2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance

A Spatiotemporal Motion-Vector Filter for Object Tracking on Compressed Video

Ronaldo C. Moura
Monity
Av Andrômeda, 443, Torre Oeste, Sala 414,12228-900,São José dos Campos - SP, Brazil
http://www.monity.com.br

Elder Moreira Hemerly
Instituto Tecnológico de Aeronáutica
Praça Mal. Eduardo Gomes, 50, Vila das Acácias,CTA, 12228-900, São José dos Campos-SP, Brazil.
www.ele.ita.br/~hemerly/

Abstract

In this paper, a novel filter for real-time object tracking from compressed domain is presented and evaluated. The filter significantly reduces the noisy motion vectors, that do not represent a real object movement, from Mpeg family compressed videos. The filter analyses the spatial (neighborhood) and temporal coherence of block motion vectors to determine if they are likely to represent true motion from the recorded scene. Qualitative and quantitative experiments are performed displaying that the proposed spatiotemporal filter (STF) outperforms the currently widely used vector median filter. The results obtained with the spatiotemporal filter make it suitable as a first step of any system that aims to detect and track objects from compressed video using its motion vectors.

1. Introduction

Inside the broad Computer Vision research field are located the Object Tracking techniques, which consist in the ability to automatically track an object at consecutive video frames. During the last two decades several techniques have been proposed for video object tracking with applications on Video Surveillance [8], Intelligent Transportation System - ITS [2], Human Machine Interface - HMI, Video Indexing [13], [12] and Shopping Behavior Analysis.

The adoption of surveillance cameras everywhere, and interest in automatic video indexing served as stimulus for recent research on object tracking and behavior recognition, such as in AVSS (IEEE International Conference on Advanced Video and Signal based Surveillance from 1998,2001,2003,2005-2010), PETS (IEEE International Workshop on Performance Evaluation of Tracking and Surveillance from 2000-2010), CLEAR (Classification of Events, Activities and Relationship Evaluation and Workshop from 2006,2007), CBMI (IEEE International Workshop on Content-Based Multimedia Indexing from 1999,2001,2003,2005,2007-2010) and ICDCS (ACM/IEEE International Conference on Distributed Smart Cameras from 2007-2010)

Despite the increasing microprocessors computational power in recent years, the processing required by object tracking techniques still consists in a bottleneck to their wider adoption, specially on low cost embedded equipment as surveillance cameras and mobile devices. To reduce this computational power demand, some techniques that extract object motion information from compressed video streams, instead of the raw video, have been developed.

These techniques, by taking advantage of important information inside video compressed by standards like Mpeg family, are capable of tracking an object without the need to fully decompress the video data, reducing by orders of magnitude the required computational complexity. The main compressed domain information used for segmentation and tracking is the block motion vectors and the discrete cosine transform (DCT) coefficients

Nevertheless, the motion-vectors (mv) contained in compressed video are chosen to minimize video bitstream while maintaining its human perceptible quality, not to represent true objects motion. Consequently the mv can represent both a real object movement or two similar block textures in consecutive frames (fake movement). To make the motion-vector useful for further segmentation steps, it is necessary to remove the noisy ones, i.e., the mv that do not represent a real object movement.

978-0-7695-4264-5/10 $26.00 © 2010 IEEE
DOI 10.1109/AVSS.2010.82

This paper presents a novel Spatiotemporal motion vector consistency Filter(STF) to remove noisy motion vectors for object tracking purpose. In Section 2, several related works for spurious motion-vector removal are reviewed, especially the widely adopted vector median filter. In Section 3, the new STF approach is presented. In Section 4 this approach is tested and evaluated. Further research possibilities are presented in the final Section 5-Conclusion.

2. Related Works

Several approaches have been proposed to object segmentation and tracking on compressed videos.

In [2] it is presented a simple car tracking technique based on motion vectors from Mpeg2 video, using fixed cameras. First a *vector median filter* is applied to all motion vectors, then non zero motion vectors are grouped (labeled) according to their direction and magnitude proximity. Each blob is projected on previous frame according to the mean block motion vectors value, and then matched with the nearest blob.

The well referenced Favalli work [6] presents a supervised tracking techniques based in motion vectors. The first step consists in manually selecting the frame macroblocks that must be tracked. Then each selected macroblock are tracked in a frame by frame basis using its motion vectors projection. In [4] a macroblock tracking technique improves [6] by creating two independent layers on the top of macroblock grid, allowing a more fine grained tracking of object boundaries with resolution superior to macroblock size.

For a not fixed camera, i.e., performing zoom, rotation, pan, tilt or translation operations, some techniques have been proposed to determine the global camera motion estimation(GME), to neutralize it before perform object tracking. This is specially useful for segment truly moving objects (foreground), while discarding the motion vectors associated with camera motion only (background).

The Kim and Kim paper [11] presents a detailed eight parameters linear estimation model for a camera performing three-dimensional rotation, zooming, but without translation. The motion vectors with high activity in luminance signal, as edges and high textures are selected as feature point for a least-square estimator.

In [14], Roy Wang *et al*. propose a set of confidence measures for DCT and motion-vectors based object tracking, for moving camera. The motion-vectors are compared with their neighbors, resulting in separated magnitude and direction confidences. A texture confidence measure is taken by analyzing regions with low AC energy in their DCT coefficients. This lower AC energy represents lower textured regions, as roads and sky, where motion vectors are usually less reliable. All confidence measures are then weighted and used in a recursive least square global motion estimator (gme), to determine camera zoom, vertical and horizontal translations. The resulted mv are then processed by a 3-dimension vector median filter, and segmented with a K-means clustering followed by an expectation-maximization (EM) clustering.

In [15] an eight parameters bilinear equation for camera global motion estimation is presented. The parameters are iteratively calculated by a least-square estimator, removing outliers with error greater than average error.

The well referenced Mezaris *et al*. work [13] uses the global motion estimation technique from [15] to automatically segment macroblocks on frame t, as foreground or background, and then applies the macroblock tracker [6] on foreground, resulting an estimated foreground map of frame t+1. This estimated macroblock foreground map of frame t+1, and the foreground map created by application of [15] on next frame t+1, are intersected resulting the filtered foreground estimation. This process is executed during *n* consecutive frames, resulting in good macroblock tracker without the burden of infinite error propagation of [6], as the tracked region of interest is constantly reset. The background with different color tones is also segmented using dc coefficients of DCT transform (Y, CB, CR) of macroblocks presented in I-frames.

In [9] a six parameters GME from [5] to automatically segment moving objects, and then applies a median filter on foreground macroblocks along their motion trajectory in the same group of pictures, usually containing 8 frames, to filter outliers. The filtered foreground macroblocks are grouped in blobs using timed Motion History Image technique, from [3], together with a connected component analysis. Blobs tracking are performed by 20x20 pixel window search on adjacent frame from estimated position of blob (center of gravity plus average motion vector).

In [12] motion vectors and dc color coefficients are used to overcome the Mezaris *et al*. [13] limitations for tracking object motion with small differences compared with camera motion model. The [5] is also used for GME to automatically segment moving object.

In the well referenced work [1], Babu *et al*. implements object segmentation based in motion-vectors from compressed videos. The motion vectors from P and B frames are accumulated over a few frames, median filtered, interpolated and segmented with an expectation maximization (EM) algorithm.

2.1. The limitations of vector median filter

Vector median filter is widely adopted for motion vectors filtering, and is presented in older and newer works as in [1, 2, 9, 14]. Nevertheless, using two-dimensional (spatial) vector median filter presents limitations as:

- **Noise adhered to object boundary problem** - After the median filtering, the noisy motion vectors near a real object were stuck (adhered) to the object boundaries, instead of been removed. The vector median filter is expected to remove noisy vectors, but in the boundaries of moving objects, surrounded by static background (i.e. with $\vec{mv} = (0,0)$), there is a great probability that an eventual noisy-mv could be the median value between object-mv and background-mv.

- **Inability to track small objects** - Small objects, with size of about one block, are mostly incorrectly filtered (removed), as their neighborhood do not have their same moving pattern.

The attempt of reducing these problems, by taking into account also temporal information and creating a three dimensional (3D) vector median filter, produces other limitations, such as:

- **Inability to track fast moving objects** - Fast objects are likely to present significant block movement, while normal median filter expected temporally adjacent blocks represent the same data. This will cause the object to be deformed, with its front incorrectly filtered(removed) and an incorrect tail created.

The concept of the vector median filter could be significantly improved if the spatiotemporal approach considers that object movement must be compensated before the 3D filtering. That is what spatiotemporal motion-vector filter does, and it is described in the following section.

3. Spatiotemporal Motion-Vector Consistency Filter

To improve the median filter limitations, a novel spatiotemporal motion vector consistency filter (STF) is presented. Let $(x, y)^t$ represents the pixel coordinate (x, y) in frame t, and $mv(x, y)^{t \to t_{ref}}$ represents the motion vector of pixel $(x, y)^t$ to a reference frame t_{ref}. The STF filter consists in the following steps:

- **Motion Vector Normalization** - A motion vector from a P frame references a past frame. A motion vector from a B frame references a past or a future frame. To simplify their usage, motion vectors need to be normalized to make them independent of frame type. This is accomplished by dividing the motion vectors by the difference between the current frame number and the reference frame number, according to Motion Vector Normalization (N) Equation 1, similarly to the process used in [1]. If the reference frame is a future frame, the divisor will be a negative number, reversing the mv direction. The normalized motion vector is an approximation of the $mv(x, y)^{t \to (t-1)}$, i.e., the motion vector referencing the previous frame. In this paper the normalized motion vector $N(mv(x, y)^{t \to t_{ref}})$ is represented as $mv(x, y)^t$.

$$N(mv(x,y)^{t \to t_{ref}}) = \frac{mv(x,y)^{t \to t_{ref}}}{t - t_{ref}} \approx mv(x,y)^{t \to (t-1)} \tag{1}$$

- **Temporal Consistency Analysis** - Each block center $(x, y)^t$ has it position in previous frame estimated by adding to it the corresponding normalized motion vector, as described in Projection (P) Equation 2.

  Then, the motion vectors from these two related blocks, $(x, y)^t$ and its projection in previous frame $P(x, y)^{t \to (t-1)}$, have their direction and magnitude coherence simultaneously analyzed by the using the Vector Matching Ratio (R) Equation 3. This process is illustrated in Figure 1.

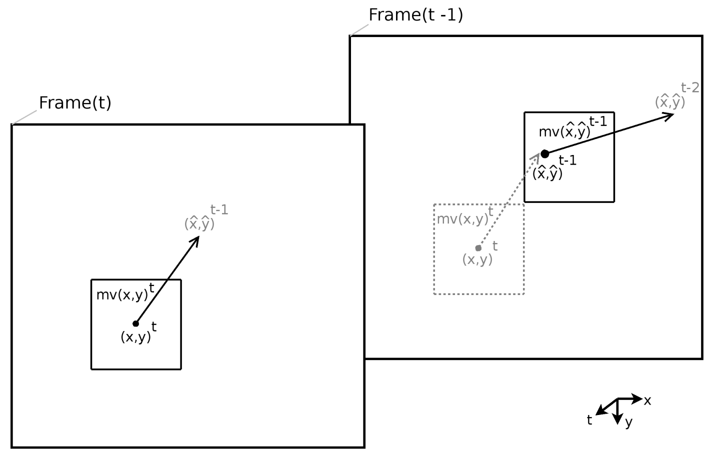


Figure 1. Illustration of the Projection (P) Equation. The Vector Matching Ratio Equation $R\left(mv(x,y)^t, mv(\hat{x},\hat{y})^{t-1}\right)$ is applied to each block center.

$$\begin{aligned} P(x,y)^{t \to (t-0)} &= (x,y)^t \\ P(x,y)^{t \to (t-1)} &= (x,y)^t + mv(x,y)^t = (\hat{x},\hat{y})^{t-1} \\ P(x,y)^{t \to (t-k)} &= P\left(P(x,y)^{t \to (t-k+1)}\right)^{\to (t-k)} \\ &\qquad \text{for } k > 2 \end{aligned} \tag{2}$$

$$R(\vec{a}, \vec{b}) = \begin{cases} 1, & \text{if } \vec{a} = \vec{b} = (0,0) \\ 1 - \frac{\|\vec{a} - \vec{b}\|}{\|\vec{a}\| + \|\vec{b}\|}, & \text{otherwise} \end{cases} \tag{3}$$

$$TCI(mv(x,y)^t) = \prod_{i=1}^{n} R\left(mv(P(x,y)^{t \to t-i+1}),\ mv(P(x,y)^{t \to t-i})\right) \quad (4)$$

$$NCI(mv(x,y)^t) = \frac{\sum_{\Delta x=-1}^{1}\sum_{\Delta y=-1}^{1}\sum_{\Delta t=-1}^{0} R\left(mv(x,y)^t,\ mv(x+\Delta x \cdot m_x, y+\Delta y \cdot m_y)^{t+\Delta t}\right) \cdot W(\Delta x, \Delta y, \Delta t)}{\sum_{\Delta x=-1}^{1}\sum_{\Delta y=-1}^{1}\sum_{\Delta t=-1}^{0} W(\Delta x, \Delta y, \Delta t)} \quad (5)$$

$$STF(mv(x,y)^t) = \begin{cases} mv(x,y)^t, & \text{if } TCI(mv(x,y)^t) >= (\tau_{tci})^n \text{ or } NCI(mv(x,y)^t) >= \tau_{nci} \\ background(x,y)^t, & \text{otherwise} \end{cases} \quad (6)$$

- **Neighbor Consistency Analysis** - Each block $(x,y)^t$ has its motion vector compared with mv of surrounding blocks, in frame t and $t-1$. Motion vectors coherence is analyzed with the Vector Matching Ratio (R) Equation 3.

The Temporal Consistency Analysis can be recursively calculated on $t-1, t-2 \ldots\ t-n$ previous frames, resulting in the Temporal Consistency Index(TCI) Equation 4.

The Neighbor Consistency Index(NCI) Equation 5 represents the proportion of surrounding mv that are consistent with a reference mv, weighted by distance to this reference mv. The variables m_x and m_y represent the block size, typically 16×16. The nearest neighbors have a greater importance in this index, as described in Neighbor Weight(W) Equation 7.

$$W(\Delta_x, \Delta_y, \Delta_t) = \begin{cases} 0 & \text{if } \Delta_x = \Delta_y = \Delta_t = 0 \\ \frac{1}{|\Delta x|+|\Delta y|+|\Delta t|}, & \text{otherwise} \end{cases} \quad (7)$$

A motion vector is considered consistent if its TCI or NCI is above a minimum threshold, as described in Spatiotemporal consistency Filter (STF) Equation 6. Good filtering results were obtained from tested sequences by setting the number of previous frames to $n = 2$, the temporal consistency threshold to $\tau_{tci} = 50\%$, and the neighbor consistency threshold to $\tau_{nci} = 50\%$. The STF filter is relatively robust to changes in thresholds, and produces similar results even after setting $\tau_{tci} = 35\%$ or $\tau_{tci} = 65\%$. This occurs because the calculated consistency index for the noisy motion vectors are usually significantly smaller than the index of the "true" motion vectors.

A motion vector classified as noise should not be considered in the following tracking stages. This can be done by setting the noisy mv to a background value, for instance, (0,0) in the case of static cameras. Another alternative would be estimate the correct value for noisy mv using the previous and surrounding mv. This last approach, nevertheless, can propagate error in highly noisy situations and, in tested sequences, produced worse results than setting mv to a background value.

4. Experimental Results

4.1. Qualitative Results

To allow the comparison between the spatiotemporal consistency filter (STF) and the vector median filter (MF), the motion vectors before and after each filter are displayed in the following images. The motion vector value is drawn over each block with the layout $\boxed{\begin{matrix} d_x \\ d_y \end{matrix}}$. The motion vectors with value (0,0), of blocks belonging to background, are not displayed to simplify the analysis.

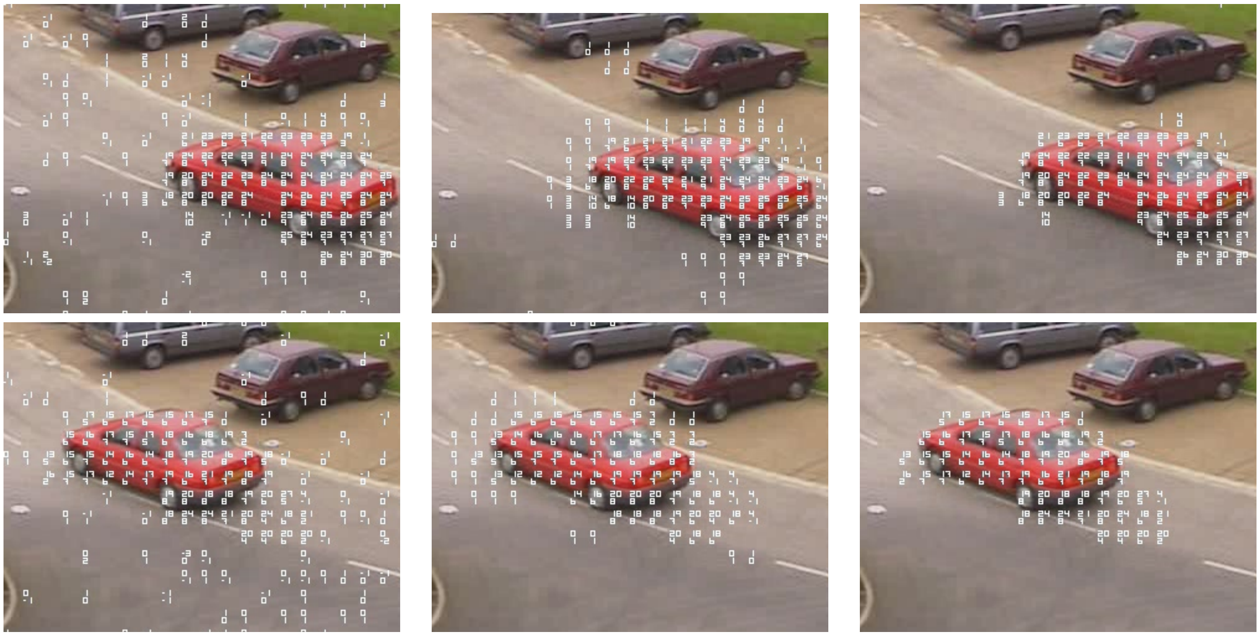
Performing significantly better than the vector median filter, the proposed spatiotemporal consistency filter does not present the "noise adhered to object boundary problem" described in section 2.1, as displayed in Figure 2, and is capable of correctly filter objects as small as one block, as displayed in Figure 3.

4.2. Quantitative Results

To numerically evaluate the ability of the median vector filter and the spatiotemporal consistency filter to correctly detect true objects motion, the CLEAR Multiple Object Detection metrics described in Kasturi *et al*. work [10] were used.

CLEAR MOD metrics notation

- N_{frames}: the number of frames in video sequence.



(a)Original mv *(b)Median Filtered mv* *(c)Spatiotemporal Filtered mv*

Figure 2. Comparison between motion vectors filters. The Spatiotemporal Consistency Filter performs better on car boundaries, keeping its correct shape, and removing the small noisy mv, as $(1, 0)$ and $(0, 1)$. Sequence Pets2000, mpeg4 compressed, with EPZS motion estimator and gop = 10. Frames 140 and 150

(a)Original mv *(b)Median Filtered mv* *(c)Spatiotemporal Filtered mv*

Figure 3. Comparison between motion vectors filters. The Spatiotemporal Consistency Filter is capable of correctly filter a person with one block size in $\frac{3}{4}$ of sampled frames (failing only after the person is occluded behind the tree). The Median Filter completely fails to perceive true motion from small objects . Sequence Pets2001 - camera 1, mpeg4 compressed, with EPZS motion estimator and gop = 50, with 48 frames sampling period: frames 1914, 1962, 2010, and 2058.

- $G_i^{(t)}$: the *i*th ground truth object in frame *t*.
- $D_i^{(t)}$: the *i*th detected (by the evaluated technique) object in frame *t*.
- $N_G^{(t)}$: number of ground truth objects in frame *t*.
- $N_D^{(t)}$: number of detected objects in frame *t*.
- $N_{mapped}^{(t)}$: number of match pairs between ground truth and detected objects in frame *t*.

The **Multiple Object Detection Accuracy - MODA** metric uses the number of missed detections m_t , the falsely identified objects fp_t , to assess the accuracy aspect of the object detection algorithm.

$$MODA = 1 - \frac{\sum_{t=1}^{N_{frames}}(m_t + fp_t)}{\sum_{t=1}^{N_{frames}} N_G^{(t)}} \quad (8)$$

The **Multiple Object Detection Precision - MODP** gives the average overlapping ratio (match ratio) between the bounding-boxes of ground-truth and detected objects, as defined in Equation 9. It does not take in consideration the missed or falsely identified objects.

$$MODP = \frac{\sum_{t=1}^{N_{frames}} \sum_{i=1}^{N_{mapped}} \frac{\left|G_i^{(t)} \cap D_i^{(t)}\right|}{\left|G_i^{(t)} \cup D_i^{(t)}\right|}}{\sum_{t=1}^{N_{frames}} N_{mapped}^{(t)}} \quad (9)$$

The following steps were used to convert the motion-vectors to objects, so they can be analyzed by CLEAR MOD metrics: a given block was considered as foreground if its motion vector has a value different from the background, i.e., (0,0) in the case of static cameras. The foreground motion vectors were grouped using an 8-connected component labeling, forming the objects.

As even after the filters a great number of small noisy motion vectors were still present, a minimum object size filter was used after the three tested configurations (STF,MF and none/no filter). The STF produced the best MOD metrics ignoring objects with one block size, and MF and ”none” configurations produced the best MOD metrics ignoring object smaller than six blocks.

The *usf_date* software from [10] was used to calculated the metrics. The video sequences and ground-truths chosen were the Dataset1/Testing/Camera1 from the PETS2001 workshop and FightOneManDown from CAVIAR dataset [7], displayed in Figures 4 and 5. The ground-truths were converted to the VIPER XML format accepted by the *usf_date* software.

The metrics results are displayed in Table 1 and Table 2. The spatiotemporal consistency filter outperforms the median filter both in bounding box overlap precision (with an average MODP of 44.5% against 27.5%), and in a lower number of false positives resulting a better accuracy (with an average MODA of 44% against 33% and -9%).

MV Filter	Missed	False Detec.	**MODP**	**MODA**
none	4803	578	31%	31%
MF	4899	380	28%	33%
STF	4089	329	42%	44%

Table 1. Multiple Object Detection metrics for sequence PETS2001 Dataset1/Testing/Camera1, with 2500 frames, 7849 objects, gop=128 and without B Frames. Obtained from objects segmented from motion vectors without any filter(none), with vector median filter(MF), and with the proposed spatiotemporal consistency filter(STF).

MV Filter	Missed	False Detec.	**MODP**	**MODA**
none	1504	947	22%	-20%
MF	1477	746	27%	-9%
STF	1058	79	47%	44%

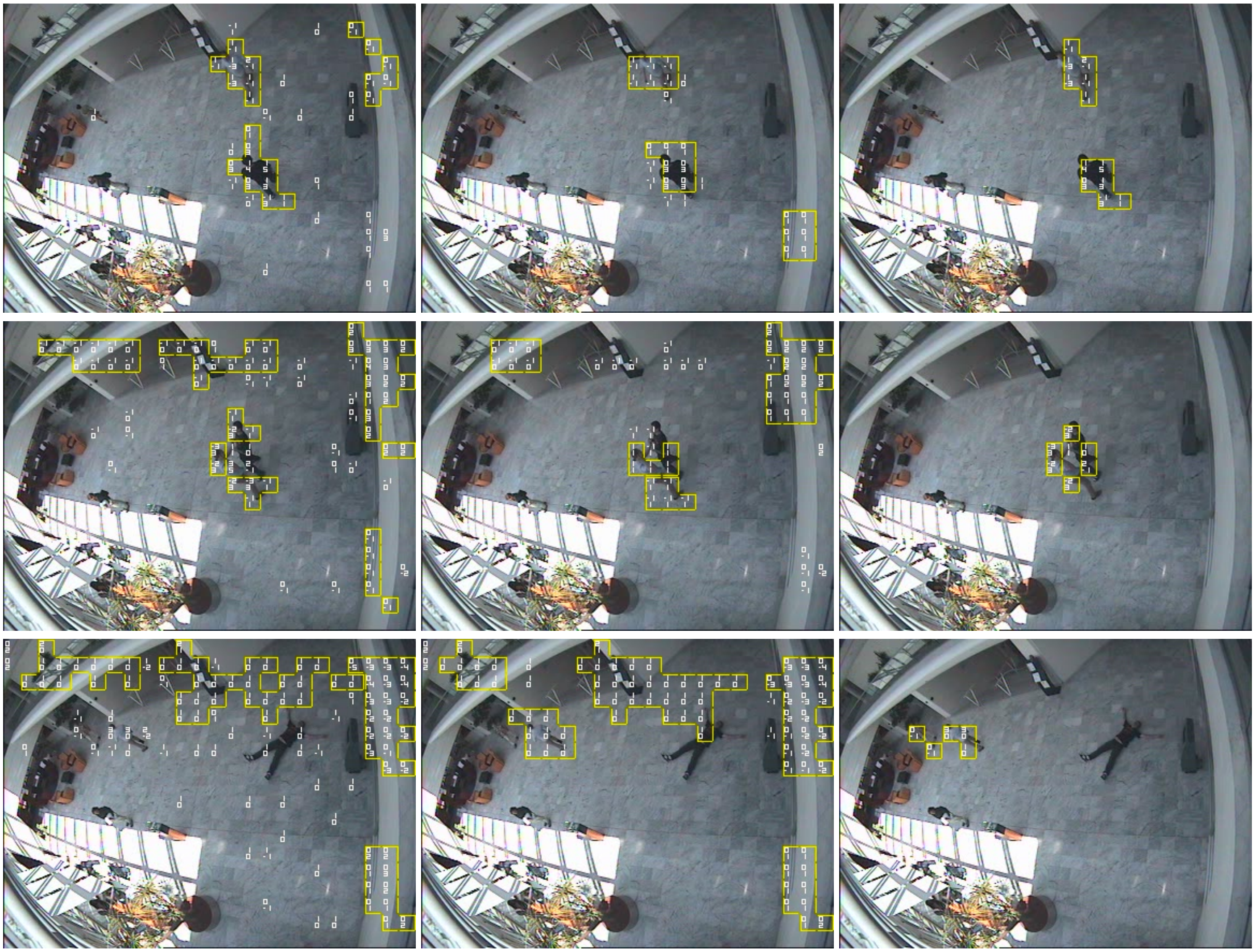
Table 2. Multiple Object Detection metrics for sequence CAVIAR Fight OneManDown, with 803 frames, 2036 objects, gop=12 and without B Frames. Obtained from objects segmented from motion vectors without any filter(none), with vector median filter(MF), and with the proposed spatiotemporal consistency filter(STF).

The Figure 5 illustrates the STF superiority to correct filter motion vectors in highly noisy situations. In the last presented frame, there are four persons in the scene: two walking, one standing still, and other lying down. There is not enough motion information to detect the persons standing and lying down, even in the original motion vectors(first column). The median filter wrongly filters (removes) one person entering in the scene, deforms other person boundaries, and is not capable of removing other noisy motion vectors in the image. Nevertheless, the STF is capable of correctly filtering two persons walking, keeping their correct boundaries, and preserving their different directions patterns, $(0, -1)$ and $(3, 0)$, what make easier the following segmentation steps.

The results obtained from the spatiotemporal filter make it suitable as a first step of any system that aims to detect and track objects from compressed video using its motion vectors.

5. Conclusion

A novel spatiotemporal motion-vector filter was presented in this paper. The proposed filter has evaluated qualitatively and quantitatively producing good results, specially when compared with the widely used vector median filter. The results obtained with the spatiotemporal filter make it suitable as a first step of any system that aims to detect and track objects from compressed video using its motion vec-



(a)No mv filter *(b)Median Filter* *(c)Spatiotemporal Filter*

Figure 5. Object blobs and resulting motion vectors in sequence CAVIAR Fight OneManDown, using different motion-vector filters. Frames displayed: 190, 290, 380

tors.

Future work based in the proposed filter should test and improve it for the moving camera scenarios. Other possible works consist in evaluating the filter as part of a complete object tracker system based in motion-vectors.

References

[1] R. V. Babu, K. R. Ramakrishnan, and S. H. Srinivasan. Video object segmentation: a compressed domain approach. *IEEE Transactions on Circuits and Systems for Video Technology*, 14:462–474, 2004. 2, 3

[2] F. Bartolini, V. Cappellini, and C. Giani. Motion estimation and tracking for urban traffic monitoring. In *Image Processing, 1996. Proceedings., International Conference on*, volume 3, pages 787–790, Lausanne, Sept. 1996. IEEE. 1, 2

[3] G. R. Bradski and J. W. Davis. Motion segmentation and pose recognition with motion history gradients. *Mach. Vision Appl.*, 13(3):174–184, 2002. 2

[4] R. De Sutter, K. DeWolf, S. Lerouge, and R. Van de Walle. Lightweight object tracking in compressed video streams demonstrated in region-of-interest coding. *EURASIP J. Appl. Signal Process.*, 2007(1):59–59, 2007. 2

[5] M. Durik and J. Benois-Pineau. Robust motion characterisation for video indexing based on mpeg2 optical flow'. In *Proceedings of Second International Workshop on Content-Based Multimedia Indexing*, pages 57–64, Brescia, Italy, 2001. 2

[6] L. Favalli, A. Mecocci, and F. Moschetti. Object tracking for retrieval applications in mpeg-2. ***IEEE : Transactions***

on Circuits and Systems for Video Technology , 10(3):427–432, 2000. 2

[7] R. Fisher. EC Funded CAVIAR project/IST 2001 37540: Context Aware Vision using Image-based Active Recognition, 2010. Available from ¡`http://homepages.inf.ed.ac.uk/rbf/CAVIAR`¿. Accessed 30 April 2010. 6

[8] I. Haritaoglu, D. Harwood, and L. S. Davis. W4: real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):809–830, Aug. 2000. 1

[9] C. Käs and H. Nicolas. An approach to trajectory estimation of moving objects in the h.264 compressed domain. In *PSIVT '09: Proceedings of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology*, pages 318–329, Berlin, Heidelberg, 2008. Springer-Verlag. 2

[10] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):319–336, Feb. 2009. 4, 6

[11] E. T. Kim and H.-M. Kim. Efficient linear three-dimensional camera motion estimation method with application to video coding. *Journal of Optical Engineering*, 37:1065–1077, Mar. 1998. 2

[12] F. Manerba, J. Benois-Pineau, R. Leonardi, and B. Mansencal. Multiple moving object detection for fast video content description in compressed domain. *EURASIP J. Adv. Signal Process*, 2008:5, 2008. 1, 2

[13] V. Mezaris, I. Kompatsiaris, N. V. Boulgouris, and M. G. Strintzis. Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(5):606–621, May 2004. 1, 2

[14] R. Wang, H.-J. Zhang, and Y.-Q. Zhang. A confidence measure based moving object extraction system built for compressed domain. In *Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on*, pages 21–24, Geneva , Switzerland, 2000. IEEE. 2

[15] T. Yu and Y. Zhang. Retrieval of video clips using global motion information. *Electronics Letters*, 37(14):893–895, July 2001. 2

(a)No mv filter *(b)Median Filter* *(c)Spatiotemporal Filter*

Figure 4. Object blobs and resulting motion vectors in sequence PETS2001 Camera1,using different motion-vector filters. Frames displayed: 570, 930, 2110