Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 4593 – 4601

doi:10.3906/elk-1704-198

Two-bit transform using local binary pattern method for low-complexity block motion estimation

Aysun TAŞYAPI ÇELEBİ*
Kocaeli University Laboratory of Integrated Systems (KUTSAL), Department of Electronics and Telecommunications Engineering, Kocaeli University, Umuttepe Campus, Kocaeli, Turkey

Received: 17.04.2017 • **Accepted/Published Online:** 20.09.2017 • **Final Version:** 03.12.2017

Abstract: Low bit-depth representation-based motion estimation approaches have been drawing considerable attention recently, mainly because of their small hardware footprint. In this paper, a new two-bit transform using a local binary pattern (LBP) for low-complexity motion estimation is proposed. The proposed approach utilizes the LBP method to obtain two-bit representations of video frames in the binarization process. Video frames are transformed into their low bit-depth representations by the LBP and then a motion estimation process is carried out using these binary frames. A Boolean exclusive-OR operation is used to calculate the number of nonmatching points metric instead of the conventional sum of absolute differences metric in the motion estimation stage. The proposed method reduces the computational complexity, especially in the binarization stage, while improving the motion estimation accuracy compared to existing one-bit and two-bit transform-based low-complexity motion estimation approaches in the literature.

Key words: Local binary pattern, low-complexity binarization, motion estimation

1. Introduction

Video compression methods are widely used in various multimedia applications and consumer electronic devices such as digital televisions, portable video units, digital camcorders, and mobile phones. Motion estimation (ME) is the main part of a typical video coding technique and it has nearly 50% of the total computational load of an encoder. Because of the limited processing power and battery capacity of mobile devices, the usage of ME methods with low complexity has been very important in video coding methods. In the literature, there are several different approaches to reduce the computational load of the full search (FS)-based ME method, which has significant computational load [1]. The FS block matching algorithm divides each current frame into fixed blocks. For each block (reference block) in the current frame, the reference frame is searched within a search window in order to find the best matching. In order to decide the similarity, it uses the sum of absolute differences (SAD) as the matching criterion. This approach is called FS-based ME. The FS-based ME approach is not suitable for mobile devices because of the enormous computational complexity and huge hardware costs.

There are several different approaches to reduce computation complexity and power consumption of FS-based ME methods in the literature. The first group of approaches proposes to check only a subgroup of all available candidate locations, for example three-step search [2], diamond search [3], and hexagonal search [4]. These approaches only check the predefined search locations instead of all possible candidates. In addition, adaptive search range [5] determination-based approaches can be put into this group because of the limited

*Correspondence: aysun.tasyapi@kocaeli.edu.tr

number of candidates. The second group of approaches proposes to utilize fewer pixels in the matching criterion computations by making use of a specific subsampling pattern such as quarter [6], quincunx [7], or 8-Queen [8].

The other group of these approaches represents image frames in a lower bit-depth representation in place of using 8-bit pixel intensity values and then uses Boolean exclusive-OR (EX-OR) operations in the computation of the matching criterion. It is known that the EX-OR-based operations can be effectively carried out in the case of hardware implementations. Thus, the whole computational load of the ME process can be reduced. Various low bit-depth approaches based on ME methods have been proposed in the literature. In [9], one-bit transform (1BT) was proposed where a multiband-pass filter is employed for binarization and EX-OR-based operations for matching criterion computations. In [10], an improved binarization approach called the multiplication-free 1BT method (MF-1BT) was introduced to reduce the computational complexity of the 1BT approach [9] by modifying the number of 1s in the filter kernel. This kernel eliminates real division operations. This kernel includes 16 nonzero components and thus the normalization operation is carried out by making use of integer shift operations. Two-bit transform (2BT)-based ME was presented in [11] to improve the ME accuracy of the one-bit depth-based methods by introducing an additional bit-plane for representing input images. This approach uses local means and standard deviations to obtain two bit planes for each image. This method is called 2BT-based ME. In [12], a constrained one-bit transform (C-1BT) method was presented where a constraint mask is used to mark the reliable pixels for the matching stage. In this method, the first bit-plane is computed as in [10], utilizing a multiplication-free filter. The second bit-plane is constructed as a constraint mask to decide the pixels that are reliable enough to include the matching criterion. C1BT provides better ME accuracy and lower computational complexity compared to 2BT-based ME approaches. In [13], enhanced 2BT-based ME was proposed, where the matching criterion is modified. In [14], enhanced C-1BT, which improves the matching criterion of the standard number of nonmatching points (NNMP), was presented. While this method improves ME performance, it increases the computational complexity as well. In [15], a weighted constraint mask was proposed instead of a binary constraint mask. This method uses a three-bit plane for ME.

In this paper, a new ME approach using LBP-based 2BT, which is suitable for hardware, is presented. The main objective is to develop a ME method that reduces the computational complexity significantly while increasing ME accuracy. This approach has more accuracy and lower complexity than other low-bit depth ME methods.

2. LBP-based binarization

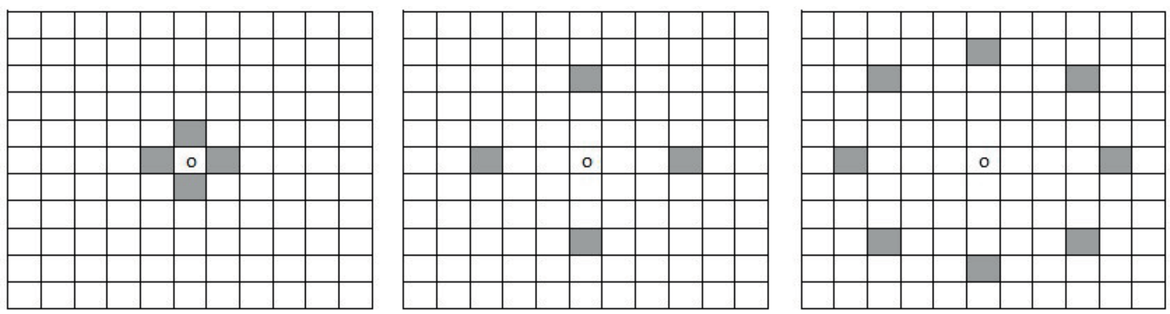
LBP was originally proposed in [16] and has started to find many applications in texture classification [17], face detection and recognition [18], video retrieval, visual inspection, remote sensing, etc. The basic advantage of LBP originates from extracting useful features for classification.

LBP, which extracts local structures of images by generating a binary code for a pixel via comparing the pixel value with the neighboring pixels, is an efficient operator. The most important properties of LBP are computational simplicity, robustness regarding monotonic illumination changes, and discriminative power. The original LBP produces a binary code for the center pixel (xc) as follows:

$$
\begin{aligned}
LPB_{P,R}(c) &= \sum_{p=0}^{P-1} sgn(x_p - x_c)2^p, \\
sgn(z) &= \begin{cases} 1, & z >= 0 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
\tag{1}
$$

where P and R are the number of neighboring pixels and the radius of the LBP, $c$is the center pixel located at point (i,j), x_c is the intensity value of the center pixel, and x_p is the intensity value of the pth neighbor. $LPB_{P,R}(c)$ is the LBP code of a pixel at point (i,j).

Features at different resolutions and scales can be extracted by changing P and R in the LBP operator. Different sample configurations of LBP are given in Figure 1. The pixel marked 'o' in Figure 1 is compared against P equally spaced pixels (highlighted in Figure 1) of radius R to have LBP$_{(P,R)}$ value. The selection of a suitable radius (R) for LBP has a significant impact on the binarization performance. If R is set to a small value, the neighbor pixels are selected too close to the center pixel. Thus, it does not capture the dominant features. If R is set to a large value, the neighbor pixels are selected too far from the center pixel and local characteristics of the image cannot be extracted.



(a) (b) (c)

Figure 1. Different local binary pattern configurations: a) LBP$_{(4,1)}$, b) LBP$_{(4,4)}$, c) LBP$_{(8,4)}$.

Recently in [19,20], LBP was proposed to be utilized for motion estimation. In [19], the proposed approach initially performs 4-Queen subsampling to decide representative pixels. Then it computes the LBP for each pixel and finally selects the pixels that belong to the edge. The proposed approach captures salient microfeatures such as edges in a local region and generates plenty of noisy pixels in binary output images. In [20], a LBP-based approach is used to convert 8-bit depth images into binary images for video stabilization. After binary images are obtained, a matching criterion is utilized to compute global motion.

In this work, a novel two-bit binarization approach by making use of a LBP-based method is presented. The proposed LBP-based binarization approach has lower complexity compared to [9,10]. This approach is called LBP-2BT-based ME. The motivation of the proposed approach is to extract distinctive pixels based on comparison results where the center pixel is compared to the surrounding pixels. In this paper, the original LBP computation is modified to obtain binary images as follows:

$$LPB_{P,R}(c) = \sum_{p=0}^{P-1} sgn(x_c - x_p - T), \tag{2}$$

where T is a fixed-threshold value. The original LBP is typically very sensitive to noise. Thus, in this work a threshold (T) is introduced when comparing the center pixel against the neighboring pixel values. If the difference between a center pixel and its neighbor is greater than T, then the cumulative sum is incremented by one. Otherwise, the cumulative sum is not changed. Next, the proposed approach generates two-bit images by making use of $LPB_{P,R}(c)$ values as

$$B_1(i,j) = \begin{cases} 1, & LPB_{P,R}(c) >= P/2 \\ 0, & otherwise \end{cases},$$

$$B_2(i,j) = \begin{cases} 0, & LPB_{P,R}(c) = \{0 \text{or} P\} \\ 1, & \text{otherwise} \end{cases}, \tag{3}$$

where $B_1(i,j)$ and $B_2(i,j)$ represent first and second bit-plane images. It becomes possible to select the pixels representing different characteristics among the surrounding pixels with this approach.

Figure 2 shows the two-bit planes constructed by the proposed method for the 45th frame of the Foreman test sequence together with the original image frame. In general, the first bit-plane extracts salient edges whereas the second bitplane provides coarse structures in binary form. The aim of using these two images together is to combine discriminated edges and coarse structures for improving ME accuracy while preserving the robustness for illumination changes and noises.

(a) (b) (c)

Figure 2. A LBP-2BT transform for a sample frame of the Foreman sequence: a) original frame, b) 1st bit plane of the LBP-2BT of a sample frame, c) 2nd bit plane of the LBP-2BT of a sample frame.

After the LBP-based binarization step presented in this work, the motion vector of a block is decided based on the number of modified nonmatching points (MNNMP) criterion as:

$$MNNMP(c_x, c_y) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \left\{B_1^t(i,j) \oplus B_1^{t-1}(i+c_x, j+c_y)\right\} + \left\{B_2^t(i,j) \oplus B_2^{t-1}(i+c_x, j+c_y)\right\}, \tag{4}$$

where $-s \le c_x,\ c_y \le s-1$, (c_x, c_y) shows the candidate motion vector, N and s respectively determine the block size and search range, $B_1^t(i,j)$ and $B_2^t(i,j)$ are first and second bit planes of the current frame, $B_1^{t-1}(i,j)$ and $B_2^{t-1}(i,j)$ are first and second bit planes of the reference frame, and $\oplus$ denotes the EX-OR operation. While the NNMP metric used in 2BT employs three Boolean operations (two EXOR and one OR), the proposed MNNMP is computed by performing the EXOR operation on bit planes separately and then adding the results.

3. Experimental results

The performances of the proposed ME method and other ME methods are measured in terms of peak signal-to-noise ratio (PSNR) between the original and estimated image frames. This approach is implemented as well with an open-loop scheme.

In order to determine the performance of the proposed LBP2BT-based method, four different LBP configurations are investigated, where P is set to 8 neighbor pixels with different radiuses (R = 4, 8, 12, and 16). In these cases, threshold value T is selected as 16 based on the experiments. When threshold value T is set as zero, similarly to the original LBP-based ME method in [20], a serious noise effect is introduced to the binary images. Thus, ME performance is degraded significantly. Table 1 gives PSNR results for these configurations. As seen from Table 1, the best ME performance is achieved when R is set to 12. This is mainly because a radius of 12 is relatively good at extracting most discriminative details from the reasonable distance between the center pixel and its neighbors. If R is set to 4, the neighbor pixels are too close, and if R is set to 16, the neighbor pixels are too far for the center pixel. Thus, local characteristics of the image cannot be extracted efficiently.

Table 1. PSNR performance (in dB) of the proposed method for different configurations (motion block size is 16 $\times$ 16 and search range is ± 16).

Video sequence	LBP-2BT$_{(8,4)}$	LBP-2BT$_{(8,8)}$	LBP-2BT$_{(8,12)}$	LBP-2BT$_{(8,16)}$
Football	21.81	22.08	22.18	22.12
Foreman	30.53	30.82	30.80	30.62
Garden	23.37	23.49	23.52	23.49
Coastguard	29.97	30.08	30.11	30.06
Tennis	28.50	28.68	28.73	28.55
Mobile	23.67	23.71	23.72	23.67

In [19], it was proposed to utilize LBP for motion estimation. This approach initially performs 4-Queen subsampling to decide representative pixels. Then it computes the LBP for each pixel and finally selects the pixels that belong to the edge. This binarization scheme generates plenty of noisy pixels in binary output images. In Figures 3a–3d, the binarization performance of the proposed approach is shown with comparison to LBP-based binarization in [19] and [20]. It is clear from Figure 3b that the method in [19] produces significantly noisy binary images, which may degrade motion estimation performance. Therefore, the ME accuracy of this method is not calculated. Figure 3c shows that the binary image obtained by the method in [20] has more microstructural information than binary images (second bit plane) obtained by the proposed method. The proposed approach provides better binarization performance than LBP-based approaches according to the visual results.

Table 2 shows the average PSNR comparison results in dB for different video sequences displaying various motion characteristics. These sequences have both camera motion and local motion. As seen from these results, the proposed approach provides better ME accuracy compared to the 2BT-based ME approach. When it is compared to the C-1BT-based ME approach, it has a slight performance loss for only two video sequences. However, the proposed method gives the best overall performance according to the average PSNR for all sequences. It is important to note that the original LBP-based binary ME method presented in [20] has significantly lower accuracy since it extracts much microstructural information and thus introduces noise into binary frames.

In order to assess the computational complexity of the proposed method, the number of operations needed per pixel (pp) is shown in Table 3. As seen from this table, the proposed method requires fewer operations, especially at the binarization stage, compared to other methods in the same category. The proposed method requires 2.35 additions, 1 subtraction, and 10 comparisons per pixel on average for the binarization of test sequences whereas 2BT-based ME needs 2.8125 additions, 1.0625 multiplications, and 0.0312 subtractions on

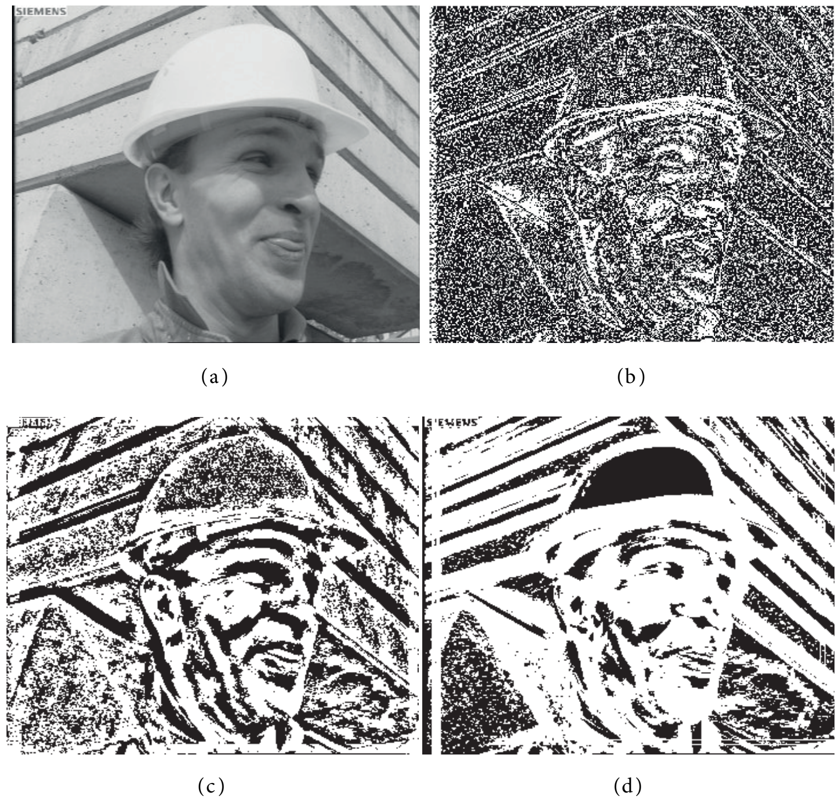


Figure 3. Binarization results of different methods: a) original Foreman sequence frame #8, b) binary image obtained by the method in [19], c) binary image obtained by the method in [20], d) binary image obtained by the proposed method (2nd bit plane).

Table 2. PSNR performance (in dB) of low-complexity ME methods.

ME method	Video sequences (frame size, sequence length)						
	Football (352×240) (125 frames)	Foreman (352×288) (300 frames)	Garden (352×240) (115 frames)	Coastguard (352×288) (300 frames)	Tennis (352×240) (150 frames)	Mobile (352×240) (300 frames)	Average of six video sequence
SAD (8-bit depth)	22.88	32.09	23.79	30.48	29.45	23.94	27.10
1BT [9]	21.83	30.32	23.31	29.83	28.11	23.61	26.17
MF-1BT [10]	21.81	30.38	23.26	29.88	28.18	23.63	26.19
2BT [11]	22.06	30.70	23.43	29.94	28.46	23.66	26.38
C-1BT [12]	22.10	30.86	23.38	29.98	28.71	23.69	26.45
LBP [20]	21.40	29.20	23.07	29.72	27.43	23.43	25.59
Proposed method (LBP-2BT)	22.18	30.80	23.52	30.11	28.73	23.72	26.51

average for binarization. The 2BT-based ME approach requires floating point multiplication and thus it is not suitable for efficient hardware implementations. The proposed method has lower computational complexity in the binarization stage compared to C-1BT, as well. As a result, the computational complexity of the proposed method is superior compared to other methods in the same category. Note that the approach in [14] is not taken into consideration for comparison since it requires a significantly higher computational load. Tables 2 and

3 show that the proposed method reduces the computational load of the binarization stage according to other methods while improving the motion estimation accuracy.

Table 3. Number of operations needed for binarization and matching (per pixel).

Stage		1BT [9]	MF1BT [10]	2BT [11]	C1BT [12]	LBP [20]	LBP-2BT (proposed)
Binarization	Addition	25	16	2.8125	16	8	2.35
	Multiplication	1	-	1.0625	-	-	-
	Shift.	-	1	-	1	-	-
	Subtraction	-	-	0.0312	1	-	1
	Comparison	1	1	3	2	9	10
	Bool	-	-	1	-	-	-
Binarization total		27	18		20	17	13.35
Matching	Bool	1	1	3	3	1	2
	Addition	-	-	-	-	-	1

Threshold value T is selected as 16 based on the experiments for PSNR results in Tables 1 and 2. In order to evaluate the effect of the proposed method on the threshold value, results are obtained for different threshold values of all sequences. Table 4 shows the PSNR values of the proposed method for different threshold values. It is seen from Table 4 that PSNR values increase for large threshold values. $T = 16$ is found to provide better results for all sequences. If T is set to 0, all microstructural information is extracted; this microstructural information behaves as noise during the matching stage and thus ME accuracy is reduced. If T is set to 20, obtained local characteristics of the image can be lost. Thus, ME accuracy can decrease for some sequences.

Table 4. PSNR performance (in dB) of the proposed method for different threshold values.

Video sequence	T = 0	T = 4	T = 8	T = 12	T = 16	T = 20
Football	21.92	21.97	22.05	22.13	22.18	22.20
Foreman	30.00	30.61	30.77	30.85	30.80	30.66
Garden	23.44	23.48	23.50	23.51	23.52	23.54
Coastguard	30.12	30.14	30.14	30.15	30.11	30.03
Tennis	28.29	28.29	28.49	28.66	28.73	28.71
Mobile	23.69	23.70	23.70	23.71	23.71	23.71
Average of six video sequences	26.24	26.36	26.44	26.50	26.51	26.47

4. Conclusion

In this work, novel 2BT-based ME using LBP is proposed. The proposed approach obtains two-bit representations of each video frame using LBP in the binarization stage. While the first bit plane shows salient edges, the second bit plane extracts coarse structures. After the binarization stage, ME is carried out using these binary frames by calculating the MNNMP criterion.

The experimental results show that the proposed 2BT-LBP-based approach provides improved ME performance compared to C-1BT- and 2BT-based two-bit approaches. Additionally, the proposed approach has lower computation complexity than other low-bit depth ME approaches. It is important to note that the proposed method has the lowest binarization cost among the compared methods. Thus, it is suitable for hardware implementations.

Acknowledgment

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant No. EEEAG/115E921.

References

[1] Chen TC, Chen YH, Tsai SF, Chien SY, Chen LG. Fast algorithm and architecture design of low-power integer motion estimation. IEEE T Circ Syst Vid 2007; 17: 568-577.

[2] Koga T, Linuma K, Hirano A, Lijima Y, Ishiguro T. Motion compensated interframe coding for video conferencing. In: Proceedings of the Natational Telecommunications Conference; 1981. pp. C9.6.1-C9.6.5.

[3] Zhu S, Ma KK. A new diamond search algorithm for fast block-matching motion estimation. IEEE T Circ Syst Vid 2000; 9: 287-290.

[4] Zhu C, Lin X, Chau LP. Hexagon-based search pattern for fast block motion estimation. IEEE T Circ Syst Vid 2002; 12: 349-355.

[5] Lee J, Choi M, Cho Y, Kim J, Cho WK. Fast H.264/AVC motion estimation algorithm using adaptive search range. In: 12th International Symposium on Integrated Circuits; 14–16 December 2009; Singapore. pp. 336-339.

[6] Bierling M. Displacement estimation by hierarchical block matching. In: SPIE Conference on Visual Communications and Image Processing; 25 October 1998; San Jose, CA, USA. pp. 942-951.

[7] Lengwehasatit K, Ortega A. Probabilistic partial-distance fast matching algorithms for motion estimation. IEEE T Circ Syst Vid 2001; 11: 139-152.

[8] Wang CN, Yang SW, Liu CM, Chiang T. A hierarchical n-queen decimation lattice and hardware architecture for motion estimation. I IEEE T Circ Syst Vid 2004; 14: 429-440.

[9] Natarajan B, Bhaskaran V, Konstantinides K. Low-complexity block-based motion estimation via one-bit transforms. IEEE T Circ Syst Vid 1997; 7: 702-706.

[10] Ertürk S. Multiplication-free one-bit transform for low-complexity block-based motion estimation. IEEE Signal Proc Let 2007; 14: 109-112.

[11] Ertürk A, Ertürk S. Two-bit transform for binary block motion estimation. IEEE T Circ Syst Vid 2005;15: 938-946.

[12] Urhan O, Ertürk S. Constrained one-bit transform for low-complexity block motion estimation. IEEE T Circ Syst Vid 2007; 17: 478-482.

[13] Choi C, Jeong J. Enhanced two-bit transform based motion estimation via extension of matching criterion. IEEE T Consum Electr 2010; 56: 1883-1889.

[14] Lee S, Jeon G, Jeong G. Fast motion estimation based on enhanced constrained one-bit transform . Electron Lett 2014; 50: 746-748.

[15] Güllü MK. Weighted constrained one-bit transform based fast block motion estimation. IEEE T Consum Electr 2011; 57: 751-755.

[16] Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classi?cation based on feature distributions. Pattern Recogn 1996; 29: 51-59.

[17] Liu L, Zhao L, Kuang G, Fieguth P. Extended local binary patterns for texture classification. Image Vision Comput 2012; 39: 86-99.

[18] Zhao XM, Zhang SQ. Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding. EURASIP J Adv Sig Pr 2012; 2012: 20.

[19] Verma R, Dabbagh MY. Binary pattern based edge detection for motion estimation in H.264/AVC. In: 26th IEEE Canadian Conference of Electrical and Computer Engineering; 2013. New York, NY, USA: IEEE. pp. 1-4.

[20] Kır B, Kurt M, Urhan O. Local binary pattern based fast digital image stabilization. IEEE Signal Proc Let 2015; 22: 341-345.