An Optimal Content-Based Pattern Generation Algorithm

Manoranjan Paul, *Member, IEEE*, and Manzur Murshed, *Member, IEEE*

Abstract—Very low bit-rate video coding algorithms using predefined regular-shaped patterns to segment out moving objects at macroblock level have exhibited good potential for improved coding efficiency when embedded in the H.264 standard as an extra mode. Even the best-matched regular-shaped pattern from a predefined codebook cannot approximate the shape of the object well, and there is no guarantee that even a regular-shaped object will have a close match with one of the limited number of predefined patterns. Intuitively, improved coding performance can be achieved if patterns are dynamically extracted from the video content. This letter presents a content-based pattern generation (CPG) algorithm for a set of macroblocks, which is shown optimal when only one pattern is allowed to represent the entire set. Coupling CPG, generating a pattern codebook after clustering the macroblocks into several disjoint sets, with any pattern selection algorithm outperforms the existing regular-shaped pattern-based coding while both embedded in H.264.

Index Terms—Arbitrary object shape, moving region, pattern matching, video coding.

I. INTRODUCTION

PATTERN-BASED video coding (PVC) [1] offers improvement at low bit rate by exploiting the principle of the MPEG-4 standard in partitioning a 16×16 -pixel macroblock (MB) containing some part of a moving object by a simplified segmentation process that avoids handling the exact shape of the object, so that popular block-based motion estimation techniques could be applied. It focuses on the moving region (MR), a binary representation of the moving pixels in an MB, through the use of a codebook of 64-pixel predefined regular-shaped pattern templates. Under a well-defined *similarity metric* [3], if the MR of an MB is well matched by a particular pattern, the MB can be coded by considering only the 64 pixels segmented by that pattern using motion compensated residual coding with the remaining 192 pixels being skipped as static background. Successful pattern matching can, therefore, theoretically attain a maximum compression ratio of 4:1 for an MB. The actual compression, however, will be lower due to the overheads of identifying this special type of MB as well as the best-matched pattern for it and the *matching error* for approximating the MR with the pattern. While avoiding MBs with too many ($\gg 64$) moving pixels [1] and extending the pattern codebook

[2] can reduce the matching error significantly to outperform the existing H.264 standard [4] (when PVC is embedded as an extra mode in H.264) at low bit rate, the general idea of using a fixed number of predefined regular-shaped patterns poses some obstacles in improving the coding efficiency even further.

An MR can be of any arbitrary shape as it comprises part of an object. So, even the best-matched regular-shaped pattern from the predefined codebook can only loosely approximate the shape of the MR. Moreover, there is no guarantee that even a regular-shaped MR will have a close match with one of the limited number of predefined patterns. Extending the codebook size further is not a viable option as the obvious increase in pattern identifier length outweighs any matching error reduction through better shape approximation.

Using a limited number of arbitrary shaped patterns that are customized on the content of the image sequence can potentially remove the above-mentioned obstacles from PVC. In this letter, an optimal content-based pattern generation (CPG) algorithm is proposed that generates an arbitrary shaped pattern set from the content of given sets of MRs such that if only one pattern were allowed to represent all the MRs derived from the corresponding set, the generated pattern would give the optimum similarity measure for the entire set.

In order to generate enough patterns to cover the entire MB area while keeping less overlapping among them, all the MRs of an image sequence are first divided into a predefined number of sets using an efficient clustering algorithm on the basis of the gravitational centers (GCs) of the regions, and then, a pattern is generated for each set. While the generated pattern guarantees the optimal similarity within each cluster, when only one pattern is allowed to represent all the MRs in corresponding cluster, similarity measure over all the MRs can still be improved if each MR is allowed to select the best-matched pattern from the generated codebook. Although the extent of such an improvement could be limited with efficient clustering, it is worthwhile to develop a two-phase arbitrary shaped PVC (ASPVC) process. In the first phase, a customized *pattern codebook* (PC) is generated using the CPG algorithm. In the second phase, this customized codebook is used with any existing best pattern selection algorithm. Experimental results show that ASPVC outperforms the existing PVC with predefined regular-shaped patterns when each embedded into H.264 as an extra mode.

II. CONTENT-BASED PATTERN GENERATION

The MR M of an MB C in the current frame is obtained using the co-located MB R in the reference frame as follows:

$$M(x, y) = T(|C(x, y) \bullet B - R(x, y) \bullet B|) \quad 0 \leq x, y \leq 15 \quad (1)$$

Manuscript received April 3, 2007; revised June 17, 2007. This work was supported by the Australian Research Council's Discovery Projects scheme under Project Number DP0666456. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Weifeng Su.

The authors are with the Gippsland School of Information Technology, Monash University, Churchill, Vic 3842, Australia (e-mail: Manoranjan.Paul@infotech.monash.edu.au; Manzur.Murshed@infotech.monash.edu.au).

Digital Object Identifier 10.1109/LSP.2007.904703

where B is a 3×3 unit matrix for the morphological closing operation $\bullet$ [5], which is applied to reduce noise, and the thresholding function $T(v) = 1$ if $v > 2$ and 0 otherwise. Let $|Q|_\ell$ be the total number of ℓ 's in the matrix Q . If $8 \leq |M|_1 < 2QP/3 + 64$, where QP is the quantization parameter, the corresponding MB, i.e., C , is defined as a candidate active-region MB (CRMB) as it has a reasonable number of moving pixels to be covered by a 64-pixel pattern, thus avoiding high matching error.

Once all such CRMBs are collected for a certain number of consecutive frames, decided by the rate-distortion optimizer [6] when the rate-distortion gain outweighs the overhead of encoding the shape of new patterns, these are divided into λ sets to generate λ patterns. In order to generate patterns with minimal overlapping, the global optimization algorithm will be computationally intractable. A simpler alternative lies in the greedy heuristic where these CRMBs are divided into λ clusters such that the average distance among the gravitational centers of CRMBs within a cluster is small, while the same among the centers of CRMBs taken from different clusters is relatively large, where the gravitational center of any MR M is calculated as

$$GC(M) = \left(\sum_{x=0}^{15} \sum_{y=0}^{15} xM(x,y), \sum_{x=0}^{15} \sum_{y=0}^{15} yA(x,y) \right) / |M|_1. \quad (2)$$

One pattern is now generated for each cluster. In order to generate this pattern from $^{256}C_{64}$ possibilities, coding efficiency must be considered. In Fig. 1, a generic CPG algorithm is presented where a μ -pixel pattern for each cluster is generated by the μ -most-frequent pixels among all the CRMBs in the cluster. An example of this pattern generation technique is detailed in Fig. 2. Fig. 2(a) shows the total number of 1's of all CRMBs in a cluster for each pixel position where the white region indicates the most frequent pixels i.e., the most of the CRMBs have MRs in this area; on the other hand, the black region indicates the least frequent pixels, i.e., a few of the CRMBs have MRs in this area. The pattern generated by the 64-most-frequent pixels is shown in Fig. 2(c).

In the first phase of the ASPVC algorithm, a customized PC is formed with all these λ content-based patterns to code all the CRMBs used to generate these patterns. This PC is then encoded in the bitstream, which is not necessary with any predefined PC as the shapes and indices of the patterns are already known at the decoding end. In the second phase, each of these CRMBs is encoded using the best-matched pattern from this PC. We have observed that PC is refreshed more frequently for high motion video sequences and bit rates, i.e., under such conditions, fewer frames are used to collect the CRMBs.

The following theorem proves an optimality issue for the patterns generated by the CPG algorithm.

Theorem 1: If only one pattern is allowed to represent the MRs of a set of CRMBs, the pattern generated by the CPG algorithm on this set results in the optimal similarity measure.

Proof: Let $M_{i,j}$ be the MR of the j th CRMB in cluster C_i and P_i be the pattern generated by the CPG algorithm to repre-

Algorithm PC = CPG(λ, μ, C)

Precondition: Given a set of CRMBs C ;

Postcondition: A pattern codebook $PC = \{P_1, \dots, P_\lambda\}$ of μ -pixel content-based patterns;

1. Divide C into λ clusters $C_1, \dots, C_\lambda$ using any clustering method based on the gravitational centers of the CRMBs.
2. FOR $i = 1, \dots, \lambda$
3. FOR $x = 0, \dots, 15$
4. FOR $y = 0, \dots, 15$
5. $P_i(x,y) = 0$;
6. $T_i(16x + y) = f_i(x,y) = \sum_{j=1}^{|C_i|} M_{i,j}(x,y)$ where $M_{i,j}$ is the MR of the j th CRMB in C_i ;
7. $\{l_0, \dots, l_{255}\} = \text{ranked indices of } T_i \text{ such that } T_i(l_j) \geq T_i(l_{j+1}) \text{ for } 0 \leq j < 255$;
8. FOR $j = 0, \dots, \mu - 1$
9. $P_i(\lfloor l_j / 16 \rfloor, l_j \bmod 16) = 1$;

Fig. 1. Content-based pattern generation algorithm.

sent this cluster. Consider the well-defined similarity metric in [3] where dissimilarity of an MR M from a pattern P is defined as $|M|_1 - |M \wedge P|_1$, which measures the number of moving pixels not covered by the pattern. Now consider the following similarity optimization problem:

$$\left. \begin{aligned} & \text{minimize}_{\text{for } P} \sum_{\forall j} (|M_{i,j}|_1 - |M_{i,j} \wedge P|_1) \\ & \Leftrightarrow \text{maximize}_{\text{for } P} \sum_{\forall j} |M_{i,j} \wedge P|_1 \\ & \Leftrightarrow \text{maximize}_{\text{for } P} \sum_{\forall j} \sum_{\forall x} \sum_{\forall y} (M_{i,j}(x,y) \wedge P(x,y)) \\ & \Leftrightarrow \text{maximize}_{\text{for } P} \sum_{\forall j} f_i(x,y) P(x,y) \end{aligned} \right\} \quad (3)$$

where $f_i(x,y) = \sum_{\forall j} M_{i,j}(x,y)$ is the frequency of pixel (x,y) in the MRs of cluster C_i . Let $\{l_0, \dots, l_{255}\}$ be the ranked indices of 256 pixels such that $f_i(\lfloor l_j / 16 \rfloor, l_j \bmod 16) \geq f_i(\lfloor l_{j+1} / 16 \rfloor, l_{j+1} \bmod 16)$. As a μ -pixel pattern has exactly μ "1"-valued pixels, the maximum value of $\sum_{\forall j} f_i(x,y) P(x,y) = \sum_{j=0}^{\mu-1} f_i(\lfloor l_j / 16 \rfloor, l_j \bmod 16)$. Hence, P_i satisfies the above optimization problem. ■

Note that when there exist $a < \mu$ and $b \geq \mu$ such that $f_i(\lfloor l_a / 16 \rfloor, l_a \bmod 16) = \dots = f_i(\lfloor l_b / 16 \rfloor, l_b \bmod 16)$, there will be $b-a+1$ $C_{\mu-a}$ possible distinct patterns generated by the CPG algorithm, each satisfying the above optimization problem.

To cover the entire 16×16 -pixel MB, we need at least four 64-pixel patterns as an MR can occupy any part of the MB. However, using only four patterns can potentially leave many MRs unsuitable to be represented by any of these patterns and hence leading to inefficiency. Contrary to predefined PC, the number of content-based patterns cannot be too high due to the overhead of encoding the shape of the patterns. A tradeoff can be achieved with eight 64-pixel patterns that are also observed empirically for all types of image sequences.

In order to find the computational overhead of the proposed ASPVC scheme, let us compare it with H.264 standard while encoding only the CRMBs. H.264 encodes each CRMB with motion compensation requiring one motion search for each mode. When ASPVC is embedded in H.264 as an extra mode, an ad-

Fig. 2. (a) 3-D representation of pixel frequency of one of the eight clusters of CRMBs obtained for *Miss America* sequence. (b) Its corresponding 2-D top view projection. (c) Generated pattern for this cluster by the CPG algorithm.

ditional one-fourth motion search is required per CRMB as pattern size is quarter of an MB. In addition, each CRMB also takes

Fig. 3. Eight 64-pixel patterns generated by the content-based pattern generation algorithm on six standard test video sequences.

part in the CPG algorithm exactly once, and the best pattern is selected at the end. Detailed analysis of the CPG and full search motion estimation algorithms reveals that even if 100 iterations are allowed with K -means clustering algorithm [7], PC generation along with pattern selection require no more than 5% (20%) operations needed by single motion estimation with full search algorithm having ± 15 (± 7) search width. Considering the seven modes used in H.264, computational overhead of ASPVC can be as low as 4.5% $((25\% + 5\%)/700\%)$, which is negligible.

Fig. 3 shows PC of eight 64-pixel patterns generated by the CPG algorithm on six standard test video sequences. Benefit of content-based pattern generation is clearly evident by the lack of similarity among the PCs, i.e., no predefined PC would be able to provide satisfactory coding efficiency for all these six sequences. As expected, the patterns within each PC cover nearly non-overlapping regions of an MB to ensure the maximum coverage of potential MRs. It is also interesting to note that each of the generated patterns is boundary-adjointed and clustered. This observation could appear striking as no such condition has been embedded in the CPG algorithm. These properties are mainly due to the fact that an MR signifies some part of a moving object covering several adjacent MBs, and thus, by nature, it has to be boundary-adjointed and clustered. While predefined regular patterns in [2] also fulfill these properties, relaxation of the shape in the CPG algorithm certainly gives the ultimate edge over the predefined regular-shape patterns in achieving superior coding efficiency.

III. SIMULATION RESULTS

In this paper, experimental results in Fig. 4 are presented using the first 100 frames of only four standard gray-scale video test sequences, due to space limitation, comprising I - and P -type frames of QCIF digital video format [8]. Full-search motion estimation with half-pel refinement and H.264 recommended “baseline” profile were employed to obtain the encoding results for standalone H.264, and ASPVC and PVC embedded in it as extra modes. We used the K -means clustering routine in MATLAB in the CPG algorithm. Fig. 4 confirms that ASPVC improved image quality by 0.50–1.20 dB and 0.25–0.50 dB compared to H.264 and PVC, respectively. The performance of ASPVC is relatively better for low motion video sequences such as *Miss America* and *Claire* due to large number of CRMBs.

Fig. 4. Rate-distortion performance on standard video sequences (a) *Miss America*; (b) *Claire*; (c) *News*; and (d) *Foreman*.

IV. CONCLUSION

A content-based pattern generation algorithm is developed in this letter where a pattern is generated for a set of moving regions by selecting the pixels in order of their frequency in that set. This algorithm has been mathematically proven to be optimal from a coding viewpoint when only one pattern is allowed to represent all the moving regions in that set. The proposed arbitrary shaped pattern-based video coding, where a pattern codebook is generated using the CPG algorithm in the first phase followed by the best pattern selection phase using that codebook, has improved the image quality by 0.5 dB against the existing regular-shaped pattern-based coding while both embedded in H.264.

REFERENCES

- [1] K.-W. Wong, K.-M. Lam, and W.-C. Siu, "An efficient low bit-rate video-coding algorithm focusing on moving regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 10, pp. 1128–1134, Oct. 2001.
- [2] M. Paul, M. Murshed, and L. Dooley, "A real-time pattern selection algorithm for very low bit-rate video coding using relevance and similarity metrics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 753–761, Jun. 2005.
- [3] M. Paul, M. Murshed, and L. Dooley, "A new efficient similarity metric and generic computation strategy for pattern-based very low bit-rate video coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2004, vol. 3, pp. 165–168.
- [4] M. Paul and M. Murshed, *Efficient H.264/AVC Video Encoder Where Pattern is Used as Extra Mode for Wide Range of Video Coding*. New York: Springer-Verlag, 2007, vol. 4352, pp. 353–362, LNC.
- [5] P. Maragos, "Tutorial on advances in morphological image processing and analysis," *Opt. Eng.*, vol. 26, no. 7, pp. 623–632, 1987.
- [6] T. Weigand, H. Schwarz, A. Joch, and F. Kossentini, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–702, Jul. 2003.
- [7] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Mathematical Statistics and Probability*, 1967, vol. 1, pp. 281–297, Univ. California Press.
- [8] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression*. New York: Wiley, 2003.