

# Generation of Video Panorama System

Kawther Abbas Sallal

Assistant Lecturer

Computer Science Dept.,  
Al-Mustansiriyah University, Baghdad, Iraq

Abdul Monem Saleh Rahma

Professor

Computer Science Dept.,  
University of Technology, Baghdad, Iraq

## ABSTRACT

The construction of large, high-resolution view is an active area of research in many fields like computer vision and computer graphics. A Panorama is the process of combining multiple images or videos with overlapping region to produce a panoramic view. it enables producing a complete view of an area that cannot fit in a single shot. This paper describes simple method for taking two or more videos and creating a panoramic video. Two steps of motion estimation is done. The first, estimates the motion between the frames of the two videos and the second estimates the motion to produce the final panoramic video. The process to generate a panoramic view can be divided into four main components: video acquisition, similar regions detection, motion estimation and merging. Geometric moment invariant produces a set of features that are invariant under shifting, scaling and rotation. it is widely used to extract the global features for pattern recognition due to its discrimination power and robustness. In this paper, moment invariant is used to determine the locations of merging. The experimental results are done on videos taken by two horizontally adjacent cameras. The results show that the proposed algorithm is fast and efficient.

## Keywords

geometric moments, motion estimation, registration, panoramic video.

## 1. INTRODUCTION

Videos are becoming an increasingly popular means of conveying information, and as a result the availability of video editing tools is growing every day. The addition of the temporal dimension to the spatial domain provides a large new range of potential problems and possibilities. Aside from the natural extensions from image editing, video sequences can be adjusted or improved in other ways[1]. constructing a large, high-resolution views becomes an active area of research in the fields of computer vision, image processing, and computer graphics. Our contribution is introducing a new direction for developing the used multimedia and devising new Medias more arousing and attracting for viewers. Some researchers worked on designing the panoramic image, but this work aims to develop an algorithm to produce films of the same type.

Panoramic technology has received a growing interest. Among its applications, such techniques may be used to efficiently represent video sequences, for example, in compression, enhancement, visualization etc. Recent years have seen a growing interest in the visualization of the world's cities and sights. For example, systems such as Google Street View enable users to browse street level imagery by presenting a panorama visualization of video captured at street level from a moving vehicle [2].

constructing full view panoramas requires taking many regular photographic images or sequential video taken by rotated camera in order to cover the whole viewing space. These samples must then be aligned and merged into complete panoramic view.

Many advanced creation methods have been developed in recent years. [1] this work focuses on a common type of video sequence, in which the videographer shoots a scene by rotating the camera to capture the entire panorama, possibly zooming into areas of particular interest. They proposed a way of re-displaying these sequences, by giving the user control over a virtual camera frame. Based on video mosaicing, a static background panorama is computed first. After segmenting and removing the foreground subjects from the original video, the remaining elements are merged into a dynamic background panorama. then redisplay this augmented video by warping and cropping the panorama. The virtual camera can have an enlarged field-of-view and a controlled camera motion.

[3] the authors demonstrated that omni-directional video applications can be created efficiently using existing MPEG-4 technology by combining a spherical or cylindrical geometry and tiled omni-directional high resolution video streams or still image textures. Furthermore, a visibility sensor was attached to each video tile to provide a pre-fetching mechanism for efficient video memory usage and guarantee real-time rendering. The experiments showed, how to select the overlapping visibility sensor area in comparison to the size of the video tiles to provide real-time rendering. [4] this paper addresses the problem of constructing mosaics from video sequences taken by rotating cameras. In particular, it investigates the widespread case where the scene is not only static but may also contain large dynamic areas, induced by moving or deforming objects. A statistical derivation of feature-based and direct method classes shows that a mixed approach may solve the problem.

[5] This paper presented a developed software, which takes a single video sequence captured by panning the camera, and generates an animated panorama. The concept is to create a cylindrical panoramic image embedded with multiple video textures. The approach is finding an approximated optimal cut-surface in a video sequence, thus the animated panorama can be generated within a practically acceptable time.

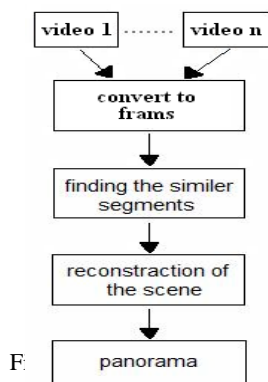
[6] This paper describes the development of a panoramic video system using off-the-shelf commodity hardware. This system captures multiple video sequences using multiple cameras, with each camera at a specified angle. Each video sequence has a limited field-of-view. The video sequences are merged to form a single wide field-of-view video. The system makes use of tools such as Barrel Transformation, Perspective Transformation, Edge Detection and Chamfer Distance Transformation. [7] proposed A method based on the invariance feature transform algorithm to stitch images captured by the turning cameras together to form panoramic

images. Based on the extracted features, the best bit first searching strategy is used to match feature points. Then in the Ransac algorithm is used to remove the mismatching feature points. Photos captured by a surveillance camera are taken as the input to test the proposed method. Irani et al., [8] propose mosaic based compression. A static panorama background is first constructed from the video sequence and then each video frame is compressed using the static panorama background as a reference. Furthermore, it detects and indexes the motion objects and provides content-based video indexing. Although they do not deal with on-demand transmission.

this work presents a simple and efficient method that takes a group of videos to generate the panoramic view by using moment invariants, which is to measure the shape characteristics. Through the use of nonlinear combination of geometric moments, one group of scale invariance, translation invariance and rotation invariance of moment invariants is obtained. Moment invariants have two roles, the first is determining the similar regions between videos and the second is determining the motion factor between the videos. The reminder of this paper is organized as follows. Section 2 describes the general system. Section 3 includes the proposed registration method. Section 4 explains the motion estimation and video panorama generation. Section 5 explains the experimental results and section 6 explains the conclusions.

## 2. THE PROPOSED SYSTEM

As illustrated in fig.(1) the first step is capturing two or more videos using normal cameras. Then the videos are divided into frames. After that, the similar segments between the frames will be determined. The final step is using the determined segments in reconstruction step to generate the panoramic scene.



The first step for making the panorama is to capture desired videos which are suitable for panorama view. The desired videos means that successive videos need to have roughly the same camera settings, enough overlap with each other and known camera parameters. A two hand held camera provides us with a 14 mega-pixel resolution is used. Since the videos are taken by the user. to satisfy the overlapping requirement, the two cameras are fixed on a slate with some distance (d) to obtain the overlapping between the captured images as shown in fig.(2) and fig.(3). The proposed algorithm is shown in fig.(4).

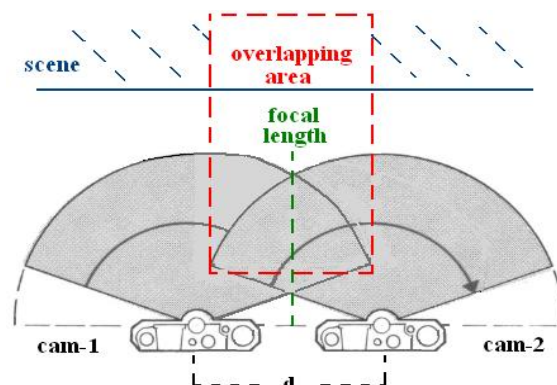


Fig.2 two videos acquisition using two cameras

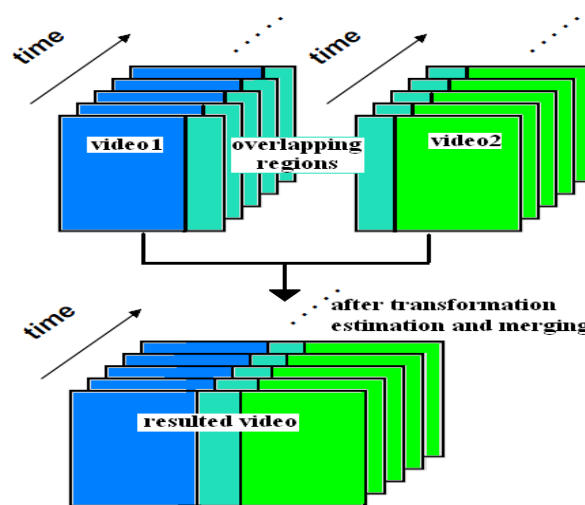


Fig.3. Panoramic video construction

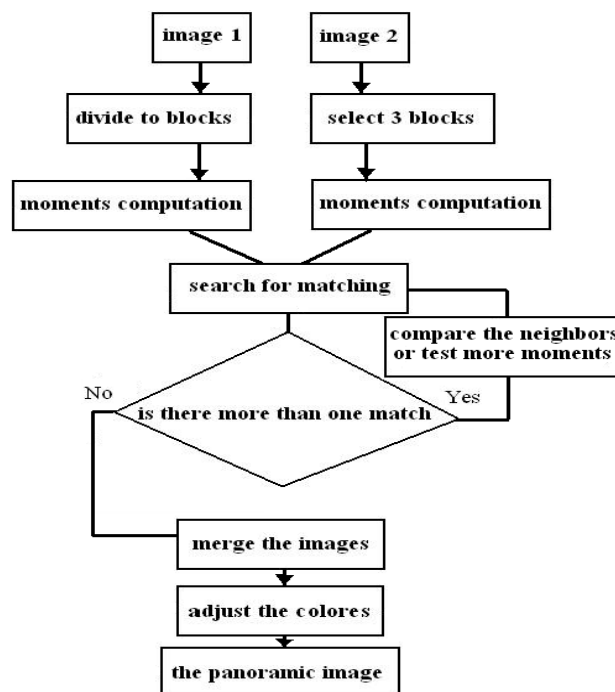


Fig.4 the proposed system

The basic steps of proposed panoramic video generation algorithm is illustrated below.

<b>Panoramic video Generation Algorithm</b>
<p><b>Input:</b> 1... N of videos (two videos are used).  <b>Output:</b> panoramic video.</p>
<p><b>Step 1:</b> convert the videos to frames.  <b>Step 2:</b> take the frames and divide them to non overlapped blocks.  <b>Step 3:</b> for each frame in the first video:                      -divide it to no. of blocks                      - compute the moments for each block.  <b>Step 4:</b> from each frame in the second video:                      -select (at least) three points not located on one straighten and compute the moments for it.  <b>Step 5:</b> compare the moments values in steps(3,4) above between every two frames in the videos to find the match.  <b>Step 6:</b> if there is one match go to step 7 , else                      If there is two moment values similar to the moment values of any one of the selected blocks then:                      - compare its neighbors or compute more moments or select another blocks.                      - Select the block that the moment values are similar to the moment values of the block in the other frame in the second video.  <b>Step7:</b> merge the frames according the extracted motion that based on the matched blocks between every two frames in the two videos.  <b>Step8:</b> adjust the color difference of the merged frames.  <b>Step 9:</b> convert the merged frames to video.  <b>END</b></p>

### 3. REGISTRATION BASED ON BLOCK'S GEOMETRIC MOMENT MATCHING

Registration is the process of matching two or more frames of the same scene. This requires estimating optimal geometric transformations to align the frames with respect to a common reference. The registration task is one of the most complex and challenging problems of image analysis, where the extreme diversity of images and working scenarios make impossible for any image registration algorithm to be suitable for all applications. The registration algorithms divided into two major types, feature-based and area-based methods.

Feature-based methods find relevant image features such as corners, line intersections, line ending or high-curvature points, that can be matched between images. Once an enough number of points have been matched by correspondence on two images, a suitable motion can be computed and applied to align them. Area-based methods is a correlation-based or template matching methods which finds correspondences between regions [9]. Area-based algorithms are appropriate in cases where registration points cannot be determined without high uncertainty. For example registration of medical and planetary images. Feature-based registration algorithms are suitable for images with high detail contents, where enough features can be detected in accurate way.

In this work, moment invariants is used to determine the matching blocks between frames and register them by determine the motion between them depending on the found matching blocks. Moment invariants are important shape descriptors. There are two approaches of shape descriptors: contour-based shape descriptors and region-based shape

descriptors. moment invariants are popular contour-based shape descriptor derived by Hu [10][11]. It was derived from the theory of algebraic invariant [12]. A 2-D continuous function  $f(x, y)$  of order  $(p+q)$  is defined as

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \dots\dots\dots (3.1)$$

For  $p, q = 0, 1, 2, \dots$

A uniqueness theorem states that if  $f(x, y)$  is piecewise continuous and has non zero values only in a finite part of  $xy$  plane, moments of all order exist and the moment sequence  $(m_{pq})$  is uniquely determined by  $f(x,y)$ . Conversely,  $(m_{pq})$  uniquely determines  $f(x,y)$ . The central moments can be expressed as

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \dots\dots\dots (3.2)$$

where  $\bar{x} = \frac{m_{10}}{m_{00}}$  and  $\bar{y} = \frac{m_{01}}{m_{00}}$

For a digital image, Equation (3.2) becomes

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \dots\dots\dots (3.3)$$

The normalized central moments, denoted  $\eta_{pq}$ , are defined as:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \dots\dots\dots (3.4)$$

where  $\gamma = \frac{p+q}{2} + 1 \dots\dots\dots (3.5)$

A set of seven invariant moments can be derived from the second and third moments:

$$\phi_1 = \eta_{20} + \eta_{02} \dots\dots\dots (3.6)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \dots\dots\dots (3.7)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \dots\dots\dots (3.8)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \dots\dots\dots (3.9)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \dots\dots\dots (3.10)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \dots\dots\dots (3.11)$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \dots\dots\dots (3.12)$$

The seven invariant moments, which are invariant to translation, scaling, mirroring and rotation, composed of the linear combination of the second-order and third-order central moments. Because seven moment invariants is relatively large, in order to facilitate comparison making use of logarithmic methods. At the same time, taking into account the possible negative moment invariants situation, you have to

get absolute value before getting logarithm. This can be done according the following equation:

$$M_i = -\log_{10}(\text{abs}(\emptyset_i)) \dots\dots\dots(3.13)$$

Where  $M_i$  is the moment descriptor, and  $\emptyset$  is the moments value and  $i= 1,2,\dots,7$ .

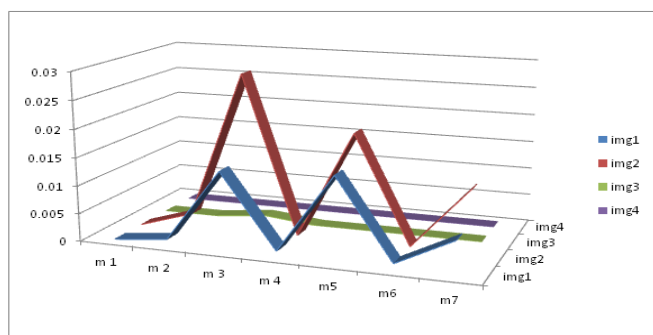
The example in fig. (5) shows the moment values for the lena image with different situations and the difference in moment values for the test images shown in fig.(6). From left top, Image1 is the original image. image2 is scaling with a scale factor 0.9. image3 is scaling with scale factor 0.4. from bottom left, image4 is scaling 0.9 and rotation with 30 degree. Image5 is rotation with 90 degree.



**Fig.5 different situation for image, top left (original, scaled 0.9, scaled 0.4). bottom left (scaling 0.9 and rotation 30, rotation 90)**

**Table1. the moment values for Lena images**

	$\emptyset 1$	$\emptyset 2$	$\emptyset 3$	$\emptyset 4$	$\emptyset 5$	$\emptyset 6$	$\emptyset 7$
Img1	2.876	8.175	11.932	10.947	23.768	15.068	22.387
Img2	2.876	8.176	11.946	10.946	23.753	15.069	22.393
Img3	2.876	8.178	11.961	10.946	23.748	15.069	22.399
Img4	2.876	8.175	11.933	10.947	23.768	15.068	22.387
Img5	2.876	8.175	11.932	10.947	23.768	15.068	22.387



**Fig.6 the difference in moment values**

The basic idea is extract the features of the first image by computing the moment values for all image blocks. Select some blocks (at least three not located on one straighten and located in the region of overlap) from the second image. These blocks are invariant for all the next video frames . The moment values are computed again for the selected blocks and then compared with the moment values of the first image

to find the match. If there is a match, the images are merged together according the extracted motion factor and according the following equation :

$$(x',y')=(((x*bs)+i)-m),(((y*bs)+i)-n))\dots\dots\dots(3.14)$$

Where  $x'$  and  $y'$  represent the exactly coordinates of the matched block in the first image,  $x$  and  $y$  represent the coordinates of the extracted block in the overlapping region,  $bs$  is the block size,  $i$  and  $j$  are the high and width of the image and  $m$  and  $n$  are the high and width of the overlapping region. Then, colors of them are adjusted. For each pair of potentially matching frames there is a set of feature matches that are geometrically consistent. Where the shape distance between the blocks of the two frames are computed to determine the matched moment values according the equation below:

$$D_{Moment}(R_1,R_2) = \sqrt{\sum_{i=1}^7 (M_i^{(1)} - M_i^{(2)})^2} \dots(3.15)$$

Where  $D$  is the distance,  $R1$  and  $R2$  is the two blocks used to find match.

similar moment values may be appears and this problem can cause false merging. Therefore, three ways are determined to find the exactly match, the first is comparing the moment values of the neighboring blocks. The second is, taking more moment values for the blocks. the third is selecting another blocks.

two images are related to each other by a geometric transformation. Geometric transformation is the relationship between two frames such that any given point in one image corresponds to only one point in the other. By recovering the geometric transformation, frames can be merged together and the panorama can be created. Alignment of frames may be imperfect due to registration errors resulting from incompatible model assumptions, dynamic scenes, etc. this problem appears when there is different transformations between the acquired images or videos. Furthermore, in most cases frames that need to be merged are not exposed evenly due to changing lighting conditions, automatic controls of cameras, printing/scanning devices, etc. therefore, further processing is needed to eliminate remaining distortions and discontinuities. These unwanted effects can be enhanced during the merging process. There are two popular methods of blending the images. The first is called alpha blending, which takes weighted average of two images. The cases that alpha blending works extremely well is when image pixels are well aligned to each other and the only difference between two images is the overall intensity shift. Alpha blending will merge two images seamlessly. The second method is Gaussian pyramid [13]. This method essentially merges the images at different frequency bands and filters them accordingly. The lower the frequency band, the more it blurs the boundary. Gaussian pyramid blurs the boundary while preserving the pixels away from the boundary. It does not work well if the two images have significantly different intensity levels. The transition is not as smooth as alpha blending for this case.

The test videos used in this work is taken by two camera positioned horizontally as shown in fig.(2). According this way of acquiring, two methods are proposed, the first is by dividing the overlapping region for the images to columns and then taking the in between columns for the two images as illustrated in fig.(7). The second is by smoothing the overlapping region between two images.

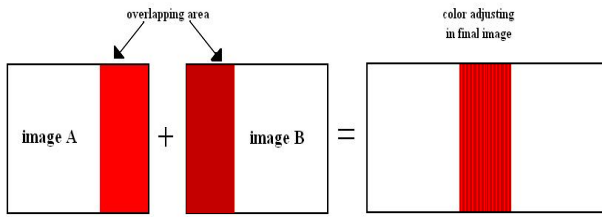


Fig. 7 the color adjusting of the merged frames

#### 4. MOTION ESTIMATION

The main objective of any motion estimation algorithm is to exploit the strong frame to frame correlation along the temporal dimension. The underlying supposition behind motion estimation is that objects and background in a frame of video sequence move within the frame to form corresponding objects on the subsequent frame[14]. Motion is described by a two-dimensional vector, known as motion vector that specifies where to retrieve a macro-block from the reference frames. The motion vector can be found using matching criterion. The motion vector that minimizes some cost measure (like SAD-Sum of Absolute Difference, MSE-Mean Square Error, MAE-Mean Absolute Error) involving the candidate and the target macro blocks is usually selected, where, The greater the similarity between the two matrices, the smaller the cost function values that result.

In block matching, the current frame is divided into blocks and each block is matched with a reference frame. The best matched block of pixels from the reference frame is then used in the current block. The simplest BMA, known as the full search or exhaustive search BMA, evaluates the cost function at every possible pixel location in the search area as in fig. (8). this way is more simple and accurate than other methods but it is computationally expensive because it evaluates cost function at every location in the search area.

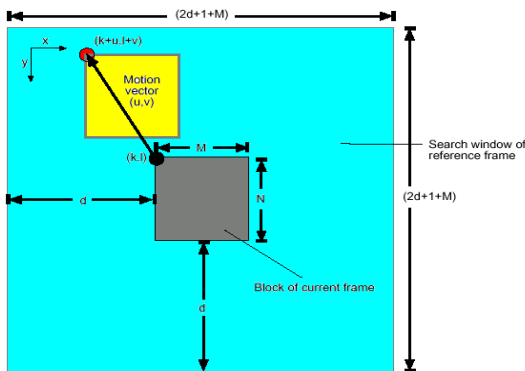


Fig.8 motion estimation

Therefore, a fast algorithm for motion estimation. In this work, the well known three step search is used. It became very popular because of its simplicity and also robust and near optimal performance. The general idea is represented in Fig.9. It starts with the search location at the center and sets the 'step size'  $S = 4$ , for a usual search parameter value of 7. It then searches at eight locations  $\pm S$  pixels around location  $(0,0)$ . From these nine locations searched so far it picks the one giving least cost and makes it the new search origin. It then sets the new step size  $S = S/2$ , and repeats similar search for two more iterations until  $S = 1$ . At that point it finds the location with the least cost function and the macro block at that location is the best match. The calculated motion vector is then saved for transmission.

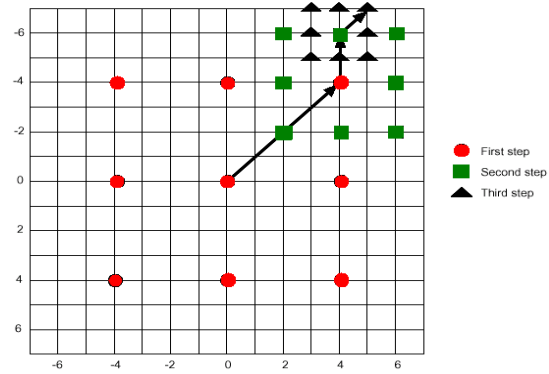


Fig.(9) three step search algorithm

Peak-Signal-to-Noise-Ratio (PSNR) given by equation (4.1) characterizes the motion compensated image that is created by using motion vectors and macro blocks from the reference frame[15].

$$PSNR = 10 \log_{10} \left( \frac{\text{Peak to peak value of original data}}{MSE} \right)^2 \quad (4.1)$$

$$MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (C_{ij} - R_{ij})^2$$

Where MSE is the mean square error. N is the side of the macro block,  $C_{ij}$  and  $R_{ij}$  are the pixels being compared in current macro block and reference macro block, respectively.

In this work, motion estimation is used in two places, the first is to estimate the motion between each two frames in the two acquired videos. The second place is in estimating the motion between the merged frames mentioned above to produce the panoramic video. In first place, the geometric moments are used to determine the true matches for merging and determining the motion factor while in the second place a three step search technique for motion estimation is used.

#### 5. EXPERIMENTAL RESULTS

The results of the algorithm are explained mainly on videos taken by two horizontally normal cameras. The test videos are converted to frames. first the RGB image is converted to grayscale image with size 512x512 to faster the execution. then the region of overlapping is determined to apply the moments invariants. Dividing the region of overlapping to blocks and then applying the moment invariants gives us fast and better results than applying the moment invariants on the complete image. After determining the matched blocks ,the exact position of them in the complete image will be determined according the equation (3.14). Fig.(10) and fig.(11) show a three frames (1, 14 and 23) from two test videos. The extracted similar regions are shown in fig.(12) and the differences between the moments value of each selected block from the second image with the moments values of the first image blocks are shown in fig.(13). The final result can be seen in fig. (14). Another example explains the results using different selection of blocks number are explained in fig.(15). As shown there is one false block found. therefore, there are three solutions to solve this problem, either selecting another blocks or examine the moment values of the block's neighbors or test more moments.



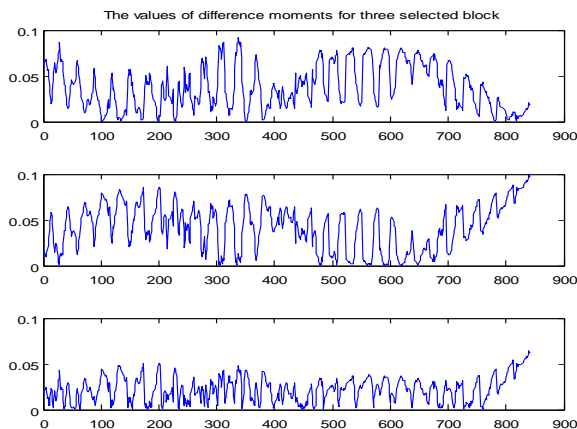
**Fig.10** some frames from video1(1,14,23)



**Fig.11** some frames from video2(1,14,23)



**Fig.12** the extracted matched blocks between two frames (frames no.14 in the two videos) in the overlapping region.



**Fig.13** the difference between the moment values of the three selected blocks in a frame with the moments values of all blocks of the overlapping region of other frame

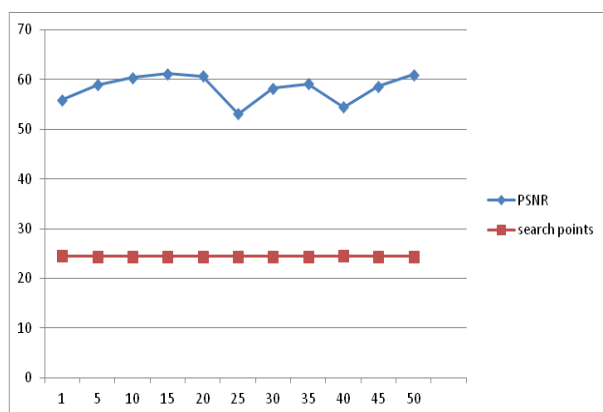


**Fig. 14** the case of false matching

merging two frames can be done in approximately (1.05) seconds for every two frames. The time increased with the number and/ or size of frames. For the two test videos each consists of 50 frames, the time for merging all frames is 52 seconds. The time for motion estimation and compensation by three step search of two merged frames is approximately 1.8 second. The result of motion compensation is shown in fig.(16) and the motion compensated image's PSNR and search point per block is shown in fig.(17).By comparing the execution time for the proposed method and other methods like SIFT and phase correlation, its found that the SIFT take about 122.5 seconds for merging 50 frames. While execution by phase correlation method expends 56.2 seconds for merging 50 frames.



**Fig. 15 motion compensation (from top to bottom: reference frame, current frame, resulted frame)**



**Fig.16 image's PSNR and search point per block performance**

## 6. CONCLUSIONS

Through the use of geometric moments, a group of scale invariance, translation invariance and rotation invariance of moment invariants are obtained. The moments method has a lower computational cost and also limited to the affine or any

simpler model. Intensity in the image does not change arbitrarily, but there may be a change in overall contrast due to changes in illumination or camera parameters. The results show that the proposed algorithm is efficient and speed.

The limitation of the proposed algorithm are, this method is not excellent while the distorted transformations occurred or the gray-value's difference between the reference image and real-time image is too large. But, as a future work, extending the proposed approach further by using matching based-on-feature which involving collection of features(such as point, line, and surface) of two or more images and descriptions of parameters to deal with this issue. Also decided to use another motion estimation method to make the generation of final video more faster than now.

## 7. REFERENCES

- [1] Hermans C. etl, "Augmented Panoramic Video" The Eurographics Association and Blackwell Publishing Ltd. Published by Blackwell Publishing ,2008.
- [2] Zheng E., Raguram R., Fite-Georgel P., "Efficient generation of multi-perspective panoramas", IEEE, International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 2011.
- [3] Heymann S., "representation coding and interactive rendering of high resolution panoramic images and video using MPEG-", Fraunhofer Institute for Telecommunications, Image Processing Department , Germany, 2005
- [4] Adrien B. etl, "From Video Sequences to Motion Panoramas", Workshop on Motion and Video Computing , United States, 2002.
- [5] Huang F. etl, "Animated Panorama from a Panning Video Sequence", IEEE, 2010.
- [6] Imran A. and James C., "A Panoramic Video System" Department of Computer Science University of the Western Cape, South Africa,2010.
- [7] Lin L. and Nan G., "Algorithm for Sequence Image Automatic Mosaic based on SIFT Feature", WASE International Conference on Information Engineering, 2010.
- [8] Irani M. and Anandan P., "Video indexing based on mosaic representations", in Proceedings of the IEEE, vol. 86, no. 5, pp. 905–921, May 1998.
- [9] Shalini G. P. "Principles of Computer Graphics: Theory and Practice Using OpenGL and Maya", Springer, 2004.
- [10] Hu M.K., "Visual pattern recognition by moments Invariants", IRE Trans. Information Theory, 8: 179-87,1962.
- [11] Rizon M. and et al., "Object Detection using Geometric Invariant Moment", American Journal of Applied Sciences , ISSN 1546-9239, 2006.
- [12] Bei J. and Chen L., "Map Matching Algorithms Based on Invariant Moments", Journal of Computational Information Systems, 2011.
- [13] Adelson E. and et al., "Pyramid methods in image processing," RCA Engineer, 29-6, 1984.
- [14] Kavitha G. and Shanmugam J., "Video Object Extraction Based on a Comparative Study of Efficient Edge Detection Techniques", The International Arab Journal of Information Technology, Vol. 6, No. 2, 2009.
- [15] Aroh B., "Block Matching Algorithms for Motion Estimation", IEEE, Final Project Paper, 2004.