A NEW EFFICIENT SIMILARITY METRIC AND GENERIC COMPUTATION STRATEGY FOR PATTERN-BASED VERY LOW BIT-RATE VIDEO CODING

Manoranjan Paul, Manzur Murshed, and Laurence Dooley

Gippsland School of Computing and IT, Monash University, Churchill Vic 3842, Australia

E-mail: {Manoranjan.Paul,Manzur.Murshed,Laurence.Dooley}@infotech.monash.edu.au

ABSTRACT

In the context of very low bit-rate video coding, pattern representations of a *moving region* (MR) in block-based motion estimation and compensation has become increasingly attractive. Generally, all existing pattern-matching algorithms apply a *similarity* metric involving elementary operations, to compute the mismatch between a MR and a particular fixed pattern in order to select the best-matching pattern from a fixed-size codebook of predefined patterns. In this paper, an efficient similarity metric together a new generic computation strategy is presented by considering only the mismatch areas of MRs. It is theoretically proven that for a specific MR in a macroblock, the new similarity metric selects exactly the same pattern as existing metrics, while the resulting computational coding efficiency is improved by between 21% and 58% compared with the H.263 low bit-rate coding standard.

1. INTRODUCTION

Reducing the transmission bit-rate while concomitantly retaining image quality continues to be a challenge for efficient very low bit-rate video compression standards, such as H.263 [4]. These standards are however unable to encode moving objects within a 16×16 pixel *macroblock* (MB) during *motion estimation* (ME), resulting in all 256 residual error values being transmitted for *motion compensation* (MC) regardless of whether there are moving objects. One solution is to sub-divide the MB and apply ME and MC to each sub-block. With a sufficient number of blocks, the shape of a moving object can be accurately represented, but this has a high processing expenditure [1].

The MPEG-4 [3] video standard first introduced the concept of content-based coding, by dividing video frames into separate segments comprising a background and one or more moving objects. The pattern-based video coding algorithms in [6]–[8] and [12] exploited the idea of partitioning the MBs, via a simplified segmentation process that avoided handling the exact shape of the moving objects, so popular MB-based ME techniques could be applied. If $C_k(x,y)$ and $R_k(x,y)$ denote the kth MB of the current and reference frames, each of size $W \text{ pixels} \times H \text{ lines}$, respectively of a video sequence, where $0 \leq x, y \leq 15$ and $0 \leq k < W/16 \times H/16$. The moving region $M_k(x, y)$ of the kth MB in the current frame is obtained as follows:

$$M_k(x, y) = T(|\ C_k(x,y) \bullet B - R_k(x,y) \bullet B\ |) \qquad (1)$$

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. As '1' indicates a *moving region* (MR) and '0' the static region of that MB, the total number of '1's is used as a MB classification criterion.
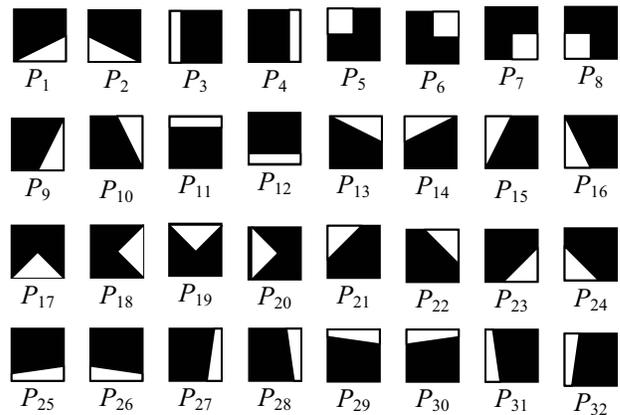

Figure 1: The pattern codebook of 32 regular shaped, 64-pixel patterns, defined in 16×16 blocks, where the white region represents 1 (motion) and black region represents 0 (no motion).

Let $|Q|_\ell$ be the total number of ℓ 's in the matrix Q. Pattern matching algorithms have traditionally classified each MB into three mutually exclusive categories:1) *Static MB* (SMB): MBs containing little or no motion; 2) *Active MB* (AMB): MBs containing moving object(s) with little static background and 3) *Active-Region MB* (RMB): MBs containing both static background and part(s) of moving object(s) such that the *MR* of the block can be considered *similar* enough to a pattern from a *pattern codebook* (PC) of 64-pixel patterns (e.g., P_1–P_{32} in Figure 1). Any MB that cannot be *directly* classified as a SMB ($0 \leq |M|_1 < 8$) or AMB ($128 < |M|_1$) [9], is first identified as a *candidate* RMB (CRMB) and a similarity metric applied to classify it as either a RMB or AMB. The first two MB types are defined in the H.263 standard [4] and treated in exactly the same way, while for the RMB classification, ME and MC is performed only for those MRs covered by a selected pattern from the codebook. Overall, this affords superior prediction and compression efficiency as well as reducing the coding time for smooth motion sequences by on average 32%, compared to H.263.

Classification of an RMB in previous algorithms [6]–[8] and [12] has used a similarity metric to identify significant

overlapping between the MR and the patterns, so the best pattern can be selected to represent the MR. Empirical results in [8] confirm that between 16% and 34% of the total MBs are classified as RMBs for smooth motion sequences [11]. The similarity metric, however, is applied much more often as the number of CRMBs will always be higher. Motion estimation, irrespective of a scene's complexity, typically comprises more than 60% of the processing overhead required to encode an inter picture with a software codec using the DCT [10], when full search is used. A corollary of this is that the computational efficiency of a similarity metric for a CRMB is critical to the overall complexity, since for example, for a codebook size of 32 patterns, the metric represents $\approx$ 55% of the ME time. Hence, any strategy that improves the computational efficiency of the metric concomitantly reduces the overall encoding complexity.

This paper presents a generic computational strategy, which can be embedded into any pattern-based coding scheme. For instance, when applied with an existing similarity metric a reduction of up to 81% in the number of operations is achieved. The paper also presents a new similarity metric, which selects the best-matched pattern by considering only the mismatched area of moving regions instead of the mismatch areas of both moving region and the pattern. The new similarity metric using this criterion requires 22% fewer operations than the existing similarity metric.

This paper is organized as follows. The existing and new similarity metrics are described in Sections 2 and 3 respectively, while the new computation strategy and complexity impact on coding are discussed in Sections 4 and 5 respectively. Some conclusions are presented in Section 6.

2. EXISTING SIMILARITY METRIC

The *dissimilarity* between a pattern P_n and the moving region M of a CRMB was measured in [6]–[8] and [12] as:-

$$S_1(M, P_n) = \sum_{x=0}^{15} \sum_{y=0}^{15} |M(x, y) - P_n(x, y)|. \qquad (2)$$

where $1 \le n \le |PC|$. If $\exists P_n \in PC : S_1(M, P_n) < T_{S_1}$, the CRMB is classified as an RMB and its MR is represented by a pattern P_i such that

$$P_i = \arg \min_{\forall P_n \in PC} (S_1(M, P_n) | S_1(M, P_n) < T_{S_1}) \qquad (3)$$

where T_{S_1} is the predefined *similarity threshold*; otherwise the CRMB is classified as an AMB. The subscript '1' signifies that threshold is dependent on a specific similarity metric.

Lemma 1: $S_1(M, P_n) = |\neg M \wedge P_n \vee M \wedge \neg P_n|_1$
$= |M|_1 + |P_n|_1 - 2|M \wedge P_n|_1$.

Proof. From Table I, it can be shown that $S_1(M, P_n)$
$= \sum_{x=0}^{15} \sum_{y=0}^{15} |\neg M(x, y) \wedge P_n(x, y) \vee M(x, y) \wedge \neg P_n(x, y)|_1$ using *sum of products* of *minterms*. As all three logical operators $\{\neg, \wedge, \vee\}$ work on the corresponding elements of the metrics, relation $S_1(M, P_n) = |\neg M \wedge P_n \vee M \wedge \neg P_n|_1$ holds. Similarly, from

columns 3 and 4 in Table I, relation $|\neg M \wedge P_n \vee M \wedge \neg P_n|_1 = |M|_1 + |P_n|_1 - 2|M \wedge P_n|_1$ holds. ∎

Table I: Equivalence table where M and P_n refer to $M(x, y)$ and $P_n(x, y)$ respectively.

M	P_n	$|M - P_n|$	$|M|_1 + |P_n|_1 - 2|M \wedge P_n|_1$	$|M \wedge \neg P_n|_1$	$|M|_1 - |M \wedge P_n|_1$
0	0	0	0	0	0
0	1	1	1	0	0
1	0	1	1	1	1
1	1	0	0	0	0

3. NEW SIMILARITY METRIC

For this new metric, the dissimilarity of pattern P_n from the moving region M of a CRMB is defined as:-

$$S_2(M, P_n) = |M \wedge \neg P_n|_1 \qquad (4)$$

where $1 \le n \le |PC|$. As in the metric in Section 2, if $\exists P_n \in PC : S_2(M, P_n) < T_{S_2}$, the CRMB is classified as an RMB and its MR is represented by a pattern P_i such that

$$P_i = \arg \min_{\forall P_n \in PC} (S_2(M, P_n) | S_2(M, P_n) < T_{S_2}) \qquad (5)$$

where T_{S_2} is the predefined similarity threshold; otherwise the CRMB is classified as an AMB.

From Table I, since the column 5 and 6 are equivalent, the following Lemma can be proven:

Lemma 2: $S_2(M, P_n) = |M|_1 - |M \wedge P_n|_1$. ∎

The key difference between the new and existing similarity metric (Section 2) is best illustrated by the example in Figure 2(a) for a pattern P_{12} and moving region M. The existing metric considers the two non-overlapping (black) regions shown in Figure 2(b) as the measure of dissimilarity of a CRMB; while the new metric uses only the non-overlapping area of M as shown in Figure 2(c). Formally, both these dissimilarity metrics are expressed as $(M - (P_{12} \cap M)) + (P_{12} - (P_{12} \cap M))$ and $M - (P_{12} \cap M)$ respectively. As the average MR size $((8+128)/2 = 68)$ is comparable to that of any predefined pattern from the codebook (64 moving pixels), intuitively the mismatch area obtained using the new similarity metric will typically be half that of the existing metric. The following heuristic is therefore justified in order to classify approximately similar number of RMBs from a set of CRMBs:

$$T_{S_2} = \frac{68 T_{S_1}}{132} \qquad (6)$$

Table II shows the empirical results for seven standard video sequences, using T_{S_1} and T_{S_2} as the existing and new similarity metrics respectively. In all examples the new metric captured more RMBs, while Table II also reveals that the classification of a CRMB differed between the two metrics.
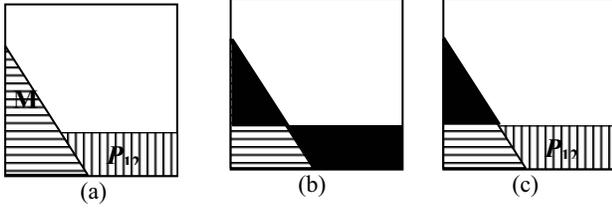
Figure 2: (a) Similarity example for of a moving region M of a CRMB and pattern P_{12}; (b) Two non-overlapping areas (black) relevant to the existing similarity metric; (c) The non-overlapping area (black) relevant to the new similarity metric.

Experiments confirmed that up to 3.6% of MBs classified as RMBs by the existing, but by not the new metric, had relatively large moving regions, approximately half of the MB. These should have actually been classified as AMBs and the new similarity metric does this. The experiments also revealed that up to 10.4% of MBs classified as RMBs by the new, but not the existing metric, had relatively small moving regions, yet were too large to be classified as SMBs and so were treated as an RMB for superior quality. The corollary of this finding is that the new similarity metric provides better control in choosing the similarity threshold in regard to whether a MB is classified as a RMB or AMB.

Table II: Percentage of RMBs generated by the ASPS algorithm [8] with respect to the total MBs using existing (S_1) and new similarity metric (S_2) and various S_1, S_2 permutations.

Video sequences	By S_1	By S_2	By S_1 not by S_2	By S_2 not by S_1
Miss America	18%	22%	1.0%	5.3%
Suzie	21%	26%	1.8%	6.9%
Mother&Daughter	24%	33%	1.0%	10.4%
Carphone	24%	27%	3.2%	5.8%
Foreman	24%	25%	3.6%	4.8%
Salesman	27%	34%	0.6%	7.4%
Claire	14%	16%	0.3%	2.1%

The following Lemma ensures that in all cases, where both metrics classify a CRMB as an RMB, the same pattern is chosen to represent the MR of the CRMB, thereby ensuring the coding efficiencies using both these metrics will be comparable.

Lemma 3: $\forall_u \forall_{v \neq u} [S_1(M, P_u) \Theta S_1(M, P_v)$

$\Leftrightarrow S_2(M, P_u) \Theta S_2(M, P_v)]$ where $\Theta \in \{=, \neq, <, >, \leq, \geq\}$.

Proof: Let P_u and P_v be two arbitrarily selected patterns in PC such that $u \neq v$. $S_1(M, P_u) \Theta S_1(M, P_v)$

$\Leftrightarrow (|M|_1 + |P_u|_1 - 2|M \wedge P_u|_1) \Theta (|M|_1 + |P_v|_1 - 2|M \wedge P_v|_1)$; By Lemma 1.

$\Leftrightarrow -|M \wedge P_u|_1 \Theta -|M \wedge P_v|_1$; $\because |P_u|_1 = |P_v|_1 = 64$.

$\Leftrightarrow (|M|_1 - |M \wedge P_u|_1) \Theta (|M|_1 - |M \wedge P_v|_1)$;

$\Leftrightarrow S_2(M, P_u) \Theta S_2(M, P_v)$; By Lemma 2. ∎

Theorem 1: *The existing and new similarity metrics S_1 and S_2 are equivalent to identifying the best pattern for any moving region.* ∎

4. NEW COMPUTATION STRATEGY

The similarity metric calculation in (2) requires 256 subtractions, 256 absolute and 255 addition operations. From Lemma 1 and 2, $|M|_1 + |P_n|_1 - 2|M \wedge P_n|_1$ and $|M|_1 - |M \wedge P_n|_1$ are the equivalent of the existing and new similarity metrics, so the flow diagrams in Figure 3 can be constructed. For a particular MR, the similarity computation (those operations highlighted in the shaded region) in Figure 3(a) and 3(b), is performed for each pattern in the PC, while those in the non-shaded region are performed just once.
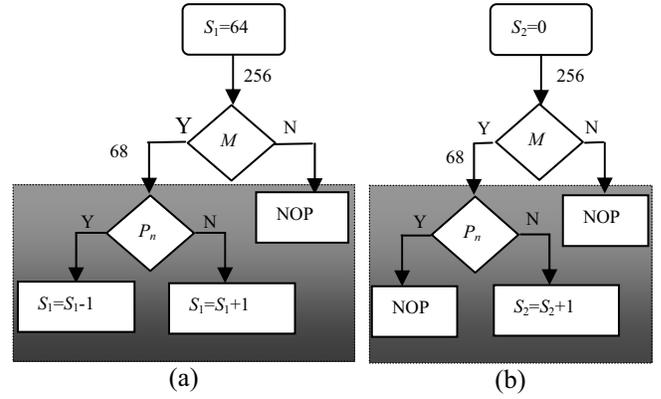

Figure 3: Flowchart of new computation strategy on (a) existing, (b) new similarity metric, where NOP means No Operation.

For the existing similarity metric in Figure 3(a), the parameter S_1 is initialised to the pattern size, namely 64. In order to find the mismatch of the moving region M, 256 compare operations are required for all CRMBs. From Section 3, since on average $M = 68$, for all CRMBs, then 68 compare operations are required. As Figure 3(a) shows, during each of these comparisons, the corresponding pattern position is checked and if it is 1, then S_1 is decremented, otherwise it is incremented. Irrespective of overlapping or non-overlapping between MR and a pattern, the number of operations required for a particular CRMBs is therefore $256 + \lambda(68 + 68)$, where λ is the pattern codebook size. In contrast, the total number of operations when this computation strategy is not applied is $(3 \times 256 - 1)\lambda$. When considering pattern matching algorithms having a maximum value of $\lambda = 32$, the new computation strategy reduces the total number of operations by approximately 81%.

Conversely, the new similarity metric in Figure 3(b), initialises $S_2 = 0$, and does not need not to perform any operations when there is '1' in the corresponding position of both MR and pattern i.e. overlapping regions. When there is a total overlap between the MR and pattern, only $(68 - 64) = 4$ operations are required. When there is no overlapping, i.e., the corresponding position is '0' both the existing and new similarity metrics require the same number of operations namely 68 (i.e., the maximum size of the MR). Thus, on average, the new similarity metric requires 32 fewer operations compared with the

existing metric i.e., $256+\lambda(68+36)$ operations, which is 22% fewer.

5. COMPUTATIOANAL IMPACT UPON PATTERN BASED CODING

To analyse the impact of this new metric on pattern-based coding, assume a MB size of $m \times m$ and maximum motion vector length d. While there is a pattern-based coding overhead, covering the selection of the best pattern for an RMB using the similarity metric, pattern identification coding and residual error arrangement, the major saving is in ME, where only a quarter of a MB needs to be searched. Table III shows that compared to H.263, an improvement of between 19% and 52% is achieved in encoding time per frame using the existing similarity metric and between 21% and 58% using the new similarity metric and generic computation strategy.

Table III: Percentage saving in coding time per frame compared to H.263 using the existing similarity metric without the generic computation strategy and the new similarity metric with the generic computation strategy.

Video sequences	Existing similarity metric	New similarity metric
Miss America	40%	45%
Suzie	24%	27%
Mother&Daughter	39%	43%
Carphone	23%	25%
Foreman	19%	21%
Salesman	52%	58%
Claire	46%	51%

6. CONCLUSIONS

This paper has presented a new similarity metric to efficiently compute the best pattern representation of a moving region in very low bit-rate, blocked-based, video coding. Unlike the existing similarity measure which considers the mismatch areas of both the moving region and pattern in selecting the best-pattern from the codebook, the new metric only considers the mismatch area of the moving region. A generic computation strategy for this similarity metric has also been presented. It has been proven that the same pattern is selected for a particular MR of macroblock using both metrics; however, the computational efficiency of the new approach provides an improvement of up to 58% compared with the H.263 coding standard.

7. REFERENCES

[1] Fukuhara, T., K. Asai, and T. Murakami, "Very low bit rate video coding with block partitioning and adaptive selection of two time-differential frame memories," *IEEE Trans. Circuits Syst. Video Tech.*, 7, 212–220, 1997.

[2] Gonzalez, R.C. and R. E. Woods, *Digital Image Processing*, Addison-Wesley, 1992.

[3] ISO/IEC N4030, MPEG-4 International Standard, 2001.

[4] ITU-T Recommendation H.263, "Video coding for low bit-rate communication," Version 2, 1998.

[5] Maragos, P., "Tutorial on advances in morphological image processing and analysis," *Opt. Eng.*, 26(7), 623–632, 1987.

[6] Paul, M., M. Murshed, and L. Dooley, "A new real-time pattern selection algorithm for very low bit-rate video coding focusing on moving regions," Proc. of IEEE Int. Con. of Acoustics, Speech, and Signal Proc. (ICASSP-03), Hong Kong, Vol-3, III_397-III_400, 2003.

[7] Paul, M., M. Murshed, and L. Dooley, "A Real Time Generic Variable Pattern Selection Algorithm for VLBR Video Coding," IEEE Int. Con. on Image Proc. (ICIP-03), Spain, 2003.

[8] Paul, M., M. Murshed, and L. Dooley, "An Arbitrary Shaped Pattern Selection Algorithm for Very Low Bit-Rate Video Coding Focusing on Moving Regions," Proc. of 4th IEEE Pacific-Rim Int. Con. on Multimedia (PCM-03), Singapore, 2003.

[9] Paul, M., M. Murshed, and L. Dooley, "Impact of Macroblock Classification on LBR Video Coding Focusing on Moving Region," Proc. of Int. Conf. of Com. and IT (ICCIT-02), Dhaka, Bangladesh, 465–470, 2002.

[10] Shanableh, T. and M. Ghanbari, "Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats," *IEEE trans. on multimedia*, 2(2), 101–110, 2000.

[11] Shi, Y.Q. and H. Sun, *Image and Video Compression for Multimedia Engineering Fundamentals, Algorithms, and Standards*, CRC Press, 1999.

[12] Wong, K.-W., K.-M. Lam, and W.-C. Siu, "An Efficient Low Bit-Rate Video-Coding Algorithm Focusing on Moving Regions," *IEEE trans. circuits and systems for video technology*, 11(10), 1128–1134, 2001.