I-076

Relevant Motion Detection From Motion Vectors in MPEG Stream

J. A. H. Garcia†, H. Morita‡, H. Yu‡, I G. B. B. Nugraha‡, H. P. Meana†

1. Introduction

Object detection and tracking is a research area in image processing field, which applications can be found on surveillance and security system, such as office building monitoring, parking area surveillance, airport security, and so on [1]. To detect objects, some algorithms have been proposed by segmenting a spatial and temporal data based on the similarity in color, texture, shape, and optical flow. However, classifying objects by using such kind of algorithms is quite expensive since they must do extensive calculation to search for an object in each video frame.

Therefore, another approach had been proposed by utilizing the information contained in a compressed video data, particularly the motion vector data. The main advantage of this approach is that decompression is not required so that it saves processing time. We only need to acquire the motion vector data of each macroblock in a frame and then do analysis based on those data.

There exist some proposed object detection algorithm based on motion vectors data, e.g., [2, 3, 4]. Reference [2] proposed independent motion detection by applying normal flow method [5] to the motion vectors data. The work in [3] also proposed an object detection algorithm by dividing a frame into grids where each grid represents a macroblock, and tracks the movement of each grid at the subsequent frames of an MPEG-2 [6] data. The tracking algorithm in [4] uses motion vectors data to track an object boundary.

In this paper, we propose an algorithm to detect human objects by taking the relevant motion vectors and filtering out the irrelevant motion vectors from a monitoring area. The relevant motion is selected by calculating the direction and the magnitude of the motion vectors data in an MPEG-2 stream where the magnitude of the motion vectors for a moving person typically has a certain range of values. By calculating on the number of such vectors in a frame, the number of persons in a scene also can be counted. Although we implement the algorithm on MPEG-2 data, it also can be used on MPEG-4 [7] video data since MPEG-4 video data also contains motion vectors information that is encoded in a similar way as in MPEG-2.

The organization of this paper is arranged as follows.

†Department of Postgraduate Students, Instituto Politecnico Nacional, Mexico City, Mexico.
Email: {jahdezg,hmpm}@calmecac.esimecu.ipn.mx
‡Graduate School of Information Systems, The University of Electro-Communications, Tokyo, Japan.
Email: {morita,baskara}@is.uec.ac.jp

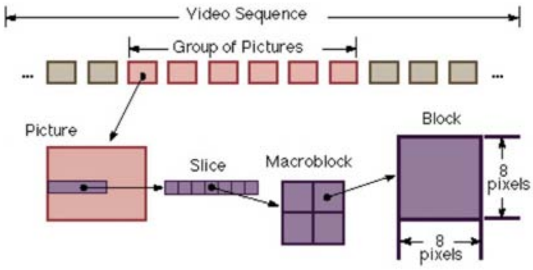


Figure 1: MPEG-2 video sequence.

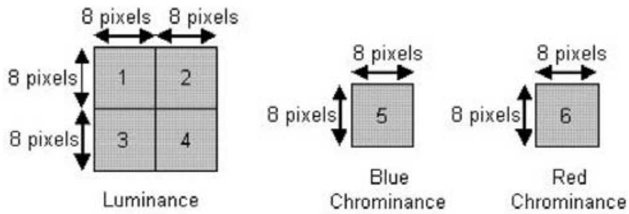


Figure 2: Macroblock structure

Section 2. gives brief explanation on the MPEG-2 data structure. Section 3. describes our algorithm. An experiment results will be shown on Section 4.. Finally, Section 5. gives the conclusion remarks.

2. MPEG Data Structure

An MPEG-2 video sequence can be divided into five layers: Group of Picture (GOP), Picture, Slice, Macroblock, and Block layer [8], as shown on Figure 1. Block is the smallest unit in MPEG-2 structure, which contains information from 8×8 pixels of an image. For example, a grey-scale image with 320×240 pixels has 1200 blocks. A macroblock contains 16×16 pixels or 4 blocks of the luminance component of the image and its corresponding chrominance blocks. Figure 2 depicts an example of a macroblock structure for a 4:2:0 picture format.

Picture layer contains information about each picture or frame. There exist three types of frames: I-frame, P-frame, and B-frame. Each block in I-frame is encoded by using intraframe coding without exploiting redundancy with the other frames. P-frame tries to exploit redundancy with the preceding frame. For each macroblock in a P-frame, a 16×16 pixels area is searched in the preceding frame to find the closest match with the macroblock. The search area is limited

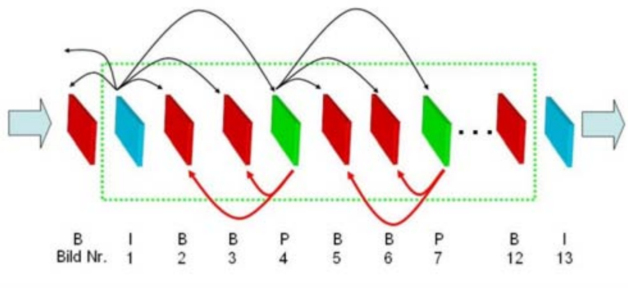


Figure 3: Group of pictures (GOP)

to 15 pixels both horizontally and vertically. The difference of the values of the pixels between the macroblock and the 16 × 16 pixels area in the preceding frame is then encoded. Information about the spatial distance of that best matching area from the macroblock is also encoded as *motion vectors*, which are composed of horizontal and vertical values. The B-frame refers to the preceding frame and the succeeding frame to exploit redundancy. Hence, B-frame has two motion vectors.

A GOP is a set of consecutive pictures or frames. The number of pictures in a GOP is usually set to be 12 or 15 frames where each GOP has one I-frame followed by a composition of P-frames and B-frames as shown on Figure 3.

3. Algorithm

A relevant motion is defined as a motion of an object to be considered. For example, if the purpose of the object detection is to count the number of vehicles the pass through a toll gate, then the relevant motions is the motions of the vehicles. Other motions, such as the motion of a tree, the people, and so on, are called irrelevant motions. The motions can be detected in the spatial domain of the images or in the compressed domain by analyzing the motion vectors data. In this paper, we propose a relevant motion detection algorithm by using the latter technique in an MPEG video data, particularly by examining the motion vectors in the P-frames of the MPEG data.

Detecting a relevant motion in by analyzing the motion vectors is not an easy task. The motion vectors in the MPEG data actually do not represent the movement of objects in the scene, but they represent the pointers to the locations of the best matched pixels values. So, although an object in the scene moves to a certain direction, the motion vectors of the macroblocks that compose the object may have various directions. Even the non-moving objects also have motion vectors because of the uncontrollable lighting variations. Figure 4 depicts the motion vectors in an MPEG-2's P-frame where the image shows someone that is moving to a direction. As the pixel values of the moving object

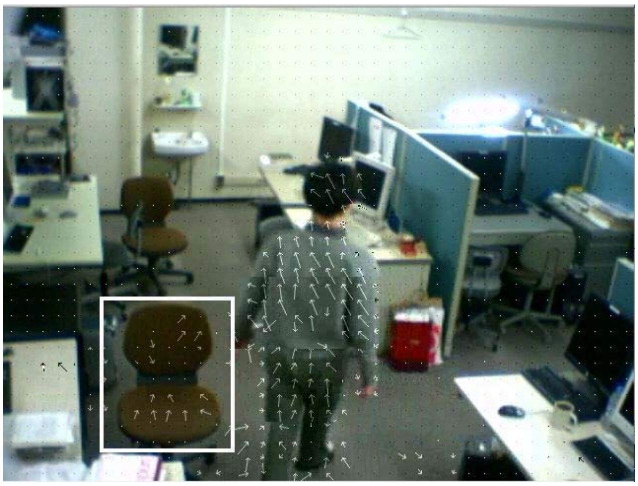


Figure 4: Motion vectors in an MPEG-2's P-frame

changes, they yield some motion vectors. However, a non-moving object also can produce motion vectors as depicted on the figure, particularly in the white rectangle part, due to color change caused by the man's shadow.

Although such noise vectors exist, a moving object can be recognized from the scene by calculating the summation of the magnitude of the motion vector in each macroblock for each frame. Figure 5(a) shows an object (human) that is moving in front of the camera and its corresponding total magnitude of the motion vectors on Figure 5(b) for that scene that consists of 250 P-frames. When the object comes onto the scene, the graph on Figure 5(b) shows that the total magnitude of motion vectors increases significantly.

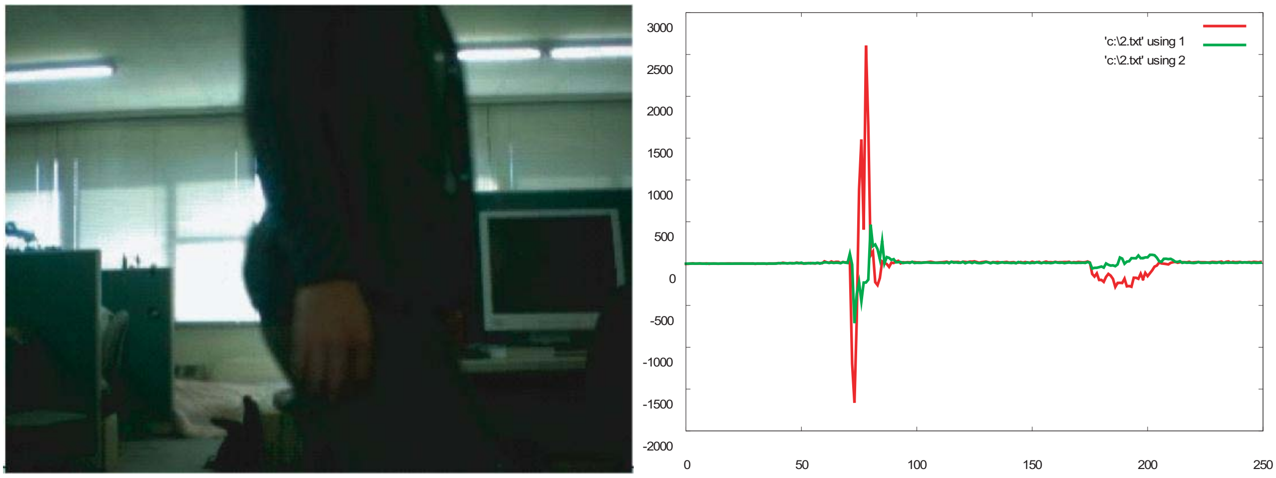
To track a moving object, we have to select the relevant motion vectors among the P-frames and calculate the centroid of the origins of the motion vectors. The centroid is the center of the moving object. However, since the vectors composition of the object may change irregularly over the time, we apply a kind of median filter to a sequence of motion vectors in every P-frames by using the following algorithm:

Step 1: Select the relevant motion vectors x_i among all motion vectors associated with macroblocks in a frame whose magnitudes are greater than a determined threshold value T and satisfy $|x_i - x_{i-1}| < \delta$ and $|x_{i-1} - x_{i-2}| < \delta$.

Step 2: Then, we calculate the centroid of those relevant motion vectors x_i.

4. Experiment Results

We implement the algorithm on an MPEG-2 data. The data is obtained from a camera that records an area in our laboratory. When someone is coming into



(a) (b)

Figure 5: (a) A moving human object on a scene and (b) the total magnitude of the motion vectors for each P-frame in that scene.

(a) (b)

Figure 6: Human movement tracking

the area, the algorithm selects the relevant motion vectors and calculate the centroid. Figure 6 depicts the result where a cross sign that indicates the centroid is shown. The algorithm is able to track the human movement inside the scene.

One other advantage of the algorithm is that it is also able to refine the trajectory of the object as shown on Figure 7. The figure shows that irregularity of the path can be smoothed out.

5. Conclusion

We have shown an object detection and tracking algorithm that works by selecting the relevant motion vectors data from an MPEG video frames, particulary P-frames. The object is tracked by calculating the centroid of the relevant motion vectors. The proposed algorithm is also able to refine the trajectory of the moving object.

In the future work, we plan to add additional information from the MPEG data, such as macroblock type and the information from the B-frames, to improve our detection algorithm.

Acknowledgment

We would like to thank to the Ministry of Education Science and Culture of Japan, the National Science and Technology Council (CONACYT) and National Polytechnic Institute Mexico City for the support during the realization of this research. This work was part of a joint research project between The University of Electro-Communications and Funai Electric Co., Ltd.

Figure 7: Refinement of the trajectory of the moving obejct

References

[1] B.J.T. Jones, ASTRA Developments Ltd, "Low-Cost Outdoor Video Motion and Non-Motion Detection, Proceeding of Security Technology, 1995, pp. 376-381.

[2] Z. Zhang, "Mining Surveillance Video for Independent Motion Detection," Proceeding of IEEE International Conference on Data Mining, 2002, pp. 741-744.

[3] L. Favalli, A. Mecocci and F. Moschetti, "Object Tracking For Retrieval in MPEG2," IEEE Trans. on Circuit and Systems for Video Technology, 2000, pp. 427-432.

[4] K. Hariharakrishnan, D. Schonfeld, P. Raffy and F. Yassa, "Video Tracking Using Block Matching," Proceeding of International Conference on Image Processing, 2003, pp. 945-948.

[5] R. Sharma, Y. Aloimonos, "Early detection of independent motion from active control of normal image flow patterns," IEEE Trans. On Systems, Man and Cybernetics, Part B, Vol. 26, Issue 1, Feb. 1996, pp. 42-52.

[6] ______, "Generic Coding of Moving Pictures and Associated Audio Information–Part 2: Video," Int. Standards Org./Int. Electrotech. Comm. (ISO/IEC), ISO/IEC 13818-2.

[7] ______, "Coding of Audiovisual Objects–Part 2: Visual," Int. Standards Org./Int. Electrotech. Comm. (ISO/IEC), 14496-2.

[8] P.N Tudor, "MPEG-2 Video Compression: Tutorial," Electronics and Communication Engineering Journal, December 1995, pp. 1-5.