Spatio-Temporal Filtering for Images and Videos

Swati Sharma

Computer Science Department,
Suresh Gyan Vihar University
Jaipur, India
swatisharma1123@gmail.com

Abstract— **Digital images and videos are compressed by removing their spatial, temporal and visual redundancies. Although the encoded signals are more compact and easier for storing or transmitting, the correlation between their pixels is distorted. This causes coding artifacts, which degrade the visual quality of the signals and cause annoyance to the viewers. Compressed images and video sequences should be enhanced prior to being sent to the displaying devices. Previous methods on quality enhancement focused on separately improving the quality of each frames. In this way, the temporal consistency between frames is not guaranteed. Furthermore, characteristics of temporal artifacts are not thoroughly studied and exploited for artifact removal. This dissertation investigates these characteristics and proposes novel methods to reduce both spatial and temporal artifacts. The paper covers three main topics in spatio-temporal filtering: quality enhancement, coding and data pruning. The dissertation starts with analyzing the usage of further information from surrounding frames beside the information in the current frames.**

INTRODUCTION

For millennia, people searched for different ways to record their experiences: writing and antiquities helped the posterity imagine their ancestor's life and journey while music and poems helped the descendants sympathize with their ancestor's feeling. Though those tools were useful, it was only when modern still images and motion pictures were discovered that life could be visualized and reflected in the most truthful way. Entertainingly, synthetic photos and animated movies even raise the human imagination to illusive life. The word 'image' is defined as 'an artificial resemblance either in painting or sculpture'. Being an artificial resemblance, it should be produced in a suitable form in order to meet the need of the human being. In the recent digital era, the aim is to map the multidimensional signal to an efficient representation with high originality and condensedness. Fig. 1.1 presents a conventional image and video system. It lays out the processes from noisy signal (after recording) to processed signal (prior to displaying).
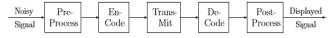

Fig. 1.1 Block Diagram of an Image and Video Processing System

First, the noisy signal is denoised in the Pre-Process phase. Then the Encode phase converts the signal to a more compressed form by removing spatial, temporal and visual redundancies. Because the noise is removed from the noised signal in Pre-Process phase, the new compressed form does not contain the noise representation and is thus more compact, compared to the compressed form without using Pre-Process phase. The new formed signal is transferred over the channel, which is symbolized by the Transmit phase. Being reconstructed into the displayable form, the encoded form is de-compressed by the Decode phase. Although redundancy removal helps reducing the number of bits needed to represent the video content, it destroys the correlation between pixels and causes coding artifacts. Post-Process phase is used to reduce these coding artifacts, as well as adjust the processed signal characteristic such as its size and frame rate into a desired form for displaying.

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts. True Type 1 or Open Type fonts are required. Please embed all fonts, in particular symbol fonts, as well, for math, etc.

I. SPECIFIC TECHNOLOGICAL RECOMMENDATIONS

The aim of this dissertation is to discuss and propose novel systems which can reduce coding artifacts in compressed images and video sequences. The system modification is not limited to the Post-Process phase only.

Block-based compressed signals suffer from blocking, ringing, mosquito and flickering artifacts, especially at low-bit-rate coding. Blocking artifacts occur at the border of neighboring blocks when each frame is processed independently in separate blocks with coarse quantization of discrete cosine transform (DCT) coefficients.

A. Topics to be Considered

1) *Information Extraction from Spatial Data Sources:* Useful spatial and temporal database systems applications are predicated on access to reliable and up-to-date source

data. However, while disk storage availability, CPU power and techniques for interpreting the semantic content of many forms of data have advanced, techniques for extracting spatial information have lagged behind. Moreover, the modelling of spatial data sources is as important as the data themselves if such data are to be interpreted correctly and the development of appropriate models for spatial data streams is an open problem. This lack of data and data model are together rendering many good ideas and prototype systems impractical.

2) *Models that survive the real-world:* The implementation of systems in the real-world is difficult and many systems have proven impractical for rea- sons such as:

- Minor but significant discrepancies between the real-world and the modelled environment (for ex- ample, road closures or changes in museum opening times)
- The loss of mobile communication coverage for systems that rely on access to central servers
- A reliance on particular system configurations or, more importantly, a particular pattern of behaviour of the user
- An inability to self-optimize for different temporal cycles and significant changes in spatial density
- Trade-offs between complexity and performance for real-world models.

3) *Terrain Models:* Terrain can be considered the step-child of spatial and temporal database systems research. Many existing spatial models do not translate well to terrain and there is a feeling that a fresh start is needed [1]. Of particular importance here is to re-conceptualize the ontological properties of terrain and the operations that should be available over such models.

4) *'Natural'/Usable User Interfaces and the Visualization of Spatio-Temporal Data:* Implementation of many spatial and temporal database systems to date has demonstrated the limitations of existing user interfaces in real-world situations. The standard WIMP interface has significant drawbacks and ideas such as augmented reality systems with head-up displays and magic wands [2] that can be used to point at real-world objects would be useful devices. Coupled with this is the need for research into effective techniques for visualizing spatio-temporal data in the context of both static and animated graphics/maps.

5) *Spatio-temporal data mining:* Data mining and knowledge discovery have become popular fields of research. A significant subset of this re- search is looking at the particular semantics of space and time and the manner

in which they can be sensibly accommodated into data mining algorithms [3]. Most of the work, although certainly not all, can be placed in one of three categories:

- Temporal Association Rule Mining, which aims to detect correlations in transactional and relational data that possess a time component. Some (more limited) work has investigated Spatial Association Rule Mining.
- Spatial Clustering, which aims to group similar objects into the same cluster while grouping dissimilar objects into different clusters. In this case similarity is influenced by both spatial and aspatial attributes of the objects as well as any obstacles that may exist.
- Time-series Analysis, which aims to detect frequent patterns in the values of an attribute over time.

Significantly, most of this work deals with one or the other of spatial or temporal semantics, with very few handling both.

6) *Spatio-Temporal Applications for Mobile, Wireless, Location-Aware Services and Sensor Networks:* Two technologies that have emerged recently offer substantial scope for application development – wireless-based, location-aware devices and networks of sensors [4]. While researchers should avoid particular technologies in this area (such as relying on a particular network protocol) there is substantial scope for systems that lever- age the mobility of the user in a way not possible be- fore. Such systems will almost inevitably require spatial database support. Sensor networks and streaming data are very hot topics right now and the field is rapidly developing. It will take an effort to keep up with developments in this area.

7) *Spatio-Temporal Modelling as a Network of Cells:* The recent papers [5] and [6] use fixed-boundary cells to model spatio-temporal data. This may be a 'back- to-the-future' solution as certainly the use of fixed grids predated R-trees. R-trees are often chosen for indexing spatial and spatio-temporal data because the code is available and the method is well-understood. However, the overlapping nature of R-trees, which causes backtracking in search, may not be best suited for every application. There may be many situations where non- overlapping search structures are both more natural and more efficient.

8) *Spatio-Temporal Vacuuming:* While disk storage cost is decreasing and data storage capacity is increasing, there is still a need to delete obsolete data. Vacuuming is used to delete data that is no longer of interest and has been investigated in terms of temporal databases [7]. Extending this to the development of models of obsolescence and to cater for spatial and spatio-temporal data is still required.

9) *Unconventional Spatio-Temporal Access Methods:*
Unconventional problems call for fresh approaches. As an example, much of the work on moving objects in networks (see for example [8] and [9]) uses graphs to model the spatial area instead of using Euclidean space representations. Road distances may be used rather than Euclidean distances. The access method must match the problem. Research on spatio-temporal access methods has mainly focused on two aspects:

- Storage and retrieval of historical information
- Future prediction. Several indexes, usually based on multi-version or 3-dimensional variations of R-trees, have been proposed towards the first goal, aiming at minimizing storage requirements and query cost. Methods for future prediction assume that, in addition to the current positions, the velocities of moving objects are known. The goal is to retrieve the objects that satisfy a spatial condition at a future timestamp (or interval) given their present motion vectors (e.g., 'based on the current information, find the cars that will be in the city center 10 minutes from now'). The only practical index in this category is the TPR-tree and its variations (also based on R-trees). Despite the large number of methods that focus explicitly on historical information retrieval or future pre- diction, currently there does not exist a single index that can achieve both goals. Even if such a 'universal' structure existed (e.g., a multi-version TPR-tree keeping all previous history of each object), it would be inapplicable for several update-intensive applications, where it is simply infeasible to continuously update the index and at the same time process queries. For instance, an up- date (i.e., deletion and re-insertion) in a TPR-tree may need to access more than 100 nodes, which means that by the time it terminates its result may already be out- dated (due to another update of the same object). Even for a small number of moving objects and a low up- date rate, the TPR-tree (or any other index) cannot 'follow' the fast changes of the underlying data. Consequently, main-memory structures seem more appropriate for update-intensive applications.

10) *Novel Query Types and Space/Time Queries:* Jim Gray gave a SIGMOD Record interview [10] that provided a context for thinking about 'management of data' beyond the context of SQL. For database researchers, there is more to life than SQL—there are very interesting problems that can only be solved with loosely-coupled system architectures (i.e., web services, grid computing) and that

there are other kinds of transactions beyond the classic Codd short transaction. Note that this interview also provides insight into finding interesting research topics for fledgling researchers.

An example of novel query type in spatio-temporal databases refers to continuous queries whose result is strongly related to the temporal context. An example of a continuous spatio-temporal query is: 'based on my current direction and speed of travel, which will be my nearest two gas stations for the next 5 minutes?' A result of the form $< \{A, B\}, [0, 1) >, < \{B, C\}, [1, 5) >$ would imply that A, B will be the two nearest neighbors during interval [0, 1), and B, C afterwards. Notice that the corresponding instantaneous query ('which are my nearest gas stations now?') is usually meaningless in highly dynamic environments; if the query point or the database objects move, the result may be invalidated immediately. Any spatial query has a continuous counterpart whose termination clause depends on the user or application needs. Consider, for instance, a window query, where the window (and possibly the database objects) moves/changes with time. The termination clause may be temporal (for the next 5 minutes), a condition on the result (e.g., until exactly one object appears in the query window, or until the result changes three times), a condition on the query window (until the window reaches a certain point in space) etc. A major difference from continuous queries in the context of traditional databases is that, in case of spatio-temporal databases, the object's dynamic behavior does not necessarily require updates, but can be stored as a function of time using appropriate indexes. Furthermore, even if the objects are static, the results may change due to the dynamic nature of the query itself (i.e., moving query window), which can be also represented as a function of time. Thus, a spatio-temporal continuous query can be evaluated instantly (i.e., at the current time) using time-parameterized information about the dynamic behavior of the query and the database objects, in order to produce several results, each covering a validity period in the future [11].

11) *Moving Object Tracking Research:* Data logging and visualization is a simple problem, how- ever, the development of analytical techniques for integrating moving and static datasets is necessary. One example can be found in [11].

12) *Approximate Queries:* In several spatio-temporal applications, the size of the data and the rapidity of updates necessitate approximate query processing. For instance, in traffic supervision systems, the incoming data are usually in the form of data streams (e.g., through sensors embedded on the road network), which are potentially unbounded in size. Therefore, materializing all data is unrealistic. Further-more, even if all the data were stored, the size of the index

would render exact query processing very expensive since any algorithm would have to access at least a complete path from the root to the leaf level. Finally, in several applications the main focus of query processing is retrieval of approximate summarized information about objects that satisfy some spatio-temporal predicate (e.g., 'the number of cars in the city center 10 minutes from now'), as opposed to exact information about the qualifying objects (i.e., the car ids), which may be unavailable, or irrelevant.

13) *Parallel Spatio-Temporal Algorithms:* With the overwhelming volumes of spatio-temporal data available in the real-world, the importance of parallel algorithms that utilize coarse-grained parallel processing environments remains undiminished. However, Jim Gray has noted in [14] that because of a number of changes in hardware and software and caching algorithms, bigger and bigger computers have been built. So one must be cautious in assuming that today's larger machines cannot solve the problem.

14) *Uncertainty:* Uncertainty is inherent is most spatio-temporal applications due to measurement/ digitization errors and missing or incomplete information. Assume, for instance, a user with a PDA inquiring about the closest restaurant in terms of road network distance. Although the user may actually be on a road segment, due to the inaccuracy of the GPS device, the system may fail to recognize this. Such a situation can be handled by defining an (application-dependent) threshold dT, so that if a point is within distance dT from a road segment, it is assumed to lie on it. Alternatively, we can snap the point to the closest road assuming incomplete information (e.g., an un-recorded alley), or we can consider it unreachable depending on the application specifications. Similar problems exist for object trajectories because while movement is continuous, measurements are discrete.

II. METHOD

Many filter-based de-noising methods have been proposed to reduce coding artifacts, most of which are frame-based enhancements for blocking and ringing artifact reduction. One method to reduce blocking artifacts is to use the lapped orthogonal transform (LOT), which increases the dependence between adjacent blocks. As the LOT based approach is incompatible with the JPEG standard, many other approaches consider pixel-domain and DCT-domain post-processing techniques. These include low-pass filtering, adaptive median filtering and nonlinear spatially variant filtering which were applied to remove the high frequencies caused by sharp edges between adjacent blocks. Other pixel-based methods are constrained least squares (CLS), and maximum a posteriori probability approach

(MAP), all of which require many iterations with high computational load. In, a projections onto convex set (POCS) based method was proposed with multi-frame constraint sets to reduce the blocking artifacts. This method required extracting the motion between frames and quantization information from the video bit-stream. Xiong et al. considered the artifact caused by discontinuities between adjacent blocks as quantization noise and used over complete wavelet representations to remove this noise. The edge information is preserved by exploiting cross-scale correlation among wavelet coefficients. The algorithm yields improved visual quality of the decoded image over the POCS and MAP methods and has lower complexity. In H.264/AVC, an adaptive de-blocking filter was proposed to selectively filter the artifacts in the coded block boundaries. Substantial objective and subjective quality improvement was achieved by this de-blocking filter.

The DCT-based methods for blocking artifact removal adjust the quantized DCT coefficients to reduce quantization error. Tien and Hang made an assumption that the quantization errors are strongly correlated to quantized coefficients in high contrast areas. Their algorithm first compares the DCT coefficients to the pre-trained quantized coefficient representatives to get the best match, then adds the corresponding quantized error pattern to reconstruct the original DCT coefficient. This method requires a large pre-defined set of quantized coefficient representatives and provides slight PSNR gain. In another approach, Jeon and Jeong defined a block boundary discontinuity and compensated for selected DCT coefficients to minimize this discontinuity. To restore the stationarity of the original image, Nosratinia averaged the original decoded image with its displacements. These displacements are calculated by compressing, decompressing and translating back shifted versions of the original decoded images. With an assumption of small changes of neighboring DCT coefficients at the same frequency in a small region, Chen et al. applied an adaptively weighted low-pass filter to the transform coefficients of the shifted blocks. The window size is determined by the block activity which is characterized by human visual system (HVS) sensitivity at different frequencies. A new type of shifted blocks across any two adjacent blocks was constituted by Liu and Bovik [12]. They also defined a blind measurement of local visibility of the blocking artifact. Based on this visibility and primary edges of the image, the block edges were divided into three categories and were processed by corresponding effective methods.

To reduce ringing artifacts, Hu et al. used a Sobel operator for edge detection, then applied a simple low-pass filter for pixels near these edges. Using a similar process, Kong et al. established an edge map with smooth, edge and texture

blocks by considering the variance of a 3×3 window centered on the pixel of interest. Only edge blocks are processed with an adaptive filter to remove the ringing artifacts close to the edges. Oguz et al. detected the ringing artifact areas that are most prominent to HVS by binary morphological operators. Then a gray-level morphological nonlinear smoothing filter is applied to these regions. Although lessening the ringing artifacts, these methods do not solve the problem completely because the high frequency components of the resulting images are not reconstructed. As an encoder-based approach, proposed a noise shaping algorithm to find the optimal DCT coefficients which adapts to the noise variances in different areas. All of these methods can only reduce ringing artifacts in each frame. To deal with the temporal characteristic of mosquito artifacts, applied the spatio-temporal median filter in the transform domain for surrounding 8×8 blocks. The improvement in this case is limited by the small correlation between DCT coefficients of the spatial neighboring 8×8 blocks as well as the lack of motion compensation in the scheme.

For flickering artifact removal, most of the current methods focused on reducing flickering artifacts in all intra-frame coding. In the quantization error is considered to obtain the optimal intra prediction mode and to help reducing the flickering artifact.

III. COPYRIGHT FORMS

You must submit the IEEE Electronic Copyright Form (ECF) as described in your author-kit message. THIS FORM MUST BE SUBMITTED IN ORDER TO PUBLISH YOUR PAPER.

ACKNOWLEDGMENT

REFERENCES

[1] W.R. Franklin. Computational and geometric cartography. In Keynote at GIScience 2002, Boulder, CO, USA, 2002.
http://www.ecse.rpi.edu/Homepages/wrf/research/giscience2002/.)

[2] M.J. Egenhofer and W. Kuhn. Beyond desktop GIS.
In GIS PlaNET, Lisbon, Portugal, 1998.
http://www.spatial.maine.edu/~max/ eyondDesktopGIS.pdf.

[3] J.F. Roddick, K. Hornsby, and M. Spiliopoulou. YABTSSTDMR- yet another bibliography of temporal, spatial and spatio-temporal data mining research.
In K.P. Unnikrishnan and R. Uthurusamy, eds, *SIGKDD Temporal Data Mining Workshop*, pages 167–175, San Francisco, CA, 2001. ACM.

[4] S. Nittel, A. Stefanidis, I. Cruz, M. Egenhofer, D. Goldin, A. Howard, A. Labrinidis, S. Madden, A. Voidard, and M. Worboys.
Report from the first workshop on geo-sensor networks. SIGMOD Record, 33(1), 2003

[5] V.P. Chakka, A.C. Everspaugh, and J.M. Patel. Indexing large trajectory data sets with SETI. CIDR, 2003.

[6] M. Hadjieleftheriou, G. Kollios, D. Gunopulos, and V. Tsotras. On-line discovery of dense areas in spatio-temporal databases. Pages 306–324.

[7] C.S. Jensen. Vacuuming. In R.T. Snodgrass, ed, *The TSQL2 Temporal Query Language*, pages Chapter 23, pp. 451–462. Kluwer Academic Pub- lishing, Boston, 1995.

[8] D. Papadias, J. Zhang, N. Mamoulis, and Y. Tao. Query processing in spatial network databases. VLDB, 2003.

[9] C. Jensen, J. Kolar, T. B. Pedersen, and I. Timko. Nearest neighbor queries in road networks. GIS, 2003.

[10] M. Winslett. Distinguished database profiles: Interview with Jim Gray. SIGMOD Record, 32(1), 2003.

[11] Y. Tao, D. Papadias, and Q. Shen. Continuous nearest neighbor search. VLDB, 2002.

[12] X. Fan, W. Gao, Y. Lu, and D. Zhao, "Flicking Reduction in All Intra Frame Coding," Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-E070, October 2002.

[13] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, pp. 560–576, July 2003.

[14] H.S. Malvar and D.H. Staelin, "The LOT: Transform Coding without Block- ing Effects," IEEE Trans. Acoust., Speech, Signal Processing, vol. 37, pp. 553–559, April 1989.

[15] T. Jarske, P. Haavisto, and I. Defee, "Post-Filtering Methods for Reducing Blocking Effects from Coded Images," IEEE Trans. Cosumer Electronics, vol. 40, pp. 521–526, August. 1994.

[16] R.L. Stevenson, "Reduction of Coding Artifacts in Transform Image Coding," Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, vol. 5, pp. 401–404, April 1993.

[17] B. Jeon and J. Jeong, "Blocking Artifacts Reduction in Image Compression with Block Boundary Discontinuity Criterion," IEEE Trans. Circuits Syst. Video Technol., vol. 8, pp. 345–357, June 1998.

[18] A. Nosratinia, "Embedded Post-Processing for Enhancement of Compressed Images," Proc. IEEE Data Compression Conf., pp. 62–71, 1999.

[19] T. Chen, H.R. Wu, and B. Qiu, "Adaptive Postfiltering of Transform Coef- ficients for the Reduction of Blocking Artifacts," IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 594–602, May. 2001.

[20] S. Liu and A.C. Bovik, "Efficient DCT-Domain Blind Measurement and Re- duction of Blocking Artifacts," IEEE Trans. Circuits Syst. Video Technol., vol. 12, pp. 1139–1149, December. 2002.