Computation of Motion Activity Descriptors in Video Segments

JungHwan Oh *Praveen Sankuratri*

Department of Computer Science and Engineering
University of Texas at Arlington
Arlington, TX 76019-0015 U. S. A.
{

Abstract

Recently, *motion activity* which is defined as amount of motion in a video sequence has been included as a descriptor in MPEG-7 standard. The motion activity descriptors (MADs) which describe this motion activity need to enable efficient content analyzing, indexing, browsing, and querying of video data. To address this issue, first, we propose a novel technique for automatic measurement of motion activity using accumulation of quantized pixel differences among the frames of given video segment. As a result, accumulated motions of shot are represented as a two dimensional *matrix*. Also, we investigate an efficient and scalable technique to compare these matrices and generate MADs that are representing various motions of shots effectively. Not only the degrees (amounts) but also the locations of motions are computed and presented accurately. Our preliminary experimental studies indicate that the proposed techniques are effective in capturing and comparing motion activities.

KEYWORDS: Motion activity, Motion activity descriptors, Video content analysis, Video similarity model, Video retrieval.

1 Introduction

As a result of the rapid advances in data compression, storage devices and communication networks, multimedia, in particular video media, has become an integral part in many fields including education, business and entertainment [1]. This brings about the researches on content analyzing and indexing of videos for effective browsing, retrieval, filtering, and summarizing [2, 1]. One of the most distinguishable features which video has is *motion*. However, this motion information is relatively less examined than the other features since the computation (i.e., optical flow) is expensive, and it is not easily applicable to natural video in which there are mostly little restrictions on camera operation, object(s) and background. The overall motion which is generated from camera motion and/or object(s) motion in video has been measured and manipulated for content analyzing and indexing purposes [3, 4, 5, 6].

Recently, *motion activity* which is defined as the perceived subjective degree of activity, or amount of motion, in a video sequence [7], has been included as a descriptor in MPEG-7 standard [8]. This motion activity which can give a little more detail information about motion than the overall motion estimation has been investigated in numerous researches. A combination of image and audio features [9], a mode of motion vector magnitudes [10], a tangent distance between consecutive frames [11], and mean, variance, and median of motion vector magnitudes [12] are used to determine the motion activity levels of video segments.

The motion activity descriptors (MADs) should be able to capture different characteristics from different video segments. To address this issue, first, we propose a novel technique for automatic measurement of motion activity using accumulation of quantized pixel differences among the frames of given video segment (i.e., shot which is defined as a collection of frames recorded from a single camera operation). As a result, accumulated motions of shot are represented as a two dimensional *matrix*. Also, we investigate an efficient and scalable technique to compare these matrices and generate MADs that are representing various motions of shots effectively. Their main contributions can be summarized as follows.

- This matrix is showing not only the amounts but also the exact locations of motions. Therefore, we can get more accurate and richer motion information of shot.

- Because the proposed matrix comparison algorithm is very efficient and scalable, it can provide various ranges of clustering for shots which is essential tool for content analyzing, indexing, browsing, and querying of video data.

- It is very cost-effective because it uses accumulation of quantized pixel differences, and expensive computation (i.e., optical flow) is not necessary.

The remainder of this paper is organized as follows. In Section 2, we propose a novel technique for automatic measurement of motion activity, and discuss how to compute MADs automatically. The experimental results are discussed in Section 3. Finally, we give our concluding remarks in Section 4.

2 Computation and Description of Motion Activity

In this section, we introduce a novel technique for automatic measurement of *motion activity (MA)* in not only two consecutive frames but also whole shot which is a collection of frames. As a result, accumulated motions of shot are represented as a two dimensional *matrix*. Then, we discuss how to generalize and describe this matrix for the purpose of indexing and comparing each other.

2.1 Motion Activity Matrix

The MA for a shot with n frames is computed using the following steps. We assume that the frame size is $c \times r$ pixels.

Step.1 The color space of each frame is quantized (i.e., from 256 to 64 or 32 colors) to reduce false detection of motion by noise which is not actually motion but detected as motion.

Step.2 An empty two dimensional matrix which has the same size ($c \times r$) of the frame is created and initialized with zeros. For convenience, this matrix is called *Motion Activity Matrix (MAM)*.

Step.3 Compare all the corresponding pixels of two consecutive frames. If they have different color, increase the matrix value in the corresponding position by one (this value may be larger according to the other conditions). Otherwise, it remains without any increasing or decreasing.

Step.4 Step.3 is repeated until all consecutive pairs of frames are compared.

To visualize the computed MAM, we can convert this MAM to an image which is called *Motion Activity Matrix Image (MAMI)*. Let us convert an MAM with the maximum value m into a 256 gray scale image as an example. If m is greater than 256, m and other values are scaled down to fit into 256 as maximum, otherwise, they are scaled up. But the value zero remains unchanged. An empty image with same size of MAM is created, and the corresponding value of MAM is assigned as a pixel value. For example, assign white pixel for the matrix value zero which means no motion, and black pixels for the matrix value 256 which means maximum motion in a given shot. Each pixel value for an $MAMI$ can be computed as follows if we assume that $MAMI$ is a 256 gray scale image.

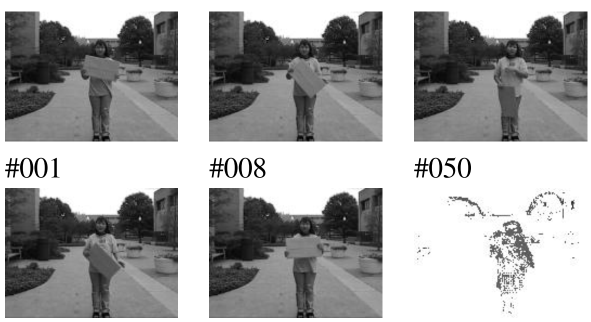
$$Each\ Pixel\ Value\ = \\ 256\ -\ Corresponding\ Matrix\ Value \quad (1)$$

Figure 1 shows the first and the eighth frames in a shot, and the $MAMI$ for those frames #1 through

#8. As seen in this figure, there is not much motion. The several frames and the $MAMI$ for whole shot (from the first frame (#1) to the last (#99)) are shown in Figure 2. These figures illustrate that the proposed technique can compute not only the exact amount (or degree) but also the region of motions in a shot.

#001 #008 #$MAMI$

Figure 1: Frames in a shot and its MAMI



#001 #008 #050

#070 #099 #$MAMI$

Figure 2: Frames in a shot and its $MAMI$

2.2 Motion Activity Descriptors

For querying or indexing purpose, we basically need to compare shots, in other words, MAMs need to be compared with other MAMs. In this subsection, we discuss a technique to compare shots in terms of amounts (degrees) as well as regions of motions.

The amount of motions can be compared by *Total Motion (TM)* computed from the corresponding MAM. Assume that MAM_A and TM_A are motion activity matrix and total motion for shot A respectively. MAM_A and TM_A are defined and computed as follows (where, a_{ij} is obtained by Step.1 through Step.4 in the previous subsection).

$$MAM_A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1c} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2c} \\ \dots & \dots & \dots & \dots & \dots \\ a_{r1} & a_{r2} & a_{r3} & \dots & a_{rc} \end{pmatrix} \quad (2)$$

$$TM_A = \sum_{i=0}^{r} \sum_{j=0}^{c} a_{ij} \quad (3)$$

Also, if we cluster the shots (MAMs, eventually) based on the *Total Motion (TM)* computed from the corresponding MAM, we can reduce inappropriate comparisons among shots which have very different amounts of motions. Consequently, we can save the computation cost significantly.

However, comparing only by TM does not give very accurate results because it ignores the locality such that where the motions occur. We introduce a technique to capture locality information without using partitioning, which is described as follows. In the proposed technique, the locality information of MAM can be captured by two one dimensional matrices which are the summation of column values and the summation of row values in MAM. These two arrays are called as *Summation of Column (SC)* and *Summation of Row (SR)* to indicate their actual meanings. The following equations show how to compute SC_A and SR_A from MAM_A.

$$SC_A = (\sum_{i=0}^{r} a_{i1} \;\; \sum_{i=0}^{r} a_{i2} \;\; \dots \;\; \sum_{i=0}^{r} a_{ic})$$
$$SR_A = (\sum_{j=0}^{c} a_{1j} \;\; \sum_{j=0}^{c} a_{2j} \;\; \dots \;\; \sum_{j=0}^{c} a_{rj})$$

Figure 3 shows some examples such that how these SC and SR can capture where the motions occur. Two SRs in Figure 3 (a) are same, which means that the vertical locations of two motions are same. Similarly, Figure 3 (b) shows that the horizontal locations of two motions are same by SCs. Figure 3 (c) is showing the combination of two, the horizontal and vertical location changes.

Therefore, we propose these TM, SC, and SR as Motion Activity Descriptors (MADs) in this paper. Now, we generalize shot similarity model using these three MADs. Assume that we compare MAM_A from shot A with MAM_B from shot

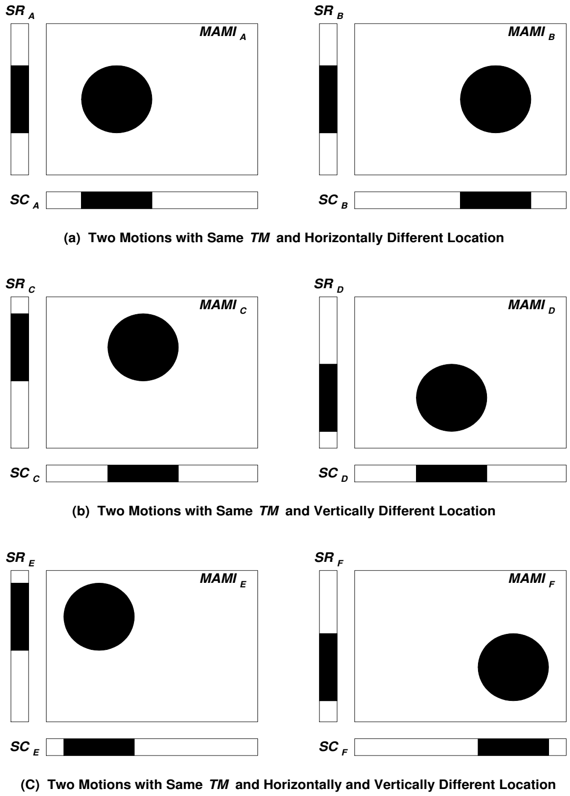


Figure 3: Comparisons of Locations of Motions

B. The similarity ($Sim_{A,B}$) between MAM_A and MAM_B can be computed using the following equation.

$$Sim_{A,B} = \omega_1 \times (|TM_A - TM_B|) + \omega_2 \times (|SC_A - SC_B|) + \omega_3 \times (|SR_A - SR_B|) \quad (4)$$

where ω_1, ω_2, and ω_3 are weighted factors, and their summation ($\omega_1 + \omega_2 + \omega_3$) is 1.0. The computation of $|SC_A - SC_B|$ is that the summation of the differences of the corresponding matrix values.

3 Experimental Results

Our video clips in the test set were originally digitized in AVI format at 30 frames/second. Their resolution is 160×120 pixels. Our test set has 68 shots which consist of total 12,399 frames as shown in Table 1. They were divided into 5 different categories, and their details are explained in the table. In this experiment, we first compute MAM (and $MAMI$) for each shot, and TM, SC and SR are extracted from this MAM of each shot. The average values of TM per category is also shown in the fourth column of Table 1.

We already showed an example of MAM (and $MAMI$) of a shot in the category 3 (see Figure 2 in the previous section). Figure 4 shows that an example of a shot with much more motions than the previous one which is indicated by the $MAMI$ in the figure. To visualized SC and SR, we plot them in Figure 5, in which x-axis presents all columns and rows, and y-axis presents the values of SC and SR. An object is walking from left to right at a constant speed in this shot, which can be interpreted that there is a constant horizontal motion. Interestingly, this content is described by the SC curve in Figure 5. The last example (Figure 6) shows a shot with a zoom out camera motion. The $MAMI$ represents this camera motion clearly. Two curves (SC and SR) in the Figure 7 also show this camera motion, in which they are more motions in the edges and less motions in the middle.

Figure 4: Frames in Shot # 53 and its $MAMI$

4 Concluding Remarks

In this paper, first, we propose a technique to measure motion activity in a shot automatically using a two dimensional matrix. Not only the degree (amount) but also the location of motions are computed and presented accurately. The other technique to compare these matrices efficiently is also proposed. In the technique, the amounts and the locations of motions are compared with by TM, and

Category No.	Total No. of Shots	Total No. of Frames	Average of TM	Category Description
1	7	1,115	10.1	No camera motion, and No specific object.
2	24	5,981	30.9	No camera motion, and One object which is not moving.
3	21	2,983	61.4	No camera motion, and One object which is moving a little.
4	12	567	462.1	No camera motion, and One or two objects which are moving much.
5	4	1,753	181.2	Camera zoom in and out
Total	**68**	**12,399**		

Table 1: Test Set of Shots and Its Results for Average of TM

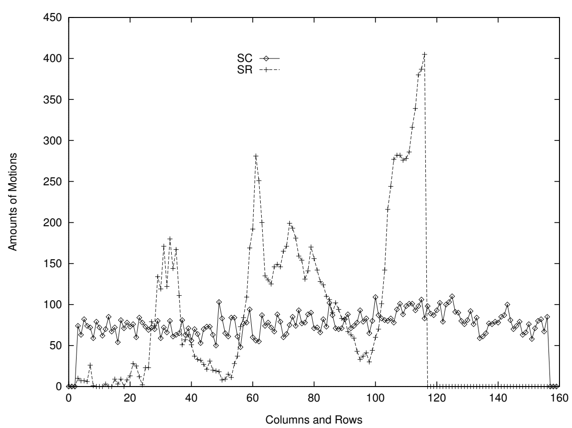


Figure 5: SC and SR for Shot # 53

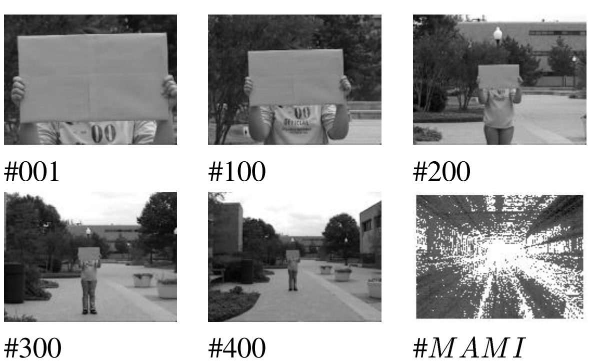


Figure 6: Frames in Shot #66 and its $MAMI$

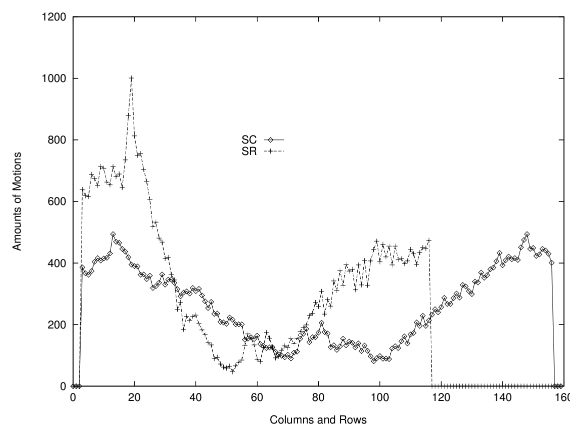


Figure 7: SC and SR for Shot# 66

SC and SR. Our preliminary experimental studies indicate that the proposed techniques are effective in capturing and comparing motion activity. We will perform further experiments in the future to study the effectiveness of matrix comparison technique using TM, SC and SR.

References

[1] R. Jain. Content-based multimedia information management. In *Proc. of 14th Int'l Conf. on Data Engineering*, pages 252–253, Orlando, Florida, February 1998.

[2] P. Aigrain, H. Zhang, and D. Petkovic. Content-based representation and retrieval of visual media : A state-of-the art review. *Multimedia Tools and Applications*, 3:179–202, 1996.

[3] Y. Tan, S. Kulkarni, and P. Ramadge. Rapid estimation of camera motion from compressed video with application to video annotation. *IEEE Transactions on Circuit and Systems for Video Technology*, 10(1):136–146, 2000.

[4] V. Kobla, D. Doermann, K. Lin, and C. Faloutsous. Compressed domain video indexing techniques using dct and motion vector information in mpeg video. In *Proc. SPIE Vol. 3022, Storage and Retrieval for Image and Video Databases VII*, pages 200–211, San Jose, CA, January 1997.

[5] E. Ardizzone et al. Video indexing using mpeg motion compensation vectors. In *Proc. IEEE International Conference on Multimedia Computing and Systems*, 1999.

[6] J. G. Kim, H. Chang, J. W. Kim, and H. Kim. Efficient camera motion characterization for mpeg video indexing. In *Proc. of 2000 IEEE International Conference on Multimedia and Expo., 2000*, pages 1171–1174, New York, NY, July 2000.

[7] A. Divakaran, K. Peker, and H. Sun. A descriptor for spatial distribution of motion activity. In *Proc. of ICIP 2000*, Canada, 2000.

[8] International Organization for Standardization. Overview of the mpeg-7 standard. In *ISO/IEC/JTC1/SC29/WG11 N4031*, Mar. 2001.

[9] S. Pfeiffer, R. Lienhart, S. Fischer, and W. Effelsberg. Abstracting digital movies automatically. *Journal of Visual Communication and Image Representation*, 7(4):345–353, December 1996.

[10] W. Wolf. Key frame selection by motion analysis. In *Proc. of ICASSP 96*, pages 1128–1231, 1996.

[11] N. Vasconcelos and A. Lippman. Towards semantically meaningful features spaces for the characterization of video content. In *Proc. of ICIP 97*, 1997.

[12] K. Peker, A. Divakaran, and T. Papathomas. Automatic measurement of intensity of motion activity of video segments. In *Proc. of SPIE conf. on Storage and Retrieval for Media Databases 2001*, pages 341–351, San Jose, CA, Jan. 2001.