

Visual Media Retrieval Using Transform-Based Layered Query Scheme

Esin Guldogan and Moncef Gabbouj

Institute of Signal Processing, Tampere University of
Technology, Tampere, Finland

E-mail: esin.guldogan@tut.fi, moncef.gabbouj@tut.fi

Olcay Guldogan

Nokia Technology Platforms,
Tampere, Finland

E-mail: olcay.guldogan@nokia.com

Abstract—This paper presents a visual media querying scheme referred to as Transform-Based Layered Query (TLQ) Scheme. The TLQ scheme mainly aims at decreasing retrieval processing time and run-time memory consumption without degrading retrieval results semantically. The scheme contains abstract layers in indexing and retrieval phases, where each indexing layer corresponds to a retrieval layer. The layers are constructed based on transformations for reducing visual frame and feature data dimensions. The proposed TLQ scheme also involves an unsupervised method for eliminating irrelevant media items between the retrieval layers. A two-layer TLQ system is implemented and integrated into MUVIS content-based multimedia indexing and retrieval framework, and its theoretical advantages are verified with dedicated experiments on image and video databases. The experiments reveal that 75% retrieval performance improvement in terms of process time can be achieved depending on transformation parameters.

Keywords—content-based indexing and retrieval; retrieval optimization; query system.

I. INTRODUCTION

Recent technology improvements along with the Internet growth have led to huge amount of digital multimedia during the recent decades. Various methods, algorithms and systems have been proposed addressing multimedia storage and management problems. Such studies revealed the indexing and retrieval concepts, which have further evolved to Content-Based Multimedia Indexing and Retrieval (CBMIR) [1], [2], [3]. Despite various successful systems, there is no perfect global solution for CBMIR in general.

CBMIR systems often analyze multimedia content via so-called low-level features for indexing and retrieval, such as color, texture and shape. Recent systems intend to combine low and high-level features for achieving significantly higher semantic performance. However, considering such combinations makes retrieval more complex and time-consuming process. Additionally, feature extraction processing time and memory requirements are becoming more important problems.

Due to high memory and processing power requirements, CBMIR has not been widely used on limited platforms, such as mobile devices or distributed systems. Nevertheless, the usage of CBMIR systems on these platforms is becoming widespread. Hence, the performance optimization of indexing and retrieval plays an important role in practical CBMIR studies. Retrieval performance optimization is more visible for

the end-user of a CBMIR system, although indexing affects retrieval directly. Query performance optimization during retrieval consists of three main groups of problems:

- Processing time and computational complexity,
- Disk and run-time memory space requirements, and
- Semantic retrieval performance.

Transform-Based Layered Query (TLQ) System is a new visual multimedia querying scheme for increasing query performance without degrading semantic performance. TLQ is further described in Section 2. A sample TLQ system implementation integrated into MUVIS [1] content-based multimedia indexing and retrieval framework is presented in Section 3. The theoretical benefits of the implemented system and its experimental results are also given in Section 3. Finally Section 4 presents the concluding remarks and discussions.

II. TRANSFORM BASED LAYERED QUERY (TLQ) SCHEME

A. TLQ System Structure

Transform-Based Layered Query (TLQ) is a querying system for multimedia databases that are indexed so-called indexing/querying layers. It mainly aims at reducing retrieval processing complexity, time and memory consumption. As shown in the transformation scheme illustrated by Figure 1, the concerning layers are constructed based on three transforms: T1, T2 and T3. T1 represents an optional transformation working on visual media, where T3 represents a similar optional transformation working only on video data. T2 represents a compulsory transformation working on feature data. Although TLQ system does not directly depend on any specific transformations, underlying framework and transformations should follow the assumptions and restrictions below for achieving overall system targets:

- Indexing process and feature extraction depends on frame size in terms of time, memory usage and complexity.
- Video indexing process also depends on video key-frames in terms of time, memory usage and complexity.
- Query process depends on feature data size in terms of time, memory usage and complexity.

- T1 transform reduces the multimedia data size while keeping its significant content.
- T2 transform reduces the feature data size while keeping the significant information, which causes the main difference between the feature data.
- T3 transform summarizes the key-frames of a video clip by eliminating redundant or insignificant key-frames.

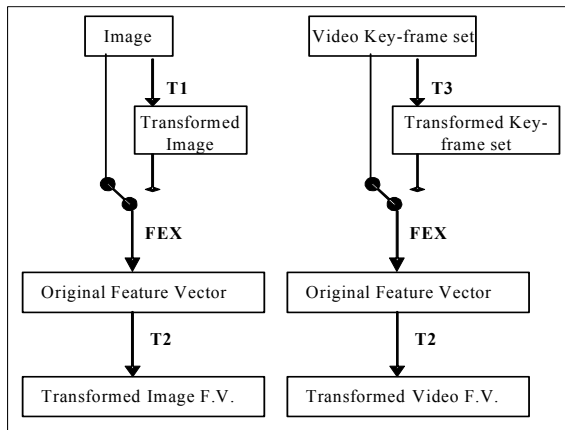


Figure 1: Overview of the transformation scheme

T1 and T3 transforms can be proposed for solving the memory consumption and processing time problems in general. Moreover, T2 leads to lower query processing time. It will also motivate using complex similarity measurement algorithms for increasing semantic query performance. These transforms on multimedia and feature data are not expected to degrade the semantic retrieval performance, since the significant information representing multimedia is kept, and may be even emphasized during the transforms.

TLQ system consists of two phases: Indexing and Retrieval. Both phases utilize the aforementioned transform-based layers.

Indexing Phase:

The TLQ system contains at least two layers, where the uppermost layer corresponds to original size feature data. The original size feature data are extracted from the original or optionally T1 and/or T3 transformed multimedia item. Reduced size feature vectors are generated for lower layers by transforming the original size feature data using T2. A lower layer always corresponds to smaller feature data than higher layer. T2 is adjusted for this purpose in each layer. Optionally, T1 and T3 can also be utilized for layer differentiation. Reduced size feature data of each layer are used in the retrieval phase for retrieving corresponding intermediate results. The intermediate results of a layer are expected to be more accurate than preceding layer's intermediate results.

Using T1 and T3 has certain benefits in the indexing phase, such as feature extraction time and memory consumption reduction. However, all these transforms will bring processing overheads. In most CBMIR systems, indexing is an offline processing that does not involve end-user interaction. Hence the overheads are not relevant for the end-users.

Retrieval Phase:

The retrieval phase of the system contains equal number of layers each corresponding to aforementioned indexing phase layer respectively. The whole retrieval process starts with the lowest (base) layer, which uses the corresponding lowest indexing layer feature data and the whole database. If the user proceeds with the higher layer, the retrieval process uses the corresponding higher indexing layer feature data. The retrieval process in the next layer is performed only within the intermediate results instead of the whole database. The number of multimedia items in the intermediate query results of each layer is based on either user-definition or the automated estimation of the system. The user may explicitly specify the number, or define the distance threshold that will lead to a certain number in each layer. The automated estimation is an unsupervised method for eliminating irrelevant media items as described in Section 2.2.

Using T2 transformed feature data brings lower retrieval processing time, since querying mostly depends on feature vector comparisons and distance calculations. Additionally, elimination of irrelevant multimedia items in each layer leads to a benefit in the higher layer's querying time.

Whenever the user is satisfied with the intermediate results of a layer, there is no need for proceeding with the higher layer. Moreover, if the underlying system allows user feedback, query parameters (e.g. feature types and weights) can be adjusted between the layers.

B. Unsupervised Elimination of Irrelevant Media Items

Unsupervised elimination method is employed between each TLQ layer. It estimates convenient number of relevant media items in a query result, and forms the intermediate result set with that amount of images for the higher layer. The estimation is based on a simple method finding the relevancy threshold within the ordered series of distances, which are mainly the query results. Relevancy threshold is the distance in the series giving the highest gradient. Additionally, higher distances refer to irrelevant results.

III. TLQ SYSTEM IMPLEMENTATION AND EXPERIMENTAL RESULTS

A two-layer TLQ system is implemented and integrated into MUVIS content-based indexing and retrieval framework. Sequential indexing and retrieval method is used for each layer in MUVIS. DCT-based downscaling [4] is employed as T1 transform, since previous studies show that downscaling does not affect image retrieval performance significantly [5]. DCT-based downscaling can only be utilized during image decoding in MUVIS, thus T1 transform is applied on JPEG images in the integrated system. T2 transform is implemented by Principal Component Analysis (PCA), which transforms a number of possibly correlated variables into smaller number of uncorrelated variables referred to as principal components. It decreases the dimension of the feature data and exposes relationship between the objects that facilitate similarity searching by eigen analysis of covariance matrix of vector. Several descriptive examples of PCA for CBMIR exist in the literature [6], [7], [8], [9].

Figure 2 illustrates the retrieval phase of a two-layer TLQ system. Such a retrieval scheme has two typical use cases, in which performance benefits can be described clearly:

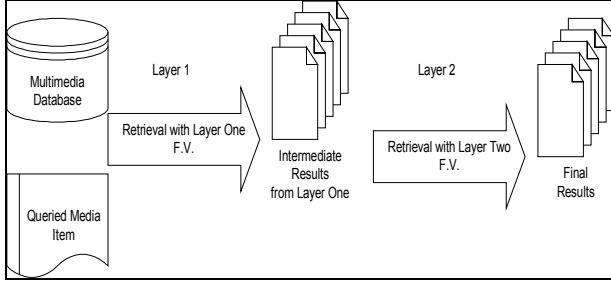


Figure 2: Retrieval phase of a two-layer TLQ system

Use Case 1 Steps:

1. User selects the query media item, adjusts the query parameters and weights, and either sets the number of items in the intermediate Layer One results or selects the “unsupervised elimination” method.
2. User starts the query for the first layer. Intermediate results are displayed for the user.
3. User wishes to proceed with more advanced second layer query. If the underlying system allows, user can give relevance feedback and modify query parameters.
4. User starts the query for the second layer. Final results are displayed for the user.

Use Case 2 Steps:

1. The first two steps are the same as in Use Case 1.
2. User is satisfied with the intermediate results, and stops the query phase.

Theoretical advantages of the implemented TLQ system can be expressed in terms of indexing and retrieval processing times by assuming that;

- Feature extraction processing time is directly proportional to image and video frame dimensions,
- Each video clip has equal number of key-frames,
- DCT-based downscaling and PCA transforms have a parameter k representing the scaling rate, and
- The number of intermediate query results is a quarter of total number of multimedia items.

Based on these assumptions, total database indexing time can be denoted with linear function:

$$F_p(K_p, n_p, m_p), \quad (1)$$

where K_p is the constant feature extraction parameter, n_p is the number of media items in the database, m_p is the average number of pixels per image and video key-frame, and p refers to the underlying system (TLQ or Ordinary).

Since T1 and T3 transforms are involved in the indexing process,

$$K_{tlq} > K_o, \text{ and } K_{tlq} - K_o \ll F_p \quad (2)$$

If $k = 4$, then

$$m_o = 16 * m_{tlq} \quad (3)$$

$$n_{tlq} = n_o \quad (4)$$

Finally,

$$F_o \sim 16 * F_{tlq} \quad (5)$$

In other words, total indexing performance gain of two-layer TLQ is approximately 90%.

Since T1 transform benefits only in offline indexing process, it may be discarded from TLQ system. In this case, F_{tlq} will be slightly higher than F_o .

Similar to F_p , total query process is also represented with linear function:

$$Q_p(R_p, n_p, v_p), \quad (6)$$

where R_p is the constant query parameter, n_p is the number of media items, and v_p is the dimension of the feature vector.

$$Q_{tlq} = Q_{L1} + [Q_{L2}], \quad (7)$$

where $L1$ and $L2$ refer to first and second layers respectively.

$$R_{L1} = R_{L2} = R_o \quad (8)$$

$$n_o = n_{L1} = 4 * n_{L2} \quad (9)$$

If $k = 4$, then

$$v_o = v_{L2} = 4 * v_{L1} \quad (10)$$

Consequently,

$$Q_o = 4 * Q_{L1}, Q_o = 4 * Q_{L2}, \text{ and } Q_o = 2 * Q_{tlq} \quad (11)$$

In other words, the performance gain is 50% for Use Case 1, and 75% for Use Case 2.

The implemented two-layer TLQ system is studied practically using the same scaling rate ($k = 4$). Multiple low-level features such as color histograms are used in the experimental studies. Features of database items are compared to corresponding features of the queried item, and the measured distances are merged with weighted mean for final query results. The experimental image database contains 10000 images, and the video database contains 300 video clips. Different queries on these two databases are used in order to show the results separately for video and image cases, although they have no practical and theoretical difference.

Table 1 presents a set of image and video query process times obtained in the experiments. The experimental results reveal that using scaling rate 4 ($k = 4$) leads to satisfactory results both semantically and practically. Table 1 also shows that, theoretical assumptions and practical test results yield to approximately same performance benefits. Moreover, it is likely to achieve further performance benefits, since the probability of user satisfaction for intermediate Layer One results is high. Figure 3 presents relatively successful Layer

One results for an image query on MUVIS, where corresponding Layer Two results are presented in Figure 4.

TABLE 1: SAMPLE QUERY PROCESS TIMES FROM THE EXPERIMENTS

Times in msec	Ordinary Query Time	TLQ Layer 1 Query Time	TLQ Layer 2 Query Time
Image 1	8062	2422	2109
Image 2	7719	2438	2407
Image 3	7860	2516	2109
Image 4	7484	2265	2172
Video 1	3065	803	782
Video 2	2320	725	574
Video 3	2353	714	355
Video 4	2299	716	556

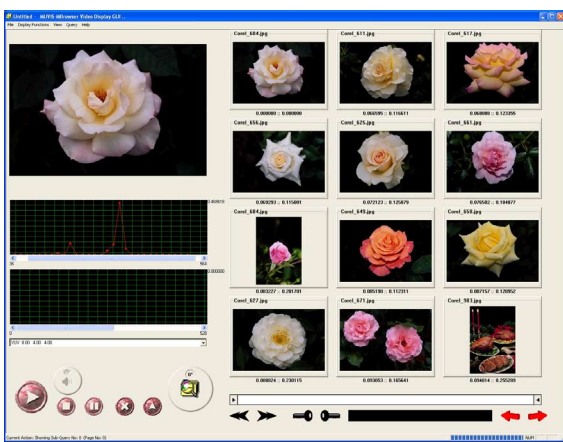


Figure 3: Sample Layer One query results

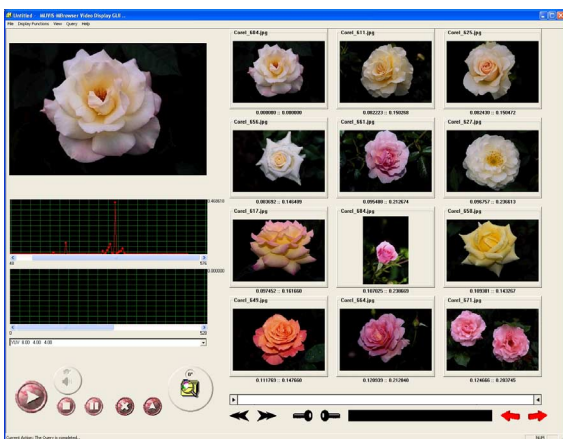


Figure 4: Sample Layer Two query results

IV. CONCLUSIONS AND FUTURE WORK

Transform-Based Layered Query Scheme is a novel retrieval approach addressing practical CBMIR system problems such as reducing query process time and memory consumption. It also involves an unsupervised method for decimating the underlying database by automatically eliminating irrelevant images. Unsupervised elimination method utilizes gradient of distance series that are the results of a query. A two-layer TLQ system is implemented and integrated into MUVIS framework for experimenting theoretical and practical outcomes of the proposed approach. The experiments do not include unsupervised elimination, since it nonlinearly brings unknown number of intermediate

results for different queries. Image and feature data transforms are implemented with DCT-based downscaling and Principal Component Analysis respectively. Expected theoretical performance gain of the integrated system is verified with the experimental results. The experiments reveal that value 4 as scaling rate ($k = 4$) leads to successful semantic and practical results. Moreover, it yields to 50 to 80% query process time benefit. Increasing k achieves higher gains, however it may also degrade semantic performance. The studies also show that satisfactory semantic results can mostly be achieved by Layer One intermediate query results.

Underlying transforms have to comply to certain restrictions in order to achieve a successful TLQ system. The essential restriction can be expressed as “Transform processes should be computationally less complex processes than feature extraction, and transforms should keep the significant information while reducing data dimensions”.

The proposed TLQ system is a flexible system, since it is independent from underlying methods for indexing and retrieval. Moreover, it allows various extensions such as relevance feedback between the layers. The system is also scalable due to multiple abstract layers approach. Such characteristics make the TLQ scheme feasible for integration into various platforms, such as mobile and distributed platforms. The advantages and optimizations of TLQ would be more valuable when applied on limited platforms.

Higher practical benefits and better semantic results can be achieved by improving the integrated two-layer TLQ system with optimized transforms and indexing and retrieval methods. Another potential improvement is integrating and testing a relevance feedback scheme on the TLQ.

REFERENCES

- [1] S. Kiranyaz, K. Caglar, O. Guldogan, and E. Karaoglu, “MUVIS: A Multimedia Browsing, Indexing and Retrieval Framework”, *Proc. of Third International Workshop on Content Based Multimedia Indexing, CBMI'03*, France, September 2003.
- [2] A. Pentland, R.W. Picard, S. Sclaroff, “Photobook: tools for content based manipulation of image databases”, *Proc. of SPIE Storage and Retrieval for Image and Video Databases II*, 1994.
- [3] S. F. Chang, W. Chen, J. Meng, H. Sundaram and D. Zhong, “VideoQ: An Automated Content Based Video Search System Using Visual Cues”, *Proc. of ACM Multimedia*, Seattle, 1997.
- [4] W. Chen, C. H. Smith and S. C. Fralick, “A Fast Computational Algorithm for the Discrete Cosine Transform”, *IEEE Trans. on Communications*, Vol. COM-25, pp. 1004-1009, 1977.
- [5] E. Guldogan, O. Guldogan and M. Gabbouj, “DCT-Based Downscaling Effects On Color And Texture-Based Image Retrieval”, *Proc. of the EWIMT, IEE European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, London, November 2004.
- [6] K. Fukunaga, Introduction to Statistical Pattern Recognition, *Academic Press*, 1990.
- [7] L. Rui, Z. Yongsheng; F. Yonghong, D. Xueqing, “Research on content-based remote sensing image retrieval: the strategy for visual feature selection, extraction, description and similarity measurement”, *Proc. of International Conferences on Info-tech & Info-net, ICII*, Beijing, 2001.
- [8] Li, X.Q. King, I. “Information retrieval using local linear PCA”, *Proc. of 6th International Conference on Neural Information Processing, ICONIP '99*, 1999.
- [9] Linh V. T., Reiner L. “PCA-based representation of color distributions for color based image retrieval”, *Proc. of International Conference Image Processing, ICIP* 2001.