



Методы сопоставления видео

Глазистов Иван

*CS MSU Graphics & Media Lab
(Video Group)*



Оглавление

➤ Введение

- Где используются методы сопоставления видео?
- Основные подходы к описанию видео
- Алгоритмы сопоставления
- Заключение, планы

Где используются методы сопоставления видео?

- ❑ Поиск дубликатов в базе фильмов
- ❑ Генерация сценариев для быстрого просмотра
- ❑ Поиск фрагментов в базе



Оглавление

- ❑ Введение
- **Алгоритмы сопоставления**
 - Алгоритм GMLabVideoGroup
 - ❑ Сопоставление строк
 - ❑ Иерархический метод
 - ❑ Поиск оптимального сопоставления в графе
- ❑ Заключение, планы



Алгоритм GMLabVideoGroup

Основные этапы:

1. Описание видео:
 - a. Разбиение видео на сцены
 - b. Построение дерева сцен
2. Сравнение фильмов

Разбиение на сцены

1. Каждый кадр описываемой последовательности разбивается на блоки 32x32 пиксела
2. Для каждого блока строится гистограмма распределения яркости

$$H(x, y, t) = \{h_k(x, y), k = \overline{1, m}\},$$

где (x, y) — координата блока в кадре,

$m = 64$ — количество уровней яркости в гистограмме.

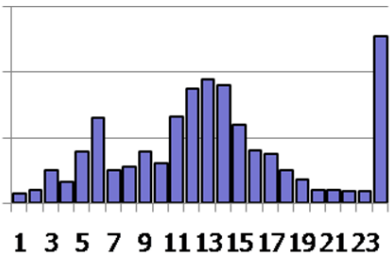
Разбиение на сцены

3. Вычисляется мера разности соседних кадров

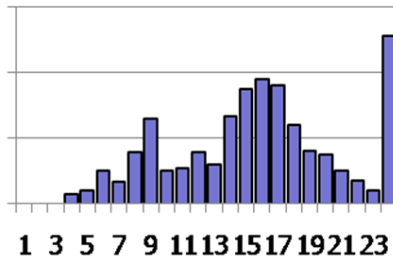
$$dH(x, y, t + 1) = \{h_k(x, y, t + 1) - h_k(x, y, t), k = \overline{1, m}\},$$

$$Diff(t + 1) = \sum_{x, y} S(G(dH(t + 1))), \text{ где } G(\cdot) - \text{низкочастотный фильтр, } S(H) = \sum_{k=1}^m |h_k|.$$

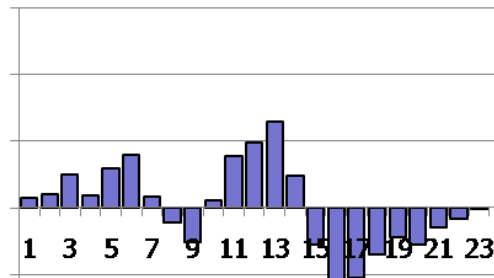
$H(t)$:



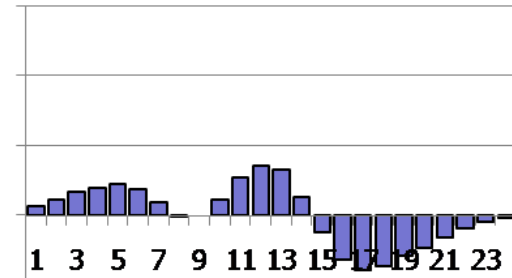
$H(t + 1)$:



$dH(t + 1)$:



$G(dH(t + 1))$:



Разбиение на сцены

4. Для более устойчивой работы в условиях стабильно высоких разностей кадров массив обрабатывается высокочастотным фильтром:

$$F = \{-1, -1, -1, -1, 8, -1, -1, -1, -1\}.$$

5. Определение смен сцены:

Если $F(\text{Diff}(t_i)) > Th$, тогда t_i — положение смены сцены.

Th — некоторый порог.

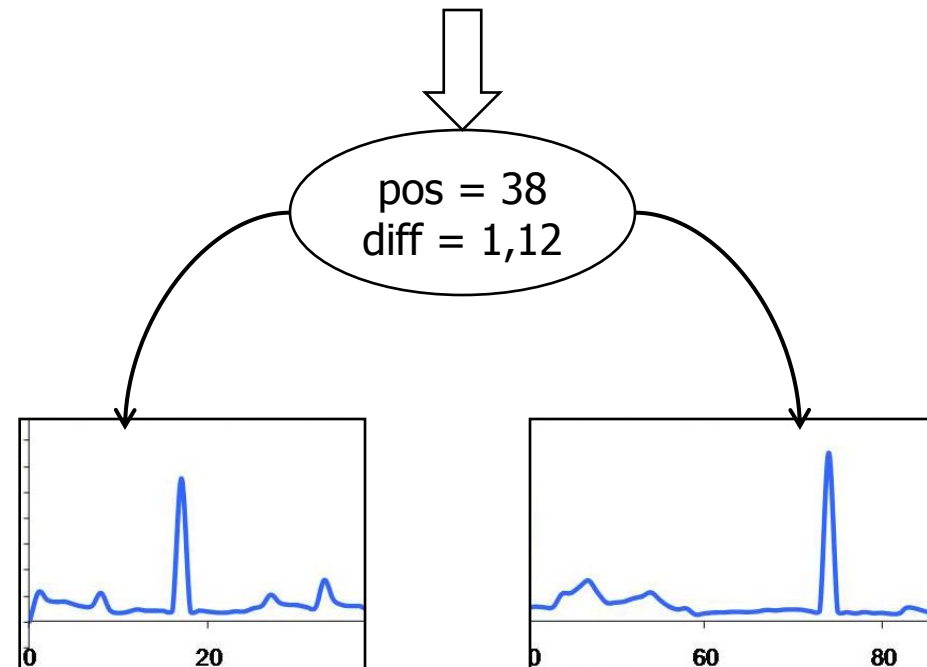
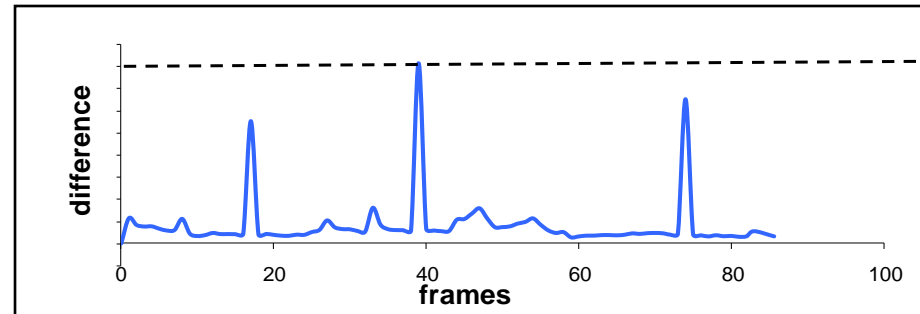
При построении дерева учитываются только разности, удовлетворяющие этому условию.

Алгоритм GMLabVideoGroup



Построение дерева сцен

1. Находится максимальное значение в массиве разностей
2. Строится вершина дерева
3. Для правой и левой частей массива применяется п.1



Алгоритм GMLabVideoGroup



Сравнение фильмов

Основные этапы:

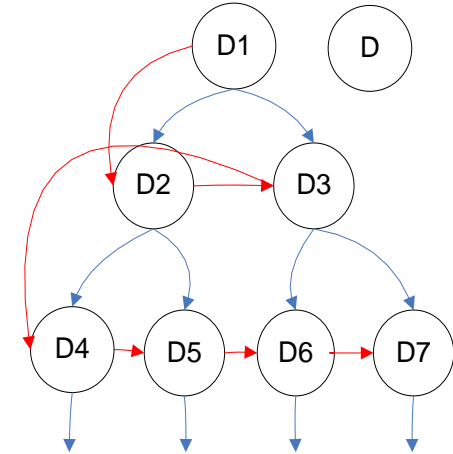
1. Позиционирование короткого фильма более длинном
2. Построение оценки разности
3. Выбор наилучшего сопоставления
4. Построение дерева наилучшего кандидата
5. Обратный поиск
6. Усреднение оценок прямого и обратного поиска

Сравнение фильмов

Позиционирование

1. Позиционирование короткого фильма более длинном

- a) Выбирается доминирующая сцена сцены
- b) Составляется вектор сцен кандидатов (обход в ширину)



Если $|D - D_i| \leq Th$, то сцена помечается как кандидат

- c) Для каждой сцены кандидата координаты смен сцен сопоставляются и выполняется п.2

Сравнение фильмов

Оценка разности

2. Построение оценки разности

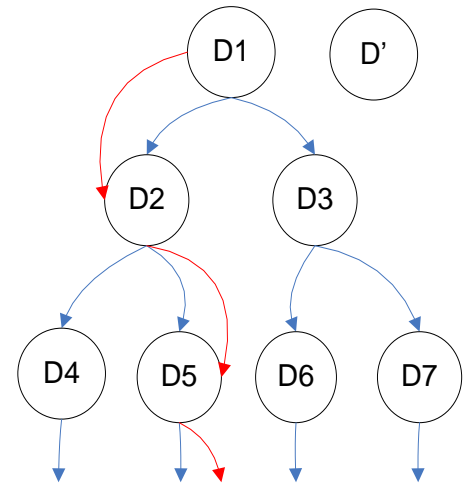
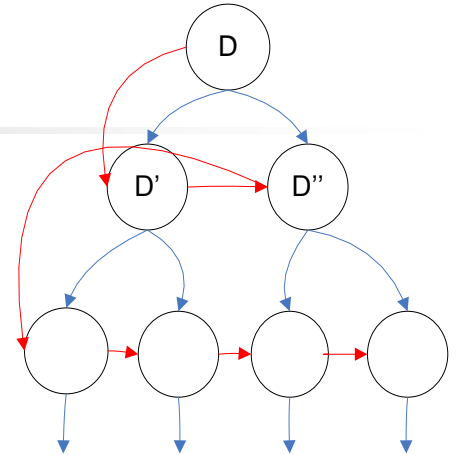
- Находится смещение дочерней по отношению к доминирующей сменой сцены (D') в длинном фильме
- Производится поиск близких по координате и силе смен сцены, начиная с корня
- Выбирается лучший кандидат по следующей мере:

$$Diff(D', D_i) = (D' - D_i)^2 + \alpha(t' - t_i)^2,$$

где t' — предположительное положение D' .

t_i — положение

α — коэффициент, выравнивающий влияние положения и силы вершины кандидата

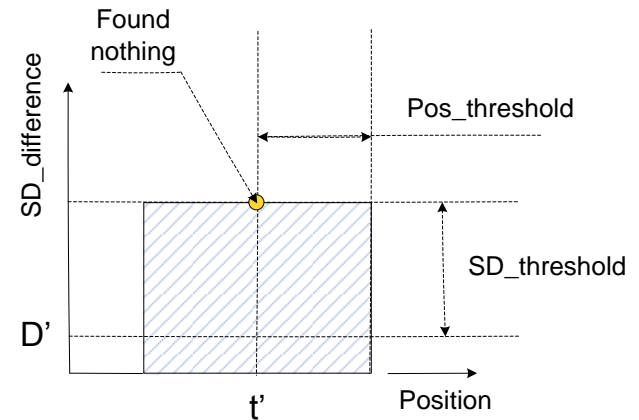
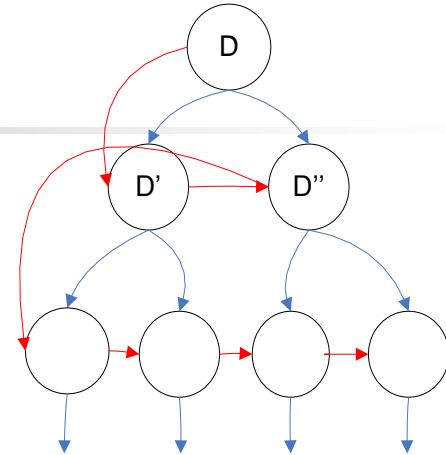


Сравнение фильмов

Оценка разности

- d) Та же самая процедура повторяется для всех вершин короткого фильма, глубина которых < 5
- e) Если не найдено близких вершин полагаем, что искомая вершина лежит на границе
- f) Если найдено < 3 вершин, то полагаем, что поиск неуспешен
- g) Вычисляем разность:

$$Diff(D, D_1) = \sum_k Diff(D'_k, D_{ik}).$$



Сравнение фильмов

Обратный поиск

3. Выбор наилучшего сопоставления
4. Построение дерева наилучшего кандидата
 - a) «Разбор» дерева длинного фильма
 - b) Получения массива смен сцены
 - c) Построение поддерева
5. Обратный поиск
6. Усреднение оценок прямого и обратного поиска

Оценка разности окончательно:

$$Diff(V_i, V_j) = (\min_k (Diff(D, D_k)) + \min_s (CrossDiff(D_1, D'_s))) / 2.$$

Алгоритм GMLabVideoGroup



Характеристики алгоритма

Время построения индекса: $O(l)$, l – длина фильма.

Скорость поиска: $O(m \log(m))$, $m = \max(p, q)$,

p, q – количества сцен в сравниваемых фрагментах

Ограничения:

Корректно работает только с достаточно большими видео, не справляется с перемонтированным видео

Алгоритм GMLabVideoGroup



Результаты

Алгоритм тестировался на базе из ~900 фильмов, разбитых на 5 групп.

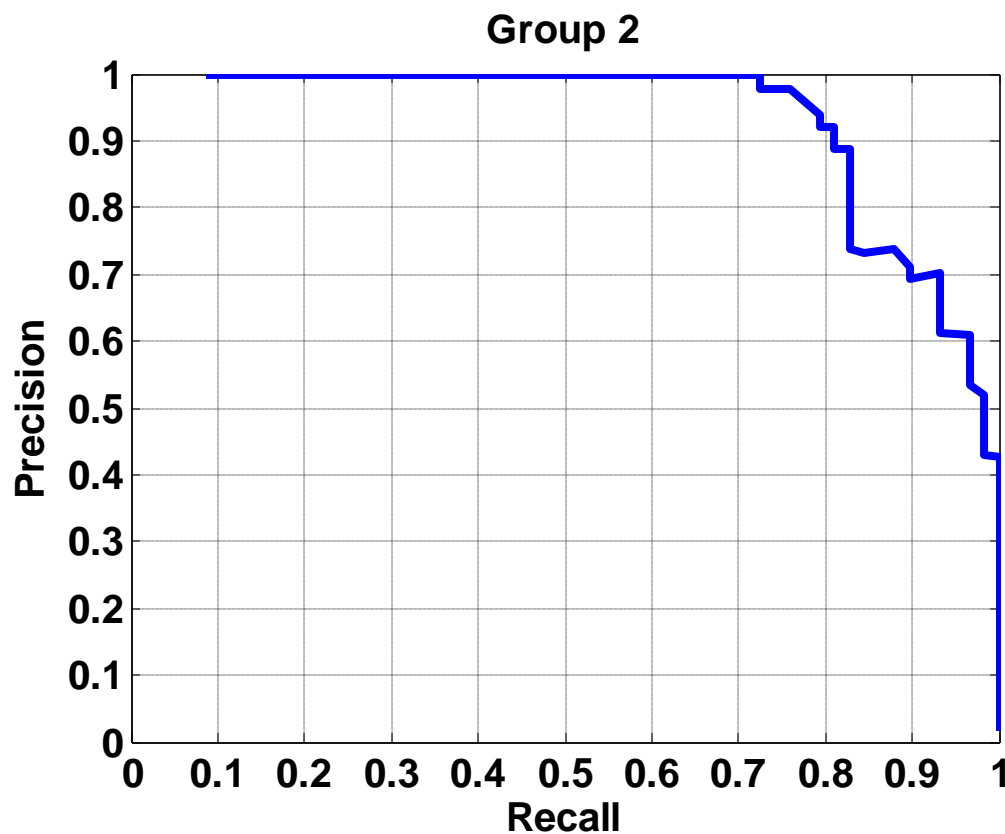
Размер групп фильмов и их описаний

Номер группы	Размер группы (Mb)	Размер индекса (Mb)	Размер после сжатия (Mb)
1	7170	4,6	0,86
2	870	1,0	0,18
3	2690	0,9	0,18
4	3050	2,5	0,43
5	480	0,74	0,05
Объединенная группа	13900	9,74	1,7

Алгоритм GMLabVideoGroup



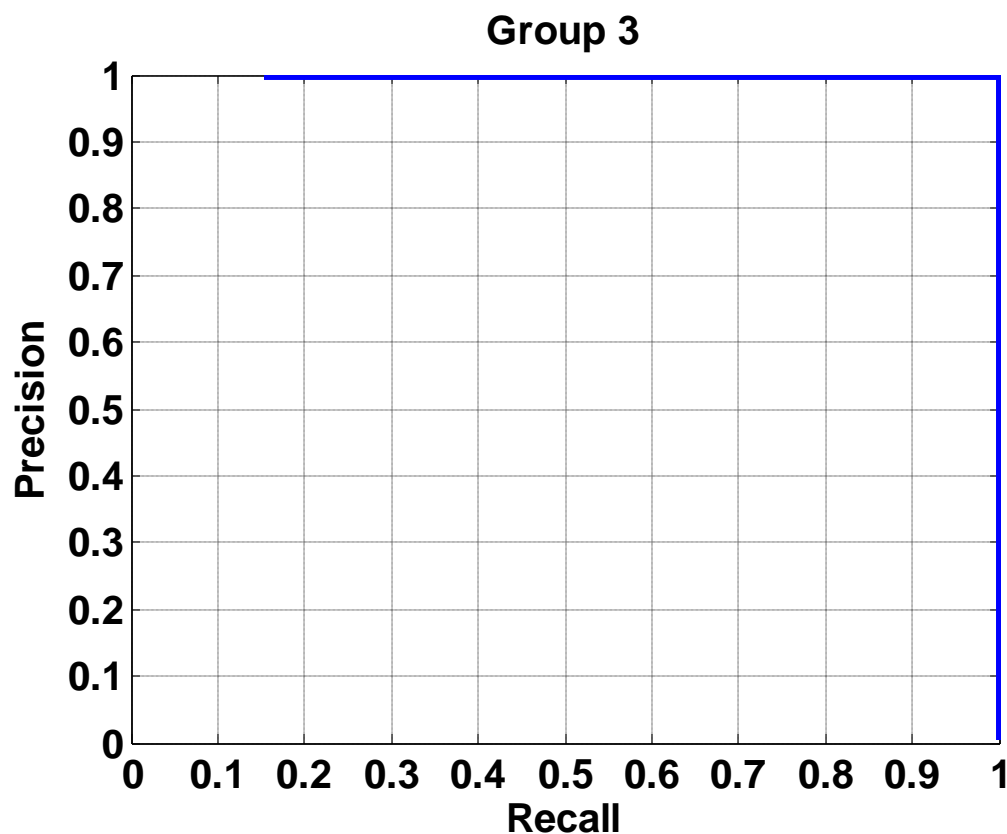
Результаты



Алгоритм GMLabVideoGroup



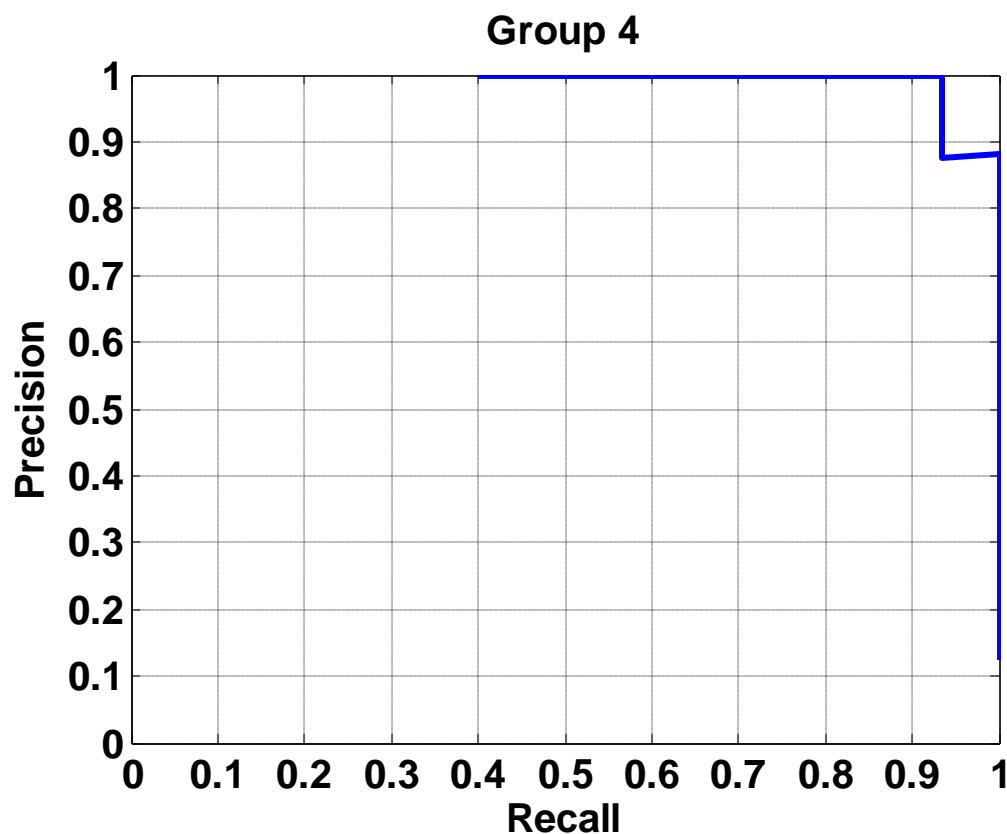
Результаты



Алгоритм GMLabVideoGroup

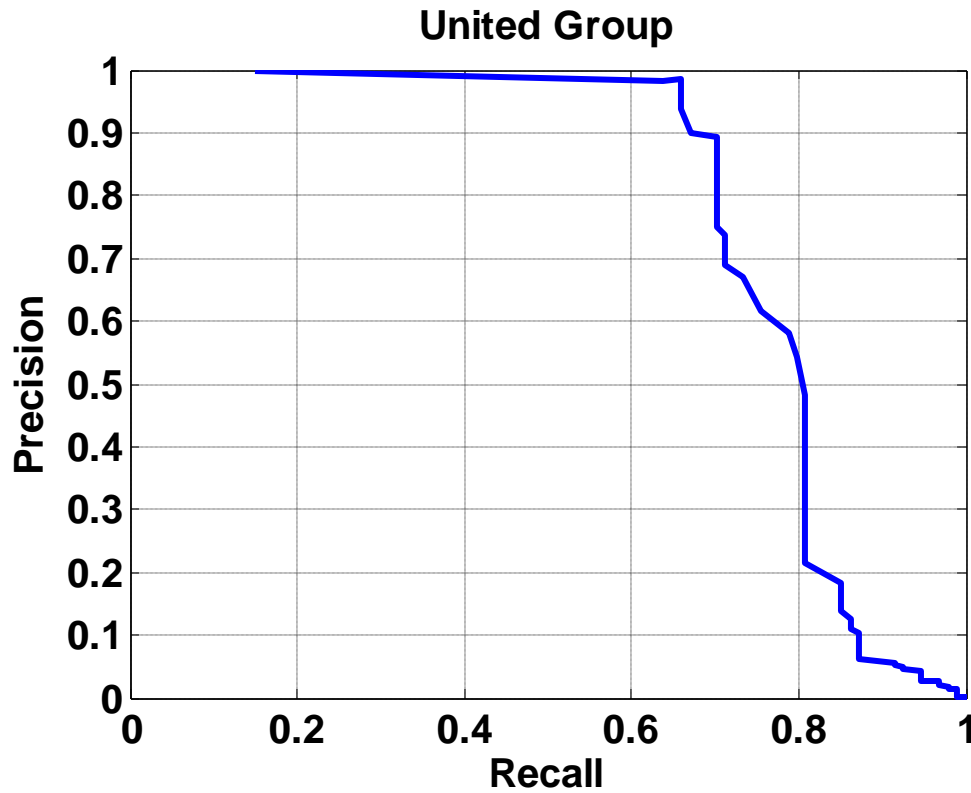


Результаты



Алгоритм GMLabVideoGroup

Результаты





Оглавление

- ❑ Введение
- **Алгоритмы сопоставления**
 - ❑ Алгоритм GMLabVideoGroup
 - Сопоставление строк
 - ❑ Иерархический метод
 - ❑ Поиск оптимального сопоставления в графе
- ❑ Заключение, планы



Сопоставление строк

Основная идея:

Представить видеопоследовательность в виде строки символов и свести задачу сопоставления видео к задаче сопоставления строк

Основная проблема:

Грамотный выбор весовой функции для сопоставления строк

"Detection of video sequences using compact signatures ", 23
T.C. Hoad and J. Zobel, ACM Transactions on Information
Systems (TOIS), 24(1):1-50, January 2006.



Сопоставление строк

Предложено четыре метода:

1. Представление фильма в виде массива длин сцен
2. Видео - массив покадровых изменений яркости
3. Сохранение длины вектора глобального движения от кадра к кадру
4. Комбинированный метод

У всех методов есть общая часть – алгоритм сопоставления строк

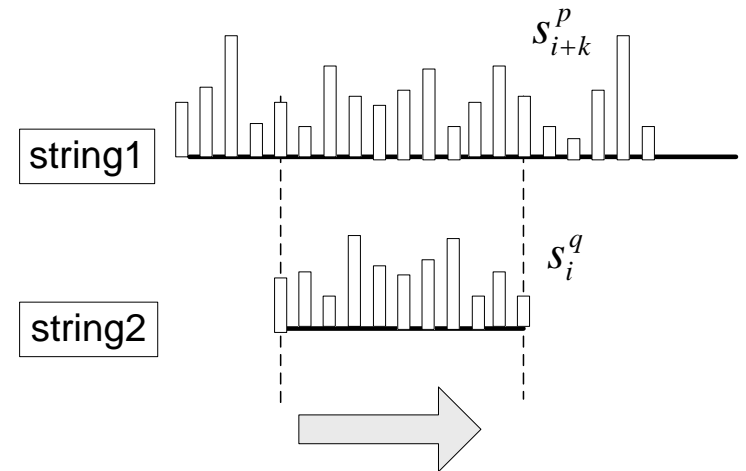
"Detection of video sequences using compact signatures ", 24
T.C. Hoad and J. Zobel, ACM Transactions on Information
Systems (TOIS), 24(1):1-50, January 2006.

Сопоставление строк

Стратегия поиска

Вычисление меры сходства:

$$Sim(V_q, V_p) = \max_k \left(\sum_{i=1}^n Sc(s_i^q, s_{i+k}^p) \right)$$



На каждом шаге алгоритма окно поиска сдвигается на 1 кадр

Сопоставление строк

Весовая функция

Обозначение: $\delta = k(s_i^q - s_{i+k}^p)$,

1. Бинарная весовая функция

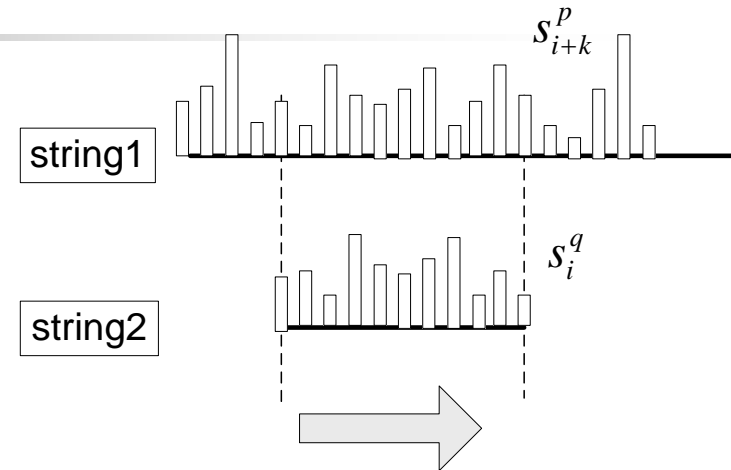
$$Sc(s_i^q, s_{i+k}^p) = \begin{cases} 20, & s_i^q = s_{i+k}^p \\ -5, & \text{иначе} \end{cases}$$

2. Линейная весовая функция

$$Sc(s_i^q, s_{i+k}^p) = 20 - \frac{3}{2} \delta,$$

3. Categorical scoring

$$Sc(s_i^q, s_{i+k}^p) = \begin{cases} 13, & \delta = 0 \\ 8, & \delta = 1, 2 \\ -3, & \delta = 3, 4 \\ -7, & \delta > 4 \end{cases}$$



Сопоставление строк

Весовая функция

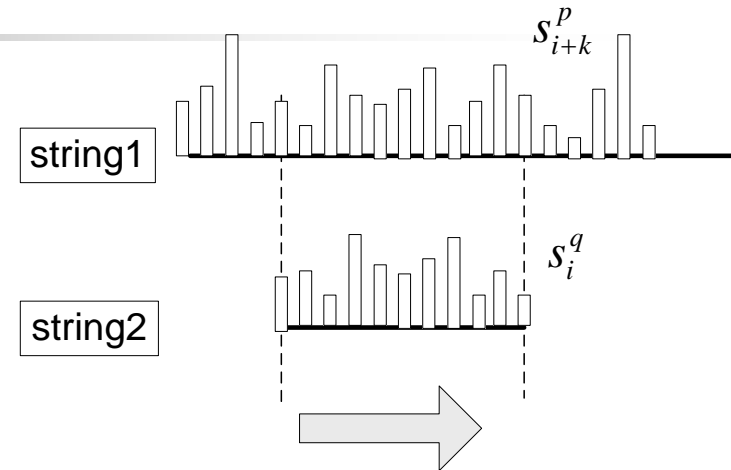
Обозначение: $\delta = k(s_i^q - s_j^p)$,

4. Кубическая весовая функция

$$Sc(s_i^q, s_{i+k}^p) = \frac{20 - (\delta - 10)^3}{50},$$

5. Mean-weighted scoring

$$Sc(s_i^q, s_{i+k}^p) = \frac{2(2 - \delta)}{m_q}, \quad m_q = \sum_{i=1}^n s_i^q.$$



Сопоставление строк

Длина сцены

Алгоритм:

1. Построение для каждого кадра 48-уровневой гистограммы распределения яркости

2. Сравнение гистограмм: $dH(t+1) = \{h_k(t+1) - h_k(t), k = \overline{1, m}\},$

$$Diff(t+1) = S(dH(t+1)) = \sum_{k=1}^m |h_k(t+1) - h_k(t)|.$$

3. Применение адаптивного порога (не описан)

4. Получение описание фильма вектором длин сцен

"Detection of video sequences using compact signatures ", 28
T.C. Hoad and J. Zobel, ACM Transactions on Information
Systems (TOIS), 24(1):1-50, January 2006.

Сопоставление строк

Длина сцены

❑ Недостатки:

- a) «Грубое сопоставление» фрагментов
- b) Применимо только достаточно «длинных» фильмов
- c) Более чувствителен к ошибкам детектора сцен, чем GMLabVideoGroup

❑ Достоинства:

- a) Стабильная работа с искаженным видео
- b) Короткое описание фильма ($\sim 5\text{kb/h}$)
- c) Быстрый поиск ($O(p(q-p))$)

"Detection of video sequences using compact signatures ",
T.C. Hoag and J. Zobel, ACM Transactions on Information
Systems (TOIS), 24(1):1-50, January 2006.

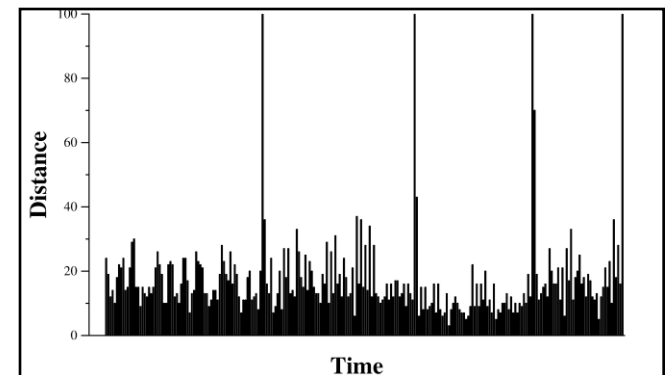
Сопоставление строк

Изменения яркости (Color-shift Signature)

Алгоритм:

- Строится гистограмма распределения яркости для каждого кадра
- Сохраняется разность гистограмм соседних кадров

Frame no.	Difference histogram			Manhattan Distance
	Y	Cr	Cb	
0	[12, 4, 8, 18]	[3, 19, 16, 6]	[17, 14, 10, 1]	128
1	[2, 4, 2, 4]	[3, 1, 6, 2]	[5, 0, 2, 3]	43
2	[1, 1, 1, 3]	[1, 2, 1, 2]	[2, 1, 1, 0]	16
3	[2, 1, 2, 1]	[0, 2, 0, 2]	[1, 0, 2, 1]	14
4	[13, 3, 18, 2]	[6, 8, 1, 1]	[6, 2, 2, 6]	68
5	[0, 2, 3, 1]	[3, 2, 3, 2]	[1, 1, 1, 1]	20



"Detection of video sequences using compact signatures ",
T.C. Hoad and J. Zobel, ACM Transactions on Information
Systems (TOIS), 24(1):1-50, January 2006.

Сопоставление строк

Информация о движении (Centroid Signature)

2. Информация о движении

- Находятся средние координаты 5% самых светлых/темных пикселей
- Сохраняется евклидова разность векторов соседних кадров



"Detection of video sequences using compact signatures ",
T.C. Hoad and J. Zobel, ACM Transactions on Information
Systems (TOIS), 24(1):1-50, January 2006.

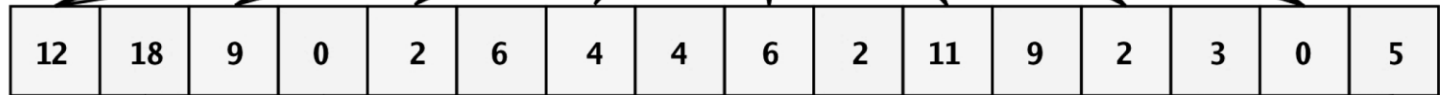
Сопоставление строк

Комбинированный метод

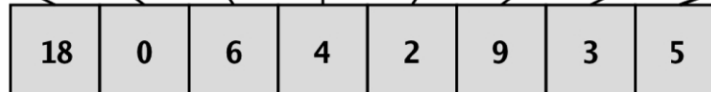
Color-shift Signature



Combined Signature



Centroid Signature



"Detection of video sequences using compact signatures ",
T.C. Hoad and J. Zobel, ACM Transactions on Information
Systems (TOIS), 24(1):1-50, January 2006.

Сопоставление строк

Методы покадрового описания видео

❑ Недостатки:

- a) Чувствительность к искажениям видео
- b) Длинное описание фильма
- c) Медленный поиск

❑ Достоинства:

- a) Возможность поиска коротких фрагментов
- b) Достоверность поиска

Сопоставление строк

Результаты

Тестовый набор:

База данных: 170 минут (1,9 Gb) broadcast-видео.

Работа проверялась на 18 запросах, имеющих в базе много дубликатов

1. Тестирование весовых функций

Scoring Method	Color Shift		Centroid		Shot Length	
	$p(n)$	$r(20)$	$p(n)$	$r(20)$	$p(n)$	$r(20)$
Binary	0.19	0.30	0.29	0.47	0.60	0.75
Categorical	0.76	0.79	0.61	0.71	0.66	0.77
Linear	0.19	0.30	0.29	0.47	0.60	0.75
Cubic	0.59	0.71	0.52	0.65	0.60	0.75
Mean-weighted	0.66	0.71	0.65	0.74	0.59	0.67

Сопоставление строк

Результаты

2. Тестирование indecision-points

Recall(20) at Indecision Point (cubic scoring)

	3	6	9	12	15	18	25	50	75	100
Color shift	0.44	0.13	0.02	0.82	0.05	0.00	0.00	0.00	0.00	0.00
Centroid	0.51	0.05	0.02	0.61	0.00	0.00	0.00	0.00	0.00	0.00
Shot length	0.00	0.00	0.00	0.75	0.00	0.00	0.00	0.00	0.02	0.02

Recall(20) at Indecision Point (mean-weighted scoring)

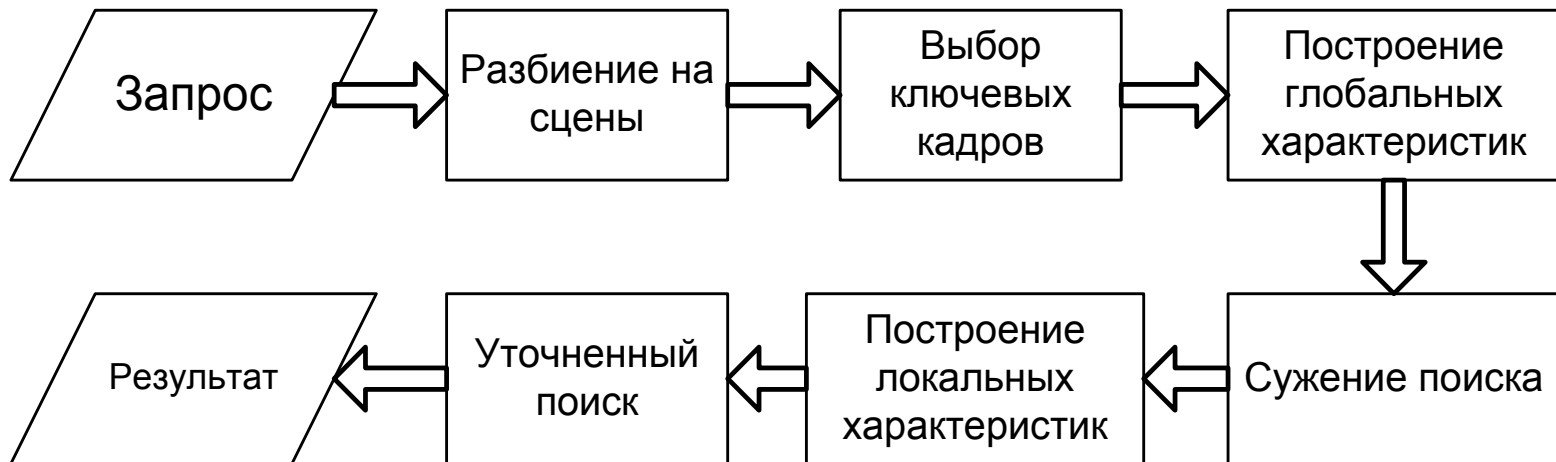
	3	6	9	12	15	18	25	50	75	90
Color shift	0.32	0.94	0.96	0.96	0.96	0.94	0.93	0.91	0.83	0.80
Centroid	0.34	0.69	0.84	0.84	0.83	0.83	0.80	0.77	0.65	0.64
Shot length	0.63	0.67	0.67	0.67	0.67	0.67	0.63	0.62	0.68	0.62



Оглавление

- ❑ Введение
- **Алгоритмы сопоставления**
 - ❑ Алгоритм GMLabVideoGroup
 - ❑ Сопоставление строк
 - Иерархический метод
 - ❑ Поиск оптимального сопоставления в графе
- ❑ Заключение, планы

Иерархический метод



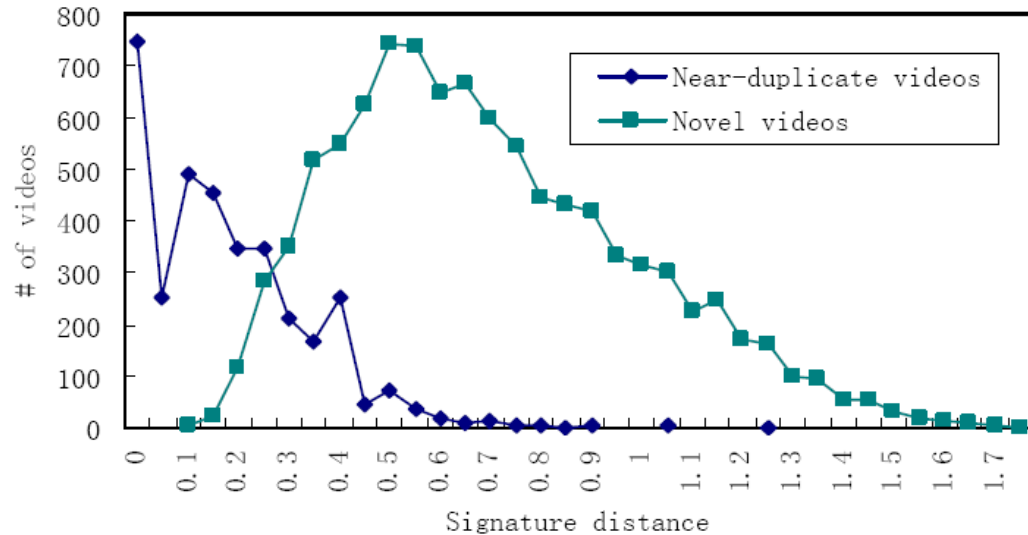
Xiao Wu, Alexander G. Hauptmann and Chong-Wah Ngo
"Practical Elimination of Near-Duplicates from Web Video Search"
ACM International Conference on Multimedia, Sep. 2007

Иерархический метод

Построение глобальных характеристик

$$VS = (s_1, s_2, \dots, s_m), s_i = \frac{1}{n} \sum_{j=1}^n h_{ij},$$

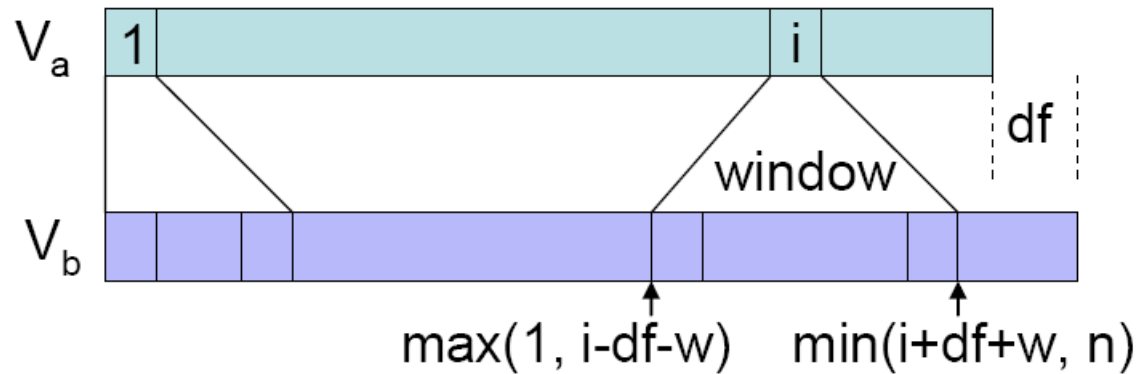
$$R(V_i | V_j) = d(VS_i, VS_j) = \sqrt{\sum_{k=1}^m (x_k - y_k)^2}.$$



Xiao Wu, Alexander G. Hauptmann and Chong-Wah Ngo
 Practical Elimination of Near-Duplicates from Web Video Search
ACM International Conference on Multimedia, Sep. 2007

Иерархический метод

Уточненный поиск



V_a – фильм-запрос,

V_b – фильм из базы,

n – длина фильма в базе,

df – разность длин двух видео,

w – размер окна поиска.

Иерархический метод

Тестирование

- ❑ Размер базы 12970 фильмов
- ❑ База построена на основе 24 YouTube-запросов
- ❑ При тестировании для каждого фильма-запроса строились графики Recall/Precision для разных значений порога сходства фильмов

Precision – процент неправильно определенных пар.

Recall – процент определенных правильно фильмов от общего количества похожих фильмов в группе в группе

Xiao Wu, Alexander G. Hauptmann and Chong-Wah Ngo
Practical Elimination of Near-Duplicates from Web Video Search
ACM International Conference on Multimedia, Sep. 2007

Иерархический метод

Выводы

❑ Недостатки:

- a) Нестабильность при работе с искаженным видео
- b) Может искать только близкие по длине фрагменты
- c) Медленный поиск

❑ Достоинства:

- a) В теории может искать незначительно перемонтированные фрагменты



Оглавление

- ❑ Введение
- **Алгоритмы сопоставления**
 - ❑ Алгоритм GMLabVideoGroup
 - ❑ Сопоставление строк
 - ❑ Иерархический метод
 - Поиск оптимального сопоставления в графе
- ❑ Заключение, планы

Поиск оптимального сопоставления в графе

Алгоритм:

1. Разбиение на сцены
2. Выделение ключевых кадров на основе анализа движения
3. Для каждого ключевого кадра строится гистограмма распределения яркости

«Clip-based Similarity Measure for Hierarchical Video Retrieval», Yuxin Peng, *hong-Wah Ngo*, MIR'04, October 15–16, 2004, New York, New York, USA

Поиск оптимального сопоставления в графе

4. Строится мера сходства сцен: $Sim(s_i, s_j) = \frac{1}{2} \{ \phi(s_i, s_j) + \hat{\phi}(s_i, s_j) \}$

$$\phi(s_i, s_j) = \max_{p=\{1,2,\dots\}, q=\{1,2,\dots\}} Intersect(r_{ip}, r_{jq}),$$

$\hat{\phi}(s_i, s_j)$ – второе по величине значение этого максимума

5. Строится двудольный граф сопоставления двух фильмов

$X = \{x_1, x_2, \dots, x_p\}, Y = \{y_1, y_2, \dots, y_q\}$ – описания двух фильмов

$G(X, Y, E)$ – двудольный граф, $E = \{(x_i, y_j) \mid Sim(x_i, y_j) > Th\}$

«Clip-based Similarity Measure
for Hierarchical Video Retrieval», Yuxin Peng, *hong-Wah Ngo*, MIR'04,
October 15–16, 2004, New York, New York, USA

Поиск оптимального сопоставления в графе

6. Выделение однозначного двудольного подграфа

1) $M = \{\}$

2) Если перебраны все вершины X , то в M содержится искомый граф

3) Пусть $x_i \in X$ - ещё не рассмотренная вершина X . Тогда

$A = \{x_i\}, B = \{\}$ - два разных множества.

4) Пусть $N(A) \in Y$ - множество вершин, в которые можно попасть из вершин в A . Если $N(A) = B$, то помечаем вершину x_i как обработанную и переходим на п.2, иначе на п.5.

5) Находим вершину $y_j \in N(A) \setminus B$.

«Clip-based Similarity Measure

for Hierarchical Video Retrieval», Yuxin Peng, hong-Wah Ngo, MIR'04,

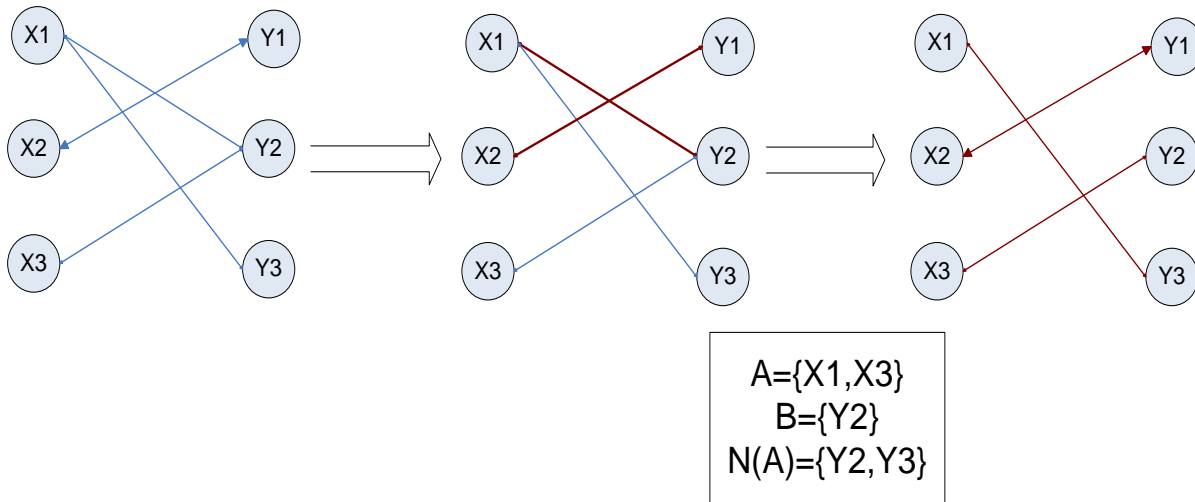
October 15-16, 2004, New York, New York, USA

Поиск оптимального сопоставления в графе

6) Если $(z, y_j) \in M$, то $A = A \cup \{z\}, B = B \cup \{y_j\}$

и переходим на п.4, иначе на п.7

7) $M = M \oplus P(x_i, y_j)$, отмечаем x_i и переходим на п.2.



Алгоритм Optimal Matching

7. Построение оптимального графа

1) Инициализация меток вершин

$$l(x_i) = \max(w_{ij}), i = \overline{1, p};$$

$$l(y_j) = 0, j = \overline{1, q}$$

2) Построение максимального графа

$$G_l = (X, Y, E), E = \{(x_i, y_j) \mid l(x_i) + l(y_j) = w_{ij}\}$$

И однозначного графа M на основе G_l

3) Если M содержит все вершины X , то оптимальный граф построен, иначе на п.4

4) Выбираем $x_i \in X, x_i \notin M, A = \{x_i\}, B = \{\}$

Алгоритм Optimal Matching

5) Если $N_{G_i}(A) = B$, то идем на п.9, иначе на п.6

6) Находим $y_j \in N_{G_i}(A) \setminus B$.

7) Если $(z, y_j) \in M$, то $A = A \cup \{z\}, B = B \cup \{y_j\}$

и переходим на п.5, иначе на п.8

8) $M = M \oplus P(x_i, y_j)$ и переходим на п.3.

9) Вычисляем $a = \min_{x_i \in A, y_j \notin N_{G_i}} \{l(x_i) + l(y_j) - w_{ij}\}$,

$$\hat{l}(v) = \begin{cases} l(v) - a, v \in A \\ l(v) + a, v \in B, \text{ вычисляем на основе } \hat{l} \text{ } G_{\hat{l}}, E_{\hat{l}}. \\ l(v), \text{ иначе} \end{cases}$$

10) $l = \hat{l}, G_{\hat{l}} = G_l$. Переход на п.6.

Построение меры сходства



фильмов

8. Построение меры сходства

1) Оценка на основе суммы весов ребер графа:

$$Sim_{OM}(XY) = \frac{\sum w_{ij}}{p},$$

2) Учет порядка сцен

$$C[i, j] = \begin{cases} 0, i = 0 \text{ or } j = 0 \\ C[i-1, j-1] + 1, i, j > 0, (x_i, y_j) \in M \\ \max\{C[i, j-1], C[i-1, j]\}, i, j > 0, (x_i, y_j) \notin M \end{cases}, \quad Sim_{DP} = \frac{C[p, q]}{p}.$$

3) Interference factor $Sim_{IF}(X, Y) = \frac{2|M|}{p+q}.$

Мера сходства:

$$Sim(X, Y) = \sum_{i=OM, DP, IF} \alpha_i Sim_i(X, Y).$$

Результаты

Размер базы: 21 час видео (19929 смен сцены)

Некоторые результаты:

Table 2: Experimental results for video clip filtering and retrieval

Query type	# of queries	Average # of shot	Average # of relevant clip	Our Approach		Liu's Approach	
				Precision	Recall	Precision	Recall
Commercial	20	12.9	4.0	0.935	1.000	0.628	0.990
News	20	18.5	4.2	0.794	0.735	0.649	0.622
Sport	10	15.0	5.1	0.765	0.684	0.601	0.509
Average	-	15.5	4.4	0.831	0.806	0.626	0.707

«Clip-based Similarity Measure
for Hierarchical Video Retrieval», Yuxin Peng, *hong-Wah Ngo*, MIR'04,
October 15–16, 2004, New York, New York, USA

Результаты

Table 3: Experimental results for the filtering and retrieval of news clips (MM+OM+DP+IF)

	Query clip	Shot #	Relevant Clip #	Our Approach		Liu's Approach	
				Precision	Recall	Precision	Recall
1	Power cut accident in London	18	7	1.000	0.429	1.000	0.286
2	Bus bomb event in Israel	12	6	1.000	0.667	0.429	0.500
3	Six-way talk about North Korea	45	6	0.833	0.833	1.000	0.167
4	The death of an Iraq aga in bomb	29	6	0.714	0.833	1.000	0.667
5	New finance policy	33	6	0.500	0.833	0.800	0.667
6	UK premier follows investigation	15	5	0.200	0.600	0.167	0.400
7	Taiwan politic issue	22	4	1.000	0.750	0.600	0.750
8	National singing competition	22	4	1.000	0.750	1.000	0.750
9	Iraq Policy	10	4	1.000	0.500	0.235	1.000
10	Resignation of a UK official	11	4	0.500	0.500	0.250	0.500
11	CCTV program promotion	11	4	1.000	1.000	0.667	1.000
12	Chinese vice president meets foreigners	19	4	0.333	1.000	0.750	0.750
13	Iraq war	9	3	1.000	0.667	1.000	0.333
14	A UN official died in Iraq	17	3	1.000	0.667	0.222	0.667
15	New policies in the ministry of police	21	3	1.000	1.000	0.667	0.667
16	Soccer association election	16	3	1.000	0.667	0.400	0.667
17	Match for solar energy bus	23	3	1.000	0.667	0.500	0.333
18	Report about blaster virus	8	3	0.400	0.667	1.000	0.667
19	Conflict between Israel and Palestine	17	3	1.000	0.667	1.000	0.667
20	Intel CEO visits China	11	2	0.400	1.000	0.286	1.000
Average		18.5	4.2	0.794	0.735	0.649	0.622

Выводы

❑ Недостатки:

- a) Большая вычислительная сложность $(O((p + q)^4) + O(pq) + O(1))$
- b) Наличие искажений может помешать корректному сравнению ключевых кадров
- c) Достаточно большой размер описаний фильмов

❑ Достоинства:

- a) Возможность поиска перемонтированных фрагментов



Планы

По алгоритму:

- Построение локальных характеристик вблизи смены сцены
- Построить более адекватную оценку вероятности сопоставления, учитывать:
 1. Количество сопоставленных смен сцены (достоверность)
 2. Плотность смен сцены
 3. Характеристики БД
- Искать N самых сильных смен сцены, а не N самых первых

По программе:

- Провести рефакторинг
- Исправить кое-какие баги



Вопросы

?