A NOVEL MULTIMEDIA RETRIEVAL TECHNIQUE: **PROGRESSIVE QUERY**
(WHY WAIT?)

*Serkan Kiranyaz and Moncef Gabbouj*

Institute of Signal Processing, Tampere University of Technology, Tampere, Finland
serkan@cs.tut.fi, moncef.gabbouj@tut.fi

**ABSTRACT**

This paper presents a novel multimedia retrieval technique, called Progressive Query (PQ). PQ is designed to bring an effective solution especially when querying large-scale multimedia databases. In addition, PQ produces intermediate query retrieval results during the execution of the query. The series of intermediate query results will finally converge to the full-scale search retrieval in a faster way and with no minimum system requirements. Experimental progressive query retrieval results show that intermediate retrieval results may be satisfactory and further query processing time may be avoided.

1. **INTRODUCTION**

It is a known fact that recent technological hardware and network improvements along with the daily usage of Internet have caused a rapid increase in the size of digital audio-visual information that is used, handled and stored via several applications. Besides several benefits and usages, such massive collection of information has brought storage and especially management problems. In order to overcome such problems several content-based indexing and retrieval techniques and applications have been developed such as MUVIS system [16], [19], Photobook [21], VisualSEEk [24], Virage [31], and VideoQ [10], some of which are designed to bring a framework structure for the multimedia items such as digital images and audio/video clips. The usual approach for indexing is to map database primitives into some high dimensional vector space that is so called feature domain. The feature domain may consist of several types of features such as visual, aural, motion, etc. as long as the database contains such items from which those particular features can be extracted. Among so many variations, careful selection of the feature sets allows capturing the semantics of the database items. Especially for large-scale multimedia databases the number of features extracted from the raw data is often kept large due to the naïve expectation that it helps to capture the semantics better. Content-based similarity between two
database items can then be assumed to correspond to the (dis-) similarity distance of their feature vectors. Henceforth, the retrieval of a similar database items with respect to a given query (item) can be transformed into the problem of finding such database items that give feature vectors, which are close to the query feature vector. This is so-called query-by-example (QBE), which is one of the most common retrieval scheme. The basic QBE operation is so called Normal Query (NQ), and works as follows: using the available aural or visual features (or both) of the queried multimedia item (i.e. an image, a video clip, an audio clip, etc.) and all the database items, the similarity distances are calculated and then merged to obtain a unique similarity distance per database item. Ranking the items according to their similarity distances (to the queried item) over the entire database yields the query result.

Such an exhaustive search for QBE is costly and CPU intensive especially for large-scale multimedia databases since the number of similarity distance calculations is proportional with the database size. This fact brought a need for indexing techniques, which will organize the database structure in such a way that the query time and I/O access amount could be reduced. During the past three decades, several indexing techniques have been proposed. Many of these techniques are formed in a hierarchical tree structure that is used to cluster (or partition) the feature space. Both KD-tree [2] and R-tree [13] are the first examples of Spatial Access Methods (SAMs). Afterwards several enhanced SAMs have been proposed. R*-tree [1] provides a consistently better performance than the R-tree and R+-tree [17] by introducing a policy called “forced reinsert”. Lin et al. proposed TV-tree [18], which uses so-called telescope vectors. Berchtold et al. [5] introduced X-tree, which is particularly designed for indexing higher dimensional data. X-tree avoids overlapping of region bounding boxes in the directory structure by using a new organization of the directory and as a result of this X-tree outperforms both TV-tree and R*-tree significantly. Still bounding rectangles can overlap in the higher dimensions. In order to prevent this, White and Jain proposed the SS-tree [30], an alternative to R-tree structure, which uses minimum bounding spheres instead of rectangles. Even though SS-tree outperforms R*-tree, the overlapping in the high dimensions still occurs. Thereafter several other SAM variants are proposed such as SR-tree [15], S²-Tree [28], Hybrid-Tree [9], A-tree [23], IQ-tree [3], Pyramid Tree [4], NB-tree [12], etc. Especially for content-based indexing and retrieval in large-scale multimedia databases, SAMs have several drawbacks and significant weaknesses. By definition an SAM-based indexing scheme partitions and works over a single feature space. However, a multimedia database can have several feature types (visual, aural, etc.), each of which might also have multiple feature subsets. Furthermore, SAMs assume that query operation time and complexity are only related to accessing a disk page (I/O access time) containing the feature vector. This is obviously not a trivial assumption for multimedia databases and consequently, no attempt in the design of SAMs has been done to
reduce the similarity distance computations (CPU time). In order to provide a more general approach to similarity indexing for multimedia databases, several efficient Metric Access Methods (MAMs) have been proposed. The generality of MAMs comes from the fact that any MAM employs the indexing process by assuming only the availability of a similarity distance function, which satisfies three trivial rules: symmetry, non-negativity and triangular inequality. Therefore, a multimedia database might have several feature types along with various numbers of feature sub-sets all of which are in different multi-dimensional feature spaces. As long as a similarity distance function that is usually treated as a “black box” by the underlying MAM, exists the database can be indexed by any MAM. Several MAMs are proposed so far. Yianilos [32] presented vp-tree that is based on partitioning the feature vectors (data points) into two groups according to their similarity distances with respect to a reference point, so called vantage point. Bozkaya and Ozsoyoglu [6] proposed an extension of vp-tree, so-called mvp-tree (multiple vantage point), which basically assigns $m$ vantage points for a node with a fan out of $m^2$. They reported 20% to 80% reduction in the similarity distance computation time compared to vp-trees. Brin [7] introduced Geometric Near-Neighbor Access Tree (GNAT) indexing structure, which chooses $k$ number of split points at the top level and each of the remaining feature vectors are associated with the closest split points. GNAT is then built recursively and the parameter $k$ value is chosen to be a different value for each feature set depending on its cardinality. The aforementioned MAMs have several shortcomings. Contrary to SAMs, these metric trees are designed only to reduce the number of similarity distance computations, paying no attention to I/O costs (disk page accesses). They are also static methods in the sense that the tree structure is built once and new insertions are not supported. Furthermore, all of them build the indexing structure from top to bottom and hence the resulting tree is not guaranteed to be balanced. Ciaccia et al. [11] proposed M-tree to overcome such problems. M-tree is a balanced and dynamic tree, which is built from bottom to top, creating a new root level only when necessary. The node size, $M$, is a fixed number and therefore, the tree height depends on $M$ and the database size. Its performance optimization concerns both CPU computational time for similarity distances and I/O costs for disk page accesses for feature vectors of the database items. Recently Traina et al. [26] proposed Slim-tree, an enhanced variant of M-trees, which is designed for improving the performance by minimizing the overlaps between nodes. They introduced two factors, “fat-factor” and “bloat-factor”, to measure the degree of overlap and proposed the usage of Minimum Spanning Tree (MST) for splitting the node. Another slightly enhanced M-tree structure, so-called M+-tree, can be found in [33].

Along with the indexing techniques addressed so far, certain query techniques are needed to speed up a QBE process. The most common query techniques developed for the aforementioned indexing techniques are as follows:
• **Range Queries**: Given a query object, \( Q \), and a maximum similarity distance range, \( \varepsilon \) and the similarity distance function \( SD \), the range query selects all indexed database items, \( Q_i \), such that \( SD(Q, Q_i) < \varepsilon \).

• **kNN Queries**: Given a query object, \( Q \), and an integer number \( k > 0 \), \( kNN \) query selects the \( k \) database items, which have the shortest similarity distance from \( Q \).

Unfortunately, both query techniques may not provide efficient retrieval scheme from the user’s point of view due to their parameter dependency. For instance, range queries require a distance parameter, \( \varepsilon \), where the user may not be able to provide such a number prior to a query process since it is not obvious to find out a suitable range value if the database contains various types of features and feature subsets. Similarly, parameter \( k \) in a \( kNN \) query may be hard to determine since it can be too small in case the database may provide many more similar (relevant) items than required, or too big if the number of similar objects is only a small fraction of the required number \( k \), which means unnecessary CPU time has been wasted for that query process. Both query techniques often require several trials to converge to a successful retrieval result and this alone might remove the speed benefit of the underlying indexing scheme, if there is any.

As mentioned before, the other alternative is the so-called **Normal Query (NQ)**, which makes a sequential (exhaustive) search due to lack of an indexing scheme. \( NQ \) for QBE is costly and CPU intensive especially for large-scale multimedia databases; however, it yields such a final and decisive result that no further trials are needed. Still, all QBE alternatives have some common drawbacks. First of all, the user has to wait until all (or some) of the similarity distances are calculated and the searched database items are ranked accordingly. Naturally, this might take a significant time if the database size (or \( k, \varepsilon \)) is large and the database contains a rich set of aural and visual features, which might further reduce the efficiency on the indexing process. Furthermore, any abrupt stopping (i.e. manual stop by the user) during the query process will cause total loss of retrieval information and essentially nothing can be saved out of the query operation so far performed. In order to speed up the query process, it is a common application design procedure to hold all features of database items into the system memory first and then perform the calculations. Therefore, the growth in the size of the database and the set of features will not only (proportionally) increase the query time (the time needed for completing a query) but it might also increase the minimum system memory requirements such as memory capacity and CPU power.

In order to eliminate such drawbacks and provide a faster query scheme, we develop a novel retrieval scheme, the **Progressive Query (PQ)**, which is implemented under MUVIS system to provide a basis for multimedia retrieval and to test the performance of the technique. \( PQ \) is a retrieval (via QBE) technique, which can be performed over the databases
with or without the presence of an indexing structure. Therefore, it can be an alternative to \( NQ \) where both produce (converge to) the same result at the end. When the database has an indexing structure, \( PQ \) can replace \( kNN \) and range queries whenever a query path over which \( PQ \) proceeds, can be formed. As its name implies, \( PQ \) provides intermediate query results during the query process. The user may browse these results and may stop the ongoing query in case the results obtained so far are satisfactory and hence no further time should unnecessarily be wasted. As expected, \( PQ \) and \( NQ \) will converge to the same (final) retrieval results at the end. Furthermore, \( PQ \) may perform the overall query process faster (within a shorter total query time) than \( NQ \). Since \( PQ \) provides a series of intermediate results, each of which obtained from a (smaller) sub-set within the database, the chance of retrieving relevant database items that would not be retrieved otherwise via \( NQ \), might be increased. Approvingly some experimental results show that it is quite probable to achieve even better retrieval performance within an intermediate sub-query than the final query state.

It is a known fact that significant performance improvements of content-based multimedia retrieval systems can be achieved by using a technique known as relevance feedback \([8],[22]\), which allows the user to rate (intermediary) retrieval results. This helps to tune the ranking and retrieval parameters and hence yields a better retrieval result at the end. Traditional, query techniques so far addressed (\( NQ, kNN \) and range queries) may allow such a feedback only after the query is completed. Since \( PQ \) provides the user with intermediate results during the query process, relevance feedback may be applied already to these intermediate results, yielding possibly faster satisfactory retrieval results from the user’s point of view.

The rest of this paper is organized as follows. In Section 2, we introduce the generic \( PQ \) design philosophy and the implementation details. Section 3 presents the experimental results and some example demonstrations. Finally, conclusions are given in Section 4.

2. PROGRESSIVE QUERY OVERVIEW

The principal idea behind the design of \( PQ \) is to partition the database items into some sub-sets within which individual (sub-) queries can be performed. Therefore, a sub-query is a fractional query process that is performed over any sub-set of database items. Once a sub-query is completed over a particular sub-set, the incremental retrieval results (belonging only to that sub-set) should be fused (merged) with the last overall retrieval result to obtain a new overall retrieval result, which belongs to the items where \( PQ \) operation so far covers from the beginning of the operation. Note that this is a continuous operation, which proceeds incrementally, sub-set by sub-set, by covering more and more group of items within the
database. Each time a new sub-query operation is completed, $PQ$ updates the retrieval results to the user. Since the previous (overall) query results are used to obtain the next (overall) retrieval result via fusion, the time consuming query operation is only performed over the (next) partitioned group of items instead of all the items where $PQ$ covered so far.

The order of the database items processed is a matter of the indexing structure of the database. If the database is not indexed at all, simply a sequential or random order can be chosen. In case the database has an indexing structure, a query path can be formed in order to retrieve the most relevant items at the beginning during a $PQ$ operation. Since there are various indexing schemes addressed in the previous section, for the sake of simplicity, we shall first explain the basics of $PQ$ for a database with no indexing structure.

Another important factor is to determine the size of each sub-set (i.e. the number of items within a sub-set where sub-query operation is performed) that is most convenient from the user’s point of view. A straightforward solution is to let the user fix the sub-set size (say e.g. 25). This would mean that the user wants updates every time 25 items are covered during the ongoing $PQ$ operation. However, this also brings the problem of uncertainty because the user cannot know how much time a sub-query will take beforehand since the sub-query time will vary due to the factors such as the amount of features present in the database and the speed of the computer where it is running, etc. So the $PQ$ retrieval updates might be too fast or too slow for the user. In order to avoid such uncertainties, the proposed $PQ$ scheme is designed over periodic sub-queries as shown in Figure 1 with a user defined period value ($t = t_p$). The period (time) is obviously more natural choice since the user can eventually expect the retrieval results will be updated every $t_p$ seconds no matter what database is involved or what computer is used. Without loss of generality, in databases without an indexing structure, $PQ$ is designed to perform sub-set partitioning sequentially with a forward direction (i.e. starting from the first item to the last one).

Next, we shall first present the formation of periodic sub-queries and the mechanism needed to partition the database into such sub-sets that each Progressive Sub-Query ($PSQ$) retrieval result can be updated within a close neighborhood of a user-defined period ($t_p$). Sub-query fusion operation will then be explained in Section 2.2. Finally, Section 2.3 is specifically devoted to $PQ$ for databases with an (hypothetical) indexing structure.
2.1. Periodic Sub-Query Formation

In order to achieve periodic sub-queries, we need to define some additional sub-query compositions:

- **Atomic Sub-Query**: The smallest sub-set size on which a sub-query is performed. Here we assume that atomic sub-query time is not significant compared to periodic sub-query time. Atomic sub-queries are the only sub-query types that have a fixed sub-set size ($S_{ASQ}$). They are only used during first periodic query and they are used in order to provide an initial sub-query per-item time ($t_r^0$), that is the time spent for the retrieval of a single database item, formulated as follows:

$$
t_r^0 = \frac{t_{ASQ}}{N_{ASQ}} \quad \text{if} \quad N_{ASQ} > 0 \quad (1)
$$

where $t_{ASQ}$ is the total time spent for atomic sub-query and $N_{ASQ}$ is the number of database items, which are involved (used) in the atomic sub-query operation. Without an indexing structure, note that $0 \leq N_{ASQ} \leq S_{ASQ}$, since the initial database items might not belong to the ongoing query type. For example in a multimedia database, there might be video-only clips and audio-only clips (and clips with both media types). So for a visual query, those audio-only clips will be totally discarded and if the initial atomic query sub-set covers such audio-only clips then naturally $N_{ASQ} \leq S_{ASQ}$. In case $N_{ASQ} = 0$, one or more atomic sub-queries have to be performed until we get a valid $t_r^0$ value (i.e. $N_{ASQ} > 0$).
• **Fractional Sub-Query:** This can be any sub-query performed over a sub-set whose size is smaller or equal to the sub-set size of the periodic sub-query. In other words a fractional sub-query time might be less than or equal to a periodic sub-query time.

As explained earlier, periodic sub-queries are periodic over time and a mechanism is needed to ensure this periodicity. This mechanism works over atomic and fractional sub-queries; it performs fusion operation over as many atomic and fractional sub-queries as necessary. First, it starts with an atomic sub-query to obtain a valid (initial) sub-query per item time, $t_r^0$, and it keeps going with atomic queries until a valid $t_r^0$ value is obtained. Once it is obtained, then one or more fractional sub-queries will be performed to complete the first periodic sub-query. The size of the fractional query ($N_{FSo}$) can then be estimated as:

$$N_{FSo}^0 \approx \frac{t_p^0}{t_r^0} \quad (2)$$

where $t_p^0 = t_p - t_{\Sigma}$ is the time left for completing the first periodic sub-query and $t_{\Sigma}$ is the total time spent for all atomic queries so far performed. Afterwards, the fractional sub-query can be performed within a sub-set of $N_{FSo}^0$ items. Once the fractional sub-query is completed, the total time ($t_{\Sigma}$) so far spent from the beginning of the operation till now is compared with the required time period of the $q^{th}$ (so far $q=0$) periodic query, $t_p^q$, where $q$ is the periodic sub-query index. If $t_{\Sigma}$ value is not within a close neighborhood (i.e. $\delta t_{w} < 0.5$ sec.) of $t_p^q$ (i.e. $t_{\Sigma} < t_p^q - \delta t_{w}$) then the operation continues with a new fractional sub-query until the condition is met. For the new fractional sub-query and for all the latter fractional sub-query operations $t_r$ value is re-estimated (updated) from the former operations such as:

$$N_{FSo}^i = \frac{t_p^q - t_{\Sigma}}{t_r^i} \implies t_r^{i+1} = \frac{t_{\Sigma}}{\sum_{i \in FSO} N_{FSo}^i} \quad \text{if} \quad \sum_{i \in FSO} N_{FSo}^i > 0 \quad (3)$$

Once one or more fractional queries form the $q^{th}$ periodic query, due to offset that is occurred from the period of $PQ$, next periodic sub-query ($q+1^{th}$) is formed with an updated (offset removed) period value:

$$t_p^{q+1} = t_p + (t_p^q - t_{\Sigma}) \quad (4)$$
where $t_p$ is the required period and $(t_p^q - t_{\sum})$ is the offset time, which is the time difference between the required period time for $q^{th}$ periodic sub-query and the total (actual) time spent so far. The flowchart of the formation of a periodic sub-query is shown in Figure 2.

![Flowchart](image)

**Figure 2: Formation of a Periodic Sub-Query.**

### 2.2. Sub-Query Fusion Operation

The overall $PQ$ operation is carried out over Progressive Sub-Queries ($PSQ$s). In principal, it can be stated that $PQ$ is a (periodic) series of $PSQ$ results. As shown in Figure 1, a new $PSQ$ retrieval is formed each time a periodic sub-query is completed and then it is fused with the previous $PSQ$ retrieval result. Once a $PSQ$ is realized, the results are shown to the user and saved during the lifetime of the ongoing $PQ$ so that the user can access them at any time. The user is shown updated retrieval results each time a new $PSQ$ is completed. The first $PSQ$ is the first periodic sub-query performed. The
fusion operation is a process of fusing two sorted sub-query results to achieve one (fused) sub-query result. Since both of the sub-query results are already sorted with respect to the similarity distances, simply comparing the consecutive items in each of the sub-query lists can perform the fusion operation. Let \( n_1 \) and \( n_2 \) be the number of items in the first and second sub-set, respectively. Since there are \( n_1 + n_2 \) items to be inserted into the fused list one at a time, the fusion operation can take maximum \( n_1 + n_2 \) comparisons. This (worst) case occurs whenever the items from both lists are evenly distributed with respect to each other. On the other hand if the maximum valued item (i.e. the last item) in the smaller list is less than the minimum valued item (i.e. the first item) in the bigger list, the number of comparisons will not exceed the number of items in the smaller list because once all of the items in it are compared with the (smallest) first item in the bigger list and henceforth inserted into the fused list, there will not be any more comparisons needed. Note that this is the direct consequence of the fact that the both lists are sorted (from minimum to maximum) beforehand and one of them is now fully depleted. Therefore, the fusion operation will take minimum \( \text{Min}(n_1, n_2) \) comparisons respectively. A sample fusion operation is illustrated in Figure 3. Note that the subsets \( X \) and \( Y \) contain 12 and 6 items, respectively, and the fusion operation performs only 12 comparisons.

**Figure 3:** A sample fusion operation between subsets \( X \) and \( Y \).
Since the fusion operation is nothing but merging two arbitrary sized sub-sets (retrieval results), it can be applied during each phase of a PQ operation. For instance, the atomic sub-queries are fused with fractional sub-queries and several fractional sub-queries are fused to obtain a periodic sub-query. Fusing the periodic sub-query with the previous PSQ retrieval produces a new PSQ retrieval, covering a larger part of the database. If the user does not stop the ongoing PQ operation, it will eventually cover the entire database at the end and therefore, it generates the overall retrieval result of the queried item. In this case PQ generates the exact same retrieval result as NQ since both of them perform the same operation, i.e. searching a queried item through the entire database and ranking the database items accordingly.

2.3. PQ in Indexed Databases

For the databases without an indexing structure, PQ can be used conveniently as an alternative query scheme to the traditional query type, NQ. As a retrieval process, PQ can also be performed over indexed databases as long as a query path can be formed over the clusters (partitions) of the underlying indexing structure. Obviously, query path is nothing but a special sequence of the database items. When the database lacks an indexing structure, it can be formed in any convenient way such as sequentially (starting from the 1st item towards the last one, or vice versa) or randomly. Otherwise, the most advantageous way to perform PQ is to use the indexing information so that the most relevant items can be retrieved in earlier PSQ steps. Since the technical details of various available indexing schemes are beyond the scope of this paper, we only show the hypothetical formation of the query path and run-time evaluation of PQ over this path.

As briefly mentioned earlier, the primary objective of indexing structures is to partition the feature domain into such (tree-based) clusters that CPU time and I/O accesses are shortened via pruning of the redundant tree nodes. Figure 4 shows a hypothetical clustering scheme and the formation of the query path over which PQ will proceed during its run-time. The illustration shows 4 clusters (partitions or nodes), which contain a certain number of items (features) and the query path is formed according to the relative (similarity) distance to the queried item and its parent cluster. Therefore, PQ will give the priority to cluster A (the host), then B (the closest), C, D, etc. Note that the query path might differ from the final retrieval result depending on the accuracy of the indexing scheme. For instance, query path gives priority to item B2 on the search with respect to item C4 but item C4 may have more similarity (relevancy) with respect to the queried item A2. When the retrieval results are formed it will eventually be ranked higher and presented earlier to the user by PQ. Even though PQ
corrects this misleading result due to the erroneous indexing (note that in this case item C4 should have belonged to cluster B, not C), as a possible consequence of this, the retrieval of C4 might be delayed to the next periodic PSQ retrieval.

At this point, one can implement two different approaches: the overall query path can be formed immediately after the query is initiated and then the PQ evolves over it with its natural supplies of periodic retrievals. This approach is only recommended for small and medium sized databases where the complete query path formation takes insignificant time. Otherwise, the query path should be dynamically (incrementally) formed along with the PQ run-time process and the time spent for it should be taken into account during the adaptive calculation of period given in (4). In this case, the adaptive period calculation for the $q + 1^{st}$ periodic sub-query period should be reformulated as follows:

$$t_p^{q+1} = t_p + (t_p^q - \sum t_{QP}^q)$$  \hspace{1cm} (5)

where $t_{QP}^q$ is the time spent for forming the query path during the formation of $q^{th}$ periodic sub-query.

![Query Path](image)

**Figure 4: Query path formation in a hypothetical indexing structure.**

As a result $PQ$ in indexed databases makes more sense than to be strictly dependant to an unknown parameter such as $k$ as in $kNN$ query or $\varepsilon$ in range query, which might cause a deficiency in the retrieval performance such as the casual need for doing multiple queries to come with a satisfactory result at the end. On the other hand there exists a certain degree
of similarity (or analogy) between $PQ$ and those conventional query techniques. For instance each $PSQ$ retrieval can be seen as a particular $kNN$ (or range) query retrieval with only one difference: the parameter $k$ (or $\varepsilon$) is not fixed beforehand, rather dynamically changing (growing) over time along with the lifetime of $PQ$ and the user has the opportunity to fix it (stop $PQ$) whenever satisfactory retrievals obtained.

3. EXPERIMENTAL RESULTS

In order to present the experimental conditions, the multimedia databases used and especially the test-bed platform used for experiments, Section 3.1 briefly introduces MUVIS and particularly $MBrowser$ application under which $PQ$ is primarily developed and tested. Later in Section 3.2 we begin the evaluation of $PQ$ with respect to $NQ$ in terms of speed, memory, accessibility and (better) relevancy in the databases with no indexing structure and hence without interference of any indexing algorithm. We shall also present the evaluation of $PQ$ with respect to the variations of the query parameters such as $PQ$ period ($t_p$), query type, etc. and show some sample visual and aural $PQ$ retrievals in image, video and audio databases.

3.1. $PQ$ within MUVIS

MUVIS framework aims to bring a unified and global approach to indexing, browsing and querying of various multimedia types such as audio/video clips and still images. MUVIS basically provides tools for real-time audio and video capturing, encoding by several widely used codecs such as MPEG-4, H.263+, MP3 and AAC. It supports several digital image types including JPEG-2000. Furthermore, it provides a well-defined interface to integrate third party feature extraction algorithms into the framework.

As shown in Figure 5, MUVIS framework is based upon three applications, each of which has different responsibilities and facilities. $AVDatabase$ is mainly responsible for real-time audio/video database creation with which audio/video clips are captured, (possibly) encoded and recorded in real-time from any peripheral audio and video devices connected to a computer. $DbsEditor$ performs the indexing of the multimedia databases and therefore, offline feature extraction process over the multimedia collections is its main task. $MBrowser$ is the primary media browser and retrieval application into which $PQ$ technique is integrated as the primary retrieval (QBE) scheme. $NQ$ is the alternative query scheme within $MBrowser$. Both $PQ$ and $NQ$ can be used for ranking the multimedia primitives with respect to their similarity to a queried media item (an audio/video clip, a video frame or an image). Due to their unknown duration, which
might cause impractical indexing times for an online query process, in order to query an (external) audio/video clip, it should first be appended (offline operation) to a MUVIS database upon which the query will be performed. There is no such necessity for images; any digital image (inclusive or exclusive to the active database) can be queried within the active database. The similarity distances will be calculated by the particular functions, each of which is implemented in the corresponding visual/aural feature extraction (FeX or AFeX) modules. More detailed information about MUVIS can be found in [16] and [19].

As shown in Figure 6, MBrowser GUI is designed to support all the functionalities that PQ provides. Once a MUVIS database is loaded into the MBrowser, the user can browse among the database items, choose any item and then initiate a query. The basic query parameters such as query type (PQ or NQ), query genre (aural or visual), PQ update period (time), the (visual and aural) set of features and their individual parameters (i.e. feature weights), etc. can be set prior to a query operation. When a (sub-) query is completed the retrieval results are then presented to the user page by page. Each page renders 12 ranked results in the descending order (from left to right and from top to bottom) and the user can browse the pages back and forth using the Next and Prev buttons on the bottom-right side of the figure (the first page with 12-best retrieval results is shown on the right side of Figure 6). If NQ is chosen, then the user has to wait till the whole process is
completed but if $PQ$ is chosen then the retrieval results will be updated periodically (with the user-defined period value) each time a new $PSQ$ is accomplished. The current $PSQ$ number (10) is shown on the $PQ$ Browser Handle and this handle can also be used to browse manually among the retrieved $PSQ$ results during (or after) an ongoing $PQ$ operation. In the snapshot shown in Figure 6, a video clip is chosen within a MUVIS (video) database and visual $PQ$ is performed. Currently the 1st page (12-best results) of the 10th $PSQ$ retrieval results is shown on the GUI window of $MBrowser$.

![MBrowser GUI showing a PQ operation where 10th PSQ is currently active (or set manually).](image)

In the experiments performed in this section, we used 4 different MUVIS databases:

1) **Open Video** Database: This database contains 1500 video clips, each of which is downloaded from “The Open Video Project” web site [25]. The clips are quite old (from 1960s) but contain color video with sound. The total duration of the database is around 46 hours.

2) **Real World** Audio/Video Database: There are 800 audio-only clips and video clips in the database with a total duration of over 36 hours. They are captured from several TV channels and the content is distributed among News, Advertisements, Talk Shows, Cartoons, etc.
3) **Sports** Hybrid Database: There are 200 video clips mainly carrying sports content such as Football, Tennis and Formula-1. There are also 495 images (in GIF and JPEG formats) showing instances from Football matches and other sport tournaments.

4) **Shape** Image Database: There are 1500 black and white (binary) images that mainly represent the shapes of different objects such as animals, cars, accessories, geometric objects, etc.

All experiments are carried out on a Pentium-4 3.06 GHz computer with 2048 MB memory. In order to have unbiased comparisons between \(PQ\) and \(NQ\), the experiments are performed using the same queried multimedia item with the same instance of \(MBrowser\) application. The evaluations of the retrieval results by QBE are performed subjectively using ground-truth method, i.e. a group of people evaluates the query results of a certain set of retrieval experiments, upon which all the group members totally agreed about the query retrieval performance. Among these a certain set of examples were chosen and presented in this article for visual verification.

### 3.2. \(PQ\) versus \(NQ\)

As explained earlier, \(PQ\) and \(NQ\) eventually converge to the same retrieval result at the end. Also in the abovementioned scenarios they are both designed to perform exhaustive search over the entire database within MUVIS. However \(PQ\) have several advantages over \(NQ\) in the following aspects:

- **System Memory Requirement**: The memory requirement is proportional to the database size and the number of features present in a \(NQ\) operation. Due to the partitioning of the database into sub-sets, \(PQ\) will reduce the memory requirement by the number of \(PSQ\) operations performed. After each periodic sub-query operation, the memory used for feature vectors in that sub-set is no longer needed and can be used for the next periodic sub-query. Figure 7 illustrates the memory usage of a retrieval example that is shown in Figure 9 by a \(PQ\) first and a \(NQ\) afterwards. As clearly seen in this experimentation, \(PQ\) only requires a fractional memory per sub-query operation and it can free it at the end of each PSQ retrieval, whereas \(NQ\) needs to keep a big amount of memory for allocations of features used and similarity distances calculated afterwards, till to the end of the sorting (ranking) and rendering (on screen) operations that are naturally reached only at the end.

We also observe that especially in very large-scale databases containing a rich set of features, \(NQ\) can exceed the system memory. Two possible outcomes may eventually occur as a result. The operating system may handle it by using virtual memory (i.e. disc) if the excessive memory requirement is not too high. In this case, the operational speed for a \(NQ\)
operation will be drastically degraded and eventually \(PQ\) can outperform \(NQ\) several times with respect to overall retrieval time. The other possibility is that \(NQ\) operation cannot be completed at all since the system is not capable of providing the excessive memory required by \(NQ\) and in this case \(PQ\) is the only feasible query operation.

![Figure 7: Memory usage for \(PQ\) and \(NQ\).](image)

- **Earlier/Better Relevant Results**: A better retrieval result occurs when more relevant items are ranked higher in the retrieval result. An earlier (and equal) result occurs when the same end result is achieved in an earlier time instance. Along with the ongoing process \(PQ\) allows intermediate query results (\(PSQ\) steps), which might sometimes show equal or ‘even better’ performance than the final (overall) retrieval result as some typical examples given in Figure 8 and Figure 9.

In Figure 8, an image retrieval example within \textbf{Shape} database via \(PQ\) using Canny Edge Histogram feature is shown. We use \(t_p = 0.2\ \text{sec}\) and \(PQ\) operation is completed in three \(PSQ\) series (i.e. \(PQ\) #1, #2 and #3). This is one particular example in which an intermediate \(PSQ\) retrieval yields a better performance than the final \(PQ\) retrieval (that is same as the retrieval result of \(NQ\)). In this example, the first 12-best retrieval results in \(PQ\) #1 are obviously better (more relevant in terms of shape similarity to the queried shape, i.e. “mobile phone”) than the ones in \(PQ\) #3 (\(NQ\)).

In Figure 9, a video retrieval example within \textbf{Real World} database via aural \(PQ\) using \textit{MFCC} (Mel-Frequency Cepstral Coefficients) as the audio features is shown. We use \(t_p = 5\ \text{sec}\) and \(PQ\) operation is completed in 12 \(PSQ\) series but only 3 \(PSQ\) retrievals (i.e. \(PQ\) #1, #6 and #12) are shown. Note that \(PQ\) #6 and the latter retrieval results till \(PQ\) #12 are identical, which means that \(PQ\) operation produces the final retrieval result (which \(NQ\) would produce) in an earlier (intermediate) \(PSQ\) retrieval.

Such “earlier and even better” retrieval results can be verified due to the fact that searching an item in a smaller data set usually yields better (detection or retrieval) performance than searching in a larger set. This is obviously an advantage for \(PQ\) since it proceeds within sub-queries performed in (smaller) sub-sets whereas \(NQ\) always has to proceed through the
entire database. Furthermore, for the databases that are not indexed such as in the examples given, this basically means that the order of the relevant items coincides with the progress of the ongoing PQ operation in the earlier PSQ steps. When the database has a solid indexing structure and a query path can be formed according to the relevancy of the queried item, the user eventually gets relevant retrieval results in a fraction of the time that is needed for a typical NQ operation.

Figure 8: PQ retrievals of a query image (left-top) within 3 PSQs ($t_p = 0.2$ sec).

Figure 9: Aural PQ retrievals of a video clip (left-top) in 12 PSQs (only 1st, 6th and 12th are shown with $t_p = 5$ sec).
• **Query Accessibility**: This is the major advantage that PQ provides. Stopping an ongoing query operation is an important capability. As shown in Figure 6, by pressing the PQ Knob during an ongoing PQ operation, the user can stop it any time (i.e. when the results are so far satisfactory). Of course, NQ can also be stopped but no retrieval result can be available afterwards since the operations such as similarity distance calculations or sorting are likely not completed.

Another important accessibility option that PQ offers is so-called PSQ Browsing. When stopped abruptly or completed at the end, the user can still browse among PSQ retrievals and visit any retrieval page of that particular PSQ since the retrieval results will be alive unless a new PQ is initiated or the application is terminated. This is obviously a significant requirement especially when better results are obtained in an earlier PSQ step than the later ones as mentioned before. On the other hand, this might still be a desirable option even if the earlier PSQ results are not better but comparable as much as the later ones. This could be relevant to the user. One particular example is shown in Figure 10, that is, a video retrieval example from the Open Video database via visual PQ using several color (YUV, HSV, etc.), texture (GLCM [20]) and shape (Canny Edge Histogram) features. We use $t_p = 3$ sec and PQ operation is completed in 4 PSQ series. Note that a retrieval performance criteria can be difficult to accomplish among these PSQ retrievals since their relevancy to the queried item is subjective.

![Figure 10: Visual PQ retrievals of a video clip (left-top) in 4 PSQs ($t_p = 3$ sec).](image)
The most important accessibility advantage that PQ can provide is that it can further improve the efficiency of any relevance feedback mechanism in certain ways. An ordinary relevance feedback technique works as follows: the user initiates a query and only after the query is completed, the user gives some feedback to the system about the retrieval results according to their relevancy (and/or irrelevancy) with respect to the queried item. Afterwards, a new query is initiated in order to get better retrieval results and this might be repeated several times until satisfactory results are obtained. This is especially a time consuming process since at each iteration the user has to wait until the query operation is completed. Due to the enhanced accessibility options that PQ provides, significant improvements can be achieved for the user with the following scenarios: First the user can employ relevant (and irrelevant) feedbacks during the query process and the incoming progressive retrievals can thus be tuned progressively. This means that during an ongoing query process the user can employ one or more relevant feedbacks anytime (within the life-time of PQ). Another alternative is that, the user can stop an ongoing PQ and then employs the relevant feedbacks with respect to the (intermediate) retrievals via PSQ Browsing and re-initiate a new (fine-tuned) PQ. Basically any relevance feedback technique can be applied along with PQ since in both scenarios PQ only provides the necessary basis for the (user) accessibility to employ the relevance feedback but otherwise stay independent from the internal structure of any individual technique employed.

- **Overall Retrieval Time (Query Speed):** The overall query time is the time elapsed from the beginning of the query to the end of the operation. For NQ, this is obviously the total time from the moment the user initiates the query until the results are ranked and displayed on the screen. However, for PQ, since the retrieval is a continuous process with PSQ series, the overall retrieval means that PQ proceeds over the entire database until the process terminates. As mentioned earlier, at this point both PQ and NQ will generate identical retrieval results for a particular queried item. The retrieval process includes the execution times for loading the database features from the disc to the computer memory, calculating the similarity measure between the query item and the database items and finally ranking the results. Both NQ (single stage) and PQ (multiple stages) would take the same amount of time to execute the first two tasks; while, PQ will execute the third task (i.e., ranking the results) faster than NQ since at each PSQ, fewer items are compared. In the limit, as the number of data partitions approaches the total number of items in the database, say N, ranking would require order N operations.

In order to verify this expectation, several audio PQ retrieval experiments in Real World database have been performed with different $t_p$ values. In order to get an unbiased measure, the experiments for each $t_p$ value are repeated 5
times and the median from 5 overall retrieval times is taken into account. \( PQ \) total execution time (overall retrieval time) and the number of \( PSQ \) updates are plotted in Figure 11. The same experiment is repeated for visual \( PQ \) operation and the result is shown Figure 12. Note that if \( PQ \) is completed with only one \( PSQ \) (corresponding to the longest \( PQ \) period in Figure 11 and 12), then it basically performs a \( NQ \) operation and therefore \( NQ \) retrieval time can also be examined in both figures.

Figure 11: Aural \( PQ \) Overall Retrieval Time and \( PSQ \) Number vs. \( PQ \) Period.

Figure 12: Visual \( PQ \) Overall Retrieval Time and \( PSQ \) Number vs. \( PQ \) Period.
It is also observed that the real $PSQ$ retrieval times are close neighborhood (i.e. $\delta t_w < 0.5 \text{ sec}$) of $t_p = 5 \text{ sec}$ (user-defined period) value. One typical example showing $PSQ$ arrival times for the $PQ$ example shown in Figure 9 is plotted in Figure 13.

![PSQ and PQ retrieval times for the sample retrieval example given in Figure 9.](image)

**Figure 13:** $PSQ$ and $PQ$ retrieval times for the sample retrieval example given in Figure 9.

### 4. CONCLUSIONS

In this paper we proposed a simple, yet efficient query technique, called Progressive Query ($PQ$). $PQ$ is particularly suited to large-scale multimedia databases. We believe that the user interaction with the active query process has the utmost importance and provides significant advantages in terms of user (and CPU) time and I/O accesses. Therefore, $PQ$ is primarily developed to provide periodic and faster retrievals along with the ongoing query process. By this way the user can get an idea about the current status of the query, immediately evaluates the available retrieval results and if satisfactory results are already achieved, the user can then halt the query process without wasting further time. We confirm this with a significant number of experiments.

As an efficient retrieval technique, $PQ$ can be especially useful for databases lacking an efficient indexing structure. Naturally the exhaustive search based query operation, $NQ$, on such databases requires significant and sometimes even infeasible retrieval times. Approvingly, an important objective achieved with the proposed $PQ$ technique is that it avoids
such implementation drawbacks, which NQ encounters. Experimental results show that PQ is not affected from such drawbacks and currently has no limitations whatsoever to the system configuration.

PQ can also be applied to indexed databases. Once a query path is formed within the indexing structure, the most relevant items can be retrieved first. Moreover, we can foresee that PQ can provide a dynamic basis for relevance feedback in such a way that the user can employ relevance feedback during the query process and the incoming progressive retrievals can achieve a better retrieval performance. Since this is a natural consequence of user interaction capability, it was the primary inspiration of developing an interactive query technique such as PQ. In our future work, we plan to develop an efficient, self-learning and MAM-based indexing structure, which is designed to provide optimal performance for PQ along with an effective relevance feedback mechanism under MUVIS framework.

5. REFERENCES


[31] Virage. [URL:www.virage.com](http://www.virage.com)
