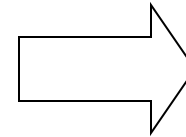
インデックス



データベース
投入



検索



データベース
使用

インデックス付けを行う

検索時に、インデックスを使う

ビデオデータベースの機能



- 索引付けと検索

- | | |
|-----------|------------|
| • 索引 | 検索 |
| • キーワード付け | キーワードによる検索 |
| • 内容情報 | 内容検索 |

- 分類

- 比較

ビデオの「キーワード付け」の課題

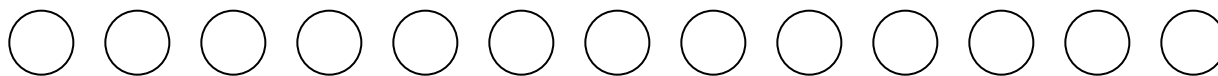
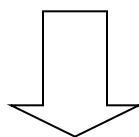


- キーワード付けは困難
 - 人手が必要
 - 主観が入る
 - 不完全
 - 見方が偏る
 - 不正確
 - 時間／費用がかかる

ビデオの「内容検索」の課題



ビデオ



フレーム

- 各フレームを「静止画像」とみなし、色、テキスト、形状などの情報を（自動的に）取り出すことは可能

→ 取り出した情報は冗長
（ビデオの「近い」フレームは中身が似ている）

ビデオの内容情報



- フレーム単独のもの
 - 色 (color)
 - テクスチャ (texture)
 - 形状 (shape)
 - エッジ (edge)

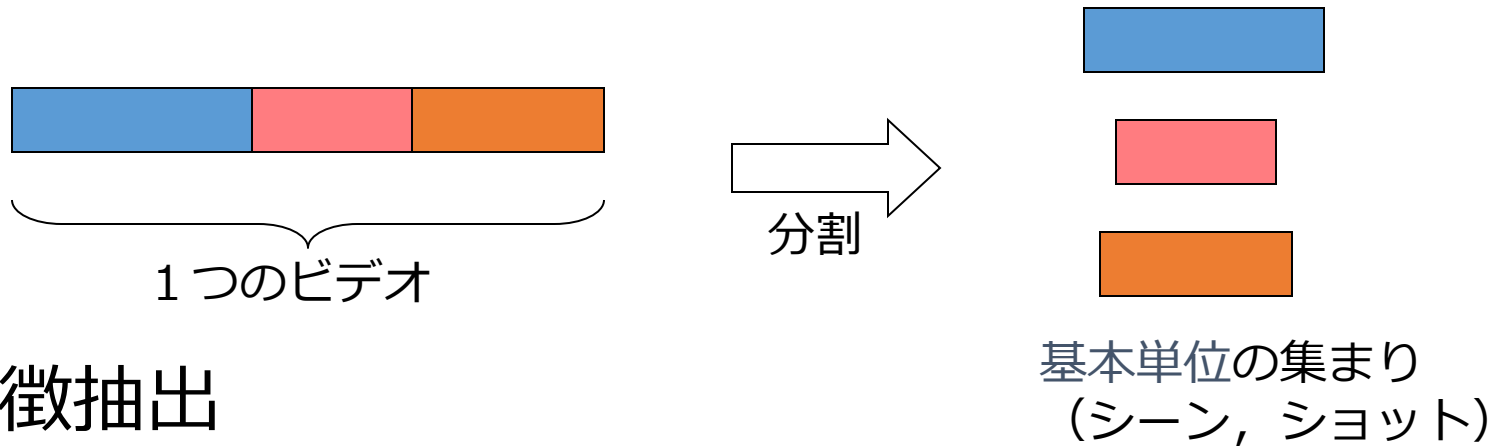
ビデオの内容情報



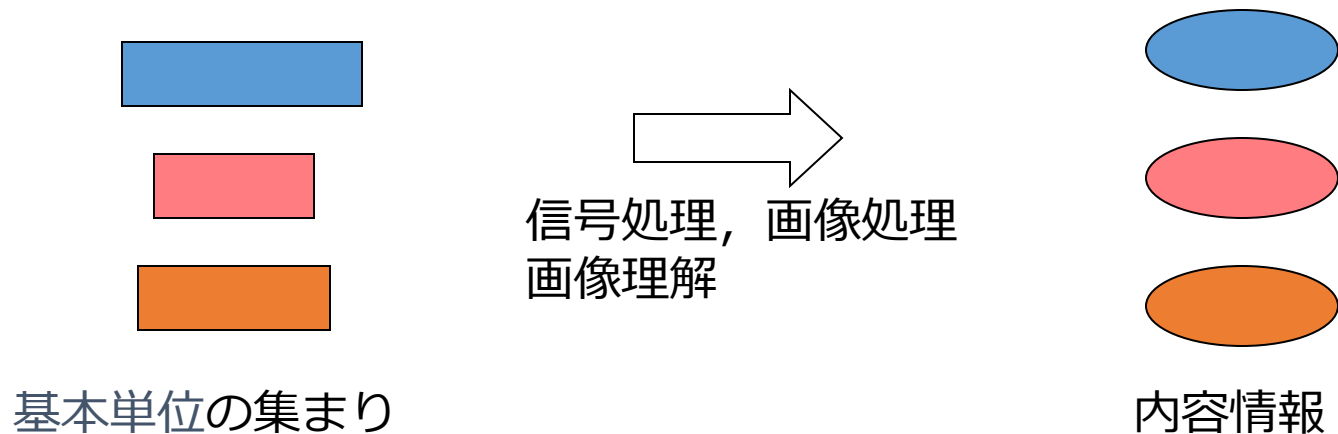
- 空間／時間の要素を含むもの
 - 動き (action)
 - 出来事 (event)
 - story
 - → フレーム単位で, 色, テクスチャ, 形状, エッジの情報を取り出すことは十分でない
 - → シーン, ショットのような, フレームより大きな単位で内容情報を扱う

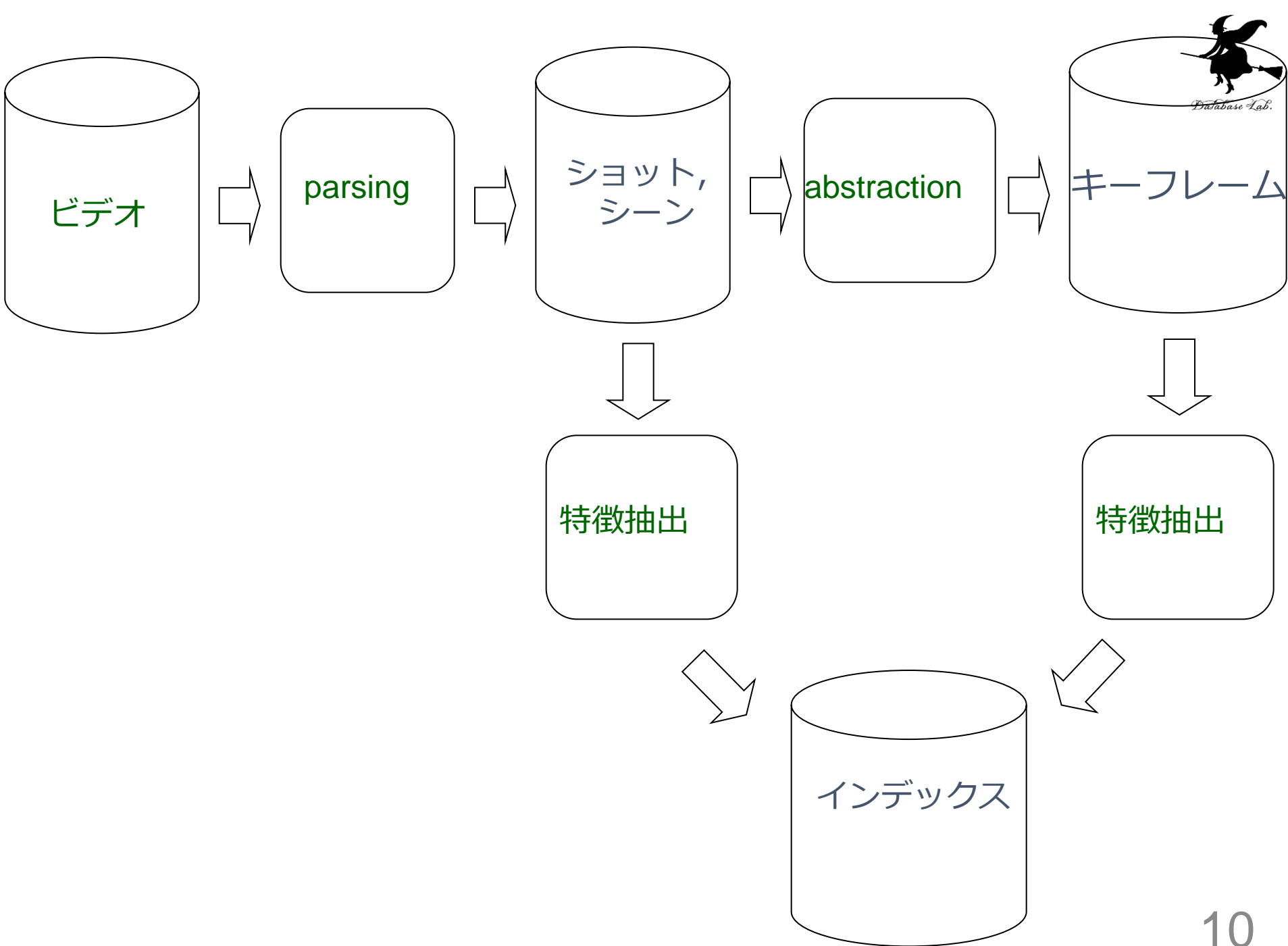
ビデオ内容検索の技術

- Parsing



- 特徴抽出





ショット, シーン

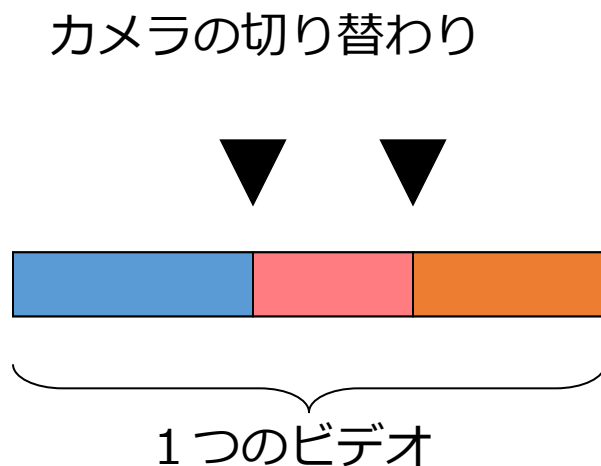


- ビデオを「時間方向」に、より小さい単位に分割した単位
 - ショット： ひと続きのフレーム
 - シーン： 意味的につながりのある、ひと続きのショット

ショット



- 「カメラの切り替わり」を単位とする
- 1つのビデオの中では、頻繁にカメラが切り替わる



カメラ操作



- 1ショット内では, パン, チルト, ズームなどのカメラ操作が行われることがある
- パン (panning) 左右に振る
- チルト (tilting) 上下に傾ける
- ズーム (zooming) 拡大／縮小

シーンを判別する手がかり



- ショットをまたがった「視覚効果」などは、「同一のシーンである」と判定するヒントになる
 - 視覚効果 (transition effect)
 - dissolve, wipe, fade-in, fade-out など
 - 効果音／音楽
 - 「同じ」ショットの繰り返し

カメラの切り替わり



ビデオの構造



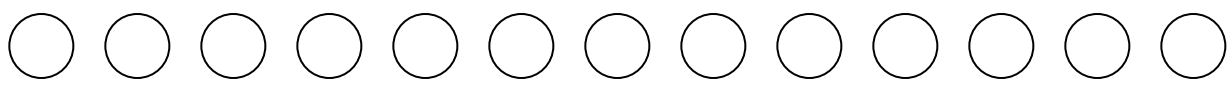
ビデオ



シーン



ショット

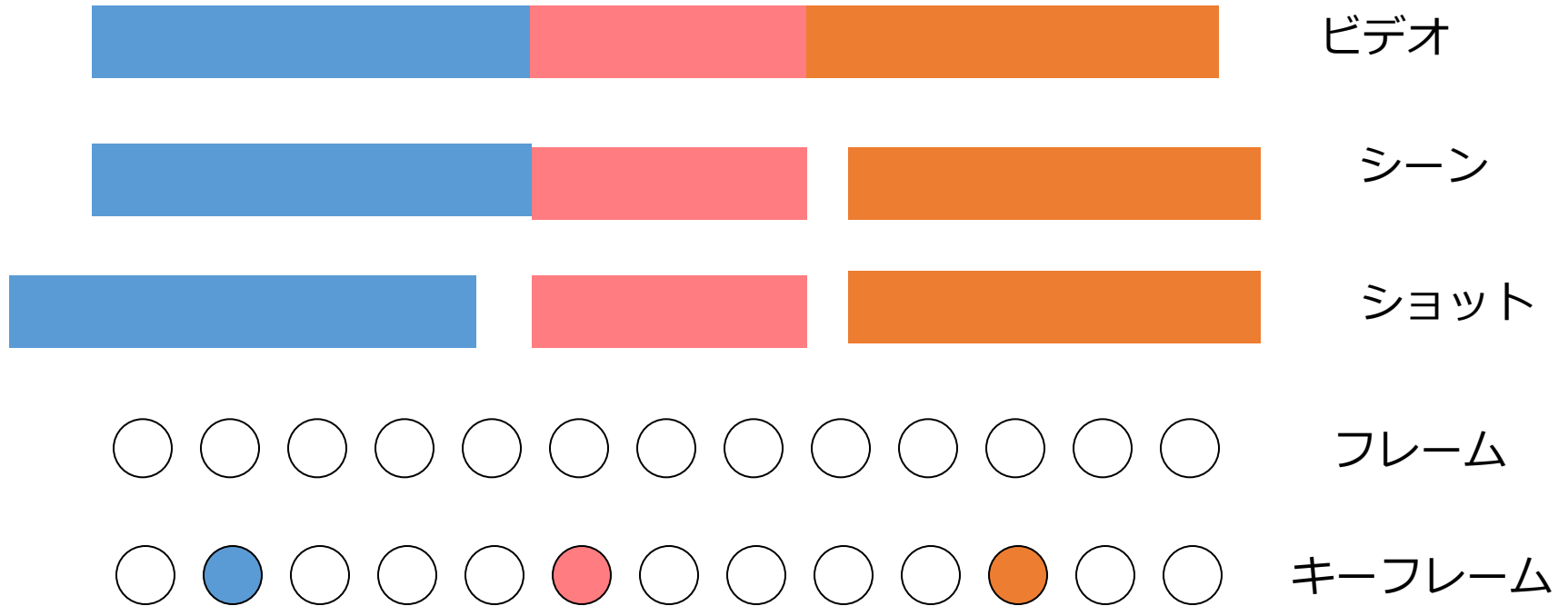


フレーム

キーフレーム



- ショットを代表する静止画像

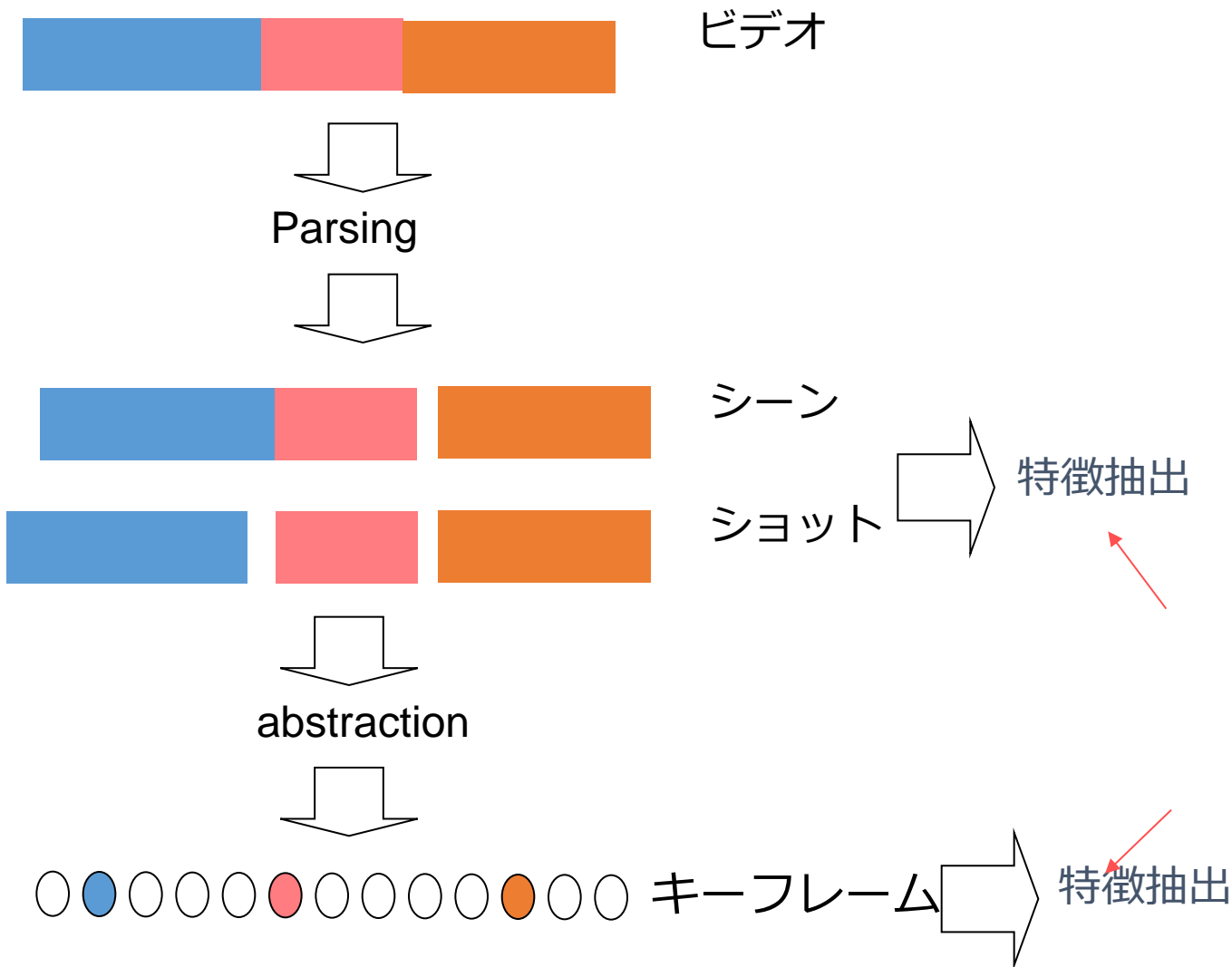


「理想」のキーフレームは？



- 面白い「もの」が写っている
- 面白い「出来事」や「動き」が写っている
- ビデオの中身そのものをよく表現している

特徴抽出までのプロセス



- Parsing
 - 時間方向の分割
- Abstraction
 - キーフレーム取り出し
 - ショットを代表する静止画像（フレーム）の取り出し
- 特徴抽出
 - シーン, ショット, キーフレームに対する特徴抽出
 - 動き, 出来事, ストーリーに関する情報も抽出する

ショットの切れ目



- ショットの切れ目でない
 - カメラ操作（パン, チルト, ズーム）
- ショットの切れ目である
 - カメラの切り替え
 - transition effect （dissolve, wipe, fade-in, fade-out など）

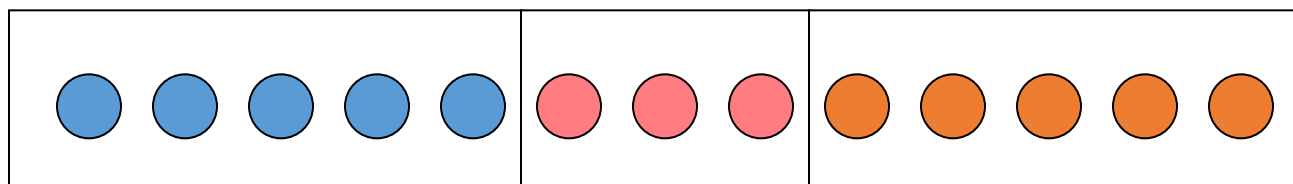
ショット検出法



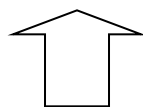
- Pixel-based
 - 画素の「変化量」を計る. ある閾値を超えて変化する画素の数を数える.
- Statistical-based
 - フレームを領域に分割. 各領域の「統計量の変化」を計る（明るさの平均, 分散など）.
- Histogram-based
 - 各フレームごとに「ヒストグラム」（明るさ, 色の分布など）を作り, 2つの連続するフレームのヒストグラムを比較
- Transform-based
 - DCT係数（MPEG データ内などに存在）を使用
- Edge-based
 - エッジの位置と数を比較
- Motion-based
 - ブロックマッチングの結果得られる motion vector を使用

ショット検出法

Pixel-based, Statistical-based, Histogram-based, Transform-based, Edge-based では



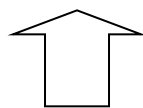
ショット検出
結果



隣り合う2フレームで比較. 閾値を
超えたら「切れ目」と判定



フレーム単位で
情報を取り出し



キーフレーム抽出法



- 特定フレームを使用
 - ショットの先頭
 - ショットの末尾
 - ショットの間
- visual content complexity indicators
- shot activity indicators
- shot motion indicators

キーフレーム抽出で 사용되는情報



- Color
 - フレームごとの色の分布
 - 明るさの平均
 - 色のばらつき具合
 - 主に登場する色の種類
- Motion
 - カメラ操作
 - 登場物の動き
- など

- ビデオの内容検索

- 特徴抽出の技術が必要（自動化したい）

- ビデオデータに含まれる「音声」, 「音」, 「テキスト」, 「顔の認識」, 「カメラの動き」, 「登場物の動き」の情報をすべて使うことが鍵になる