

# Understanding Tracking and StroMotion of Soccer Ball

Nhat H. Nguyen  
Master Student  
205 Witherspoon Hall  
Charlotte, NC 28223  
704 656 2021  
[rich.uncc@gmail.com](mailto:rich.uncc@gmail.com)

## ABSTRACT

Soccer requires rapid ball movements. These movements passionate the viewers, but sometime the viewers are confused when the ball moves too fast. Capture the soccer ball movement makes viewing soccer more enjoyable and intuitive. Furthermore, having the opponent's ball movements, coaches can adjust his team strategy accordingly. In this paper, a soccer ball tracking technique, template matching, is proposed. StroMotion, an image decomposition technique, are used to display ball movements successfully from frame to frame. Two simple experiments are performed as short video sequences where a player juggles the ball. The results are 100% accuracy when tracking within 5 pixels and 85% accuracy when tracking within 4 pixels from the ground truth.

## Keywords

Template Matching, Ball Tracking, StroMotion, Sport Video Processing.

## 1. INTRODUCTION

The amount of visual information of the soccer is unique. Soccer ball image is a white circular shape, stands out from the background which usually the grass field. Due to variations in lighting conditions and ball location on the field, the brightness of the ball might differ. Intensity threshold to segment the soccer ball becomes difficult. However, if the shape and size of the ball will not change significantly, template matching is proved to be a simple technique which can solve this problem [1]. Template matching is a technique in Digital Image Processing for finding small parts of an image which match a template image. There are different approaches to implement template matching. In this paper, I present Linear Spatial Filtering [4] method, which performs convoluted operation between the template and the image to get a response score at each pixel.

According to Dartfish, a video software consultant company, StroMotion creates stunning video footage revealing the evolution of an athlete's movement, technique, execution, and tactics [2]. In this paper, StroMotion allows the viewers to appreciate the evolution of the movements of the ball in a manner intuitive and immediate. A soccer ball movement is unfolded in time and space into a frame by frame sequence. StroMotion transform a rapid movement of an object into a series of static images along that object's trajectory. In this paper, an attempt to implement StroMotion is successfully demonstrated through the testing video sequence.

Particularly, I propose an automated technique to track the soccer ball from video sequences. I believe this method can be extended

to further applications. The proposed tracking method is robust to noise since it can be performed on an image sequence where there are other movements beside the ball. I illustrate my method is section 2. Specifically, Basic Processing, Template Matching, Connected Component Labeling and Region segmenting technique are discussed in section 2.1, 2.2, 2.3 and 2.4 respectively. Contour Tracking and StroMotion Displaying are demonstrated in section 2.5, and 2.6. The experiment, along with result, is shown in section 3. Finally, a discussion of the result and future work are in section 4.

## 2. PROPOSED METHOD

### 2.1 Basic Processing

In order to do template matching, we first perform some basic image processing techniques on each frame of the video sequence. The pre-processing techniques include grayscale converting, smoothing, and normalizing. First, an input of 3-color channel image from the video sequence is converted into grayscale where most of the computer vision techniques can be applied. Second, a median filter, a useful technique for reducing salt-and-pepper noise in the image [4], is performed in the grayscale image. Third, a normalizing operation is necessary to scale the entire image pixel intensity values to range 0 to 1. The pre-processed frame, shown in Figure 1, is now ready for segmentation.

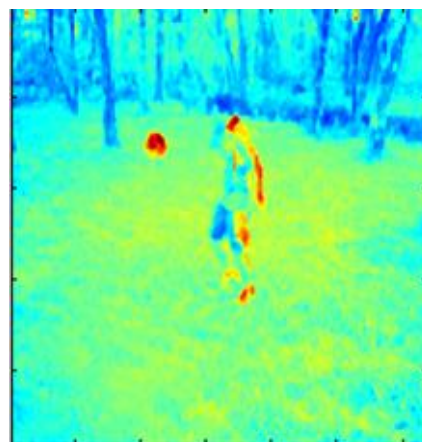


Figure 1. Pre-processed frame displayed in Jet color map

## 2.2 Template Matching

### 2.2.1 Generating Soccer ball template

Based on the general description of the soccer ball, a template, or ball kernel, is created. This template has to take into account the distance of the ball to a fixed camera to calculate the size of the ball image. In my experiment, a 20 X 20 pixel black sub image with a white 9 pixel radius circle in the middle is the best description of the ball shape.

### 2.2.2 Performing Spatial Linear Filtering

For each pixel location in the pre-processed frame, I place the center pixel of the ball kernel over so that the two pixels are overlapped. As a response score for that pixel, I calculate the sum of the products of the template pixels and the corresponding pixel of the pre-processed frame. The result of this step gets normalized to keep the pixel range between 0 and 1. The underlying principle behind this technique is the pixel and its neighbors will have a high response score if they look similar to the ball kernel [5]. As shown in Figure 2, the ball pixels response to a high score since the warm color presents a higher value in the Jet color map.

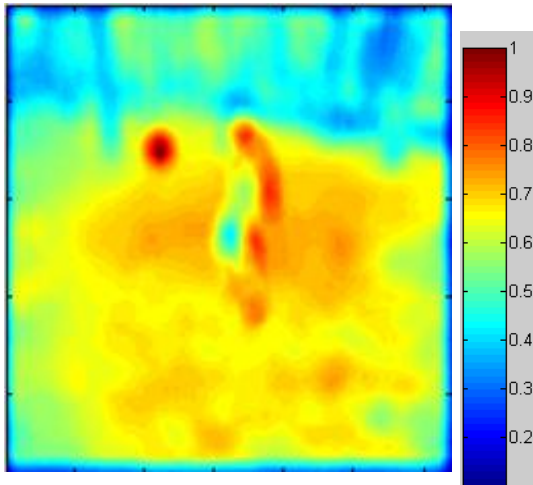


Figure 2. Responding frame after template matching operation displayed in Jet color map

## 2.3 Connected Component Labeling

The high responding scores can be analyzed in the intensity histogram as shown in Figure 3. The highest scores are at the far right-hand-side of the intensity (or horizontal) axis. I would like to segment out only a fixed number of high response pixels. This number of pixels should be close to the total number of pixels in a ball (or the area of the ball in pixels). By thresholding the histogram on the highest pixels, I have a few groups of candidate ball pixels as a binary image.

Connected components labeling scans the image and group its pixels into components based on pixel 8-connectivity [3]. After completing the scan, each group of pixels is given a unique label. The highest responding pixel groups in the histogram are labeled and shown in Figure 4.

Having several groups of ball candidate object shows some of the challenges this paper has to face.

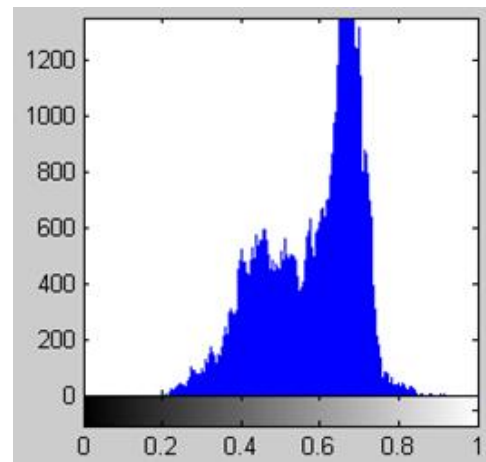


Figure 3. The histogram of the responding frame, the vertical axis is the frequency and the horizontal axis is the intensity value

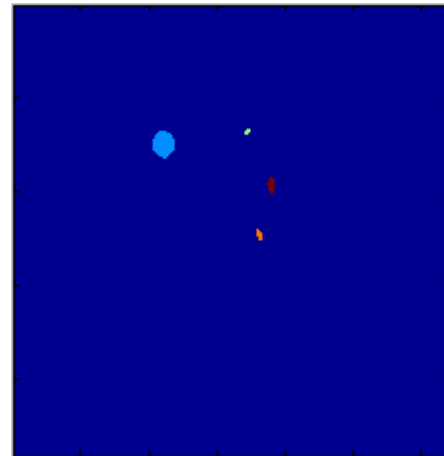


Figure 4. Group of candidate ball objects after connected component labeling

## 2.4 Region Segmenting

Region Segmenting has to make an important decision about which candidate object is the soccer ball. I apply different approaches for the first frame and for the later frames.

For the first frame, I calculate the area of each candidate object. The object which has the largest area will consider the ball object. As little as the information we get from the first frame, the method prove to the most effective. Then, I calculate the centroid, or the center of mass, of the ball region. This centroid location will be saved as a pair of x, and y coordinates.

After the first frame, a nearness constraint is considered. This constraint based on the fact that soccer ball only moves in a certain range from frame to frame. Besides acquiring the largest region, the method also take into account the distance from the centroid of a candidate object to the centroid of the detected object in the previous frame. Throughout the video sequence, there are some situations where the background region is larger

than the ball region. In those cases, the nearness constraint is critical to the algorithm to correctly acquire the ball region. For each frame, the centroid location is saved so that it can be used to compute the distance to each candidate object in the next frame.

Figure 5 demonstrates how the algorithm makes decision for the right ball object. The Euclidean distance to the last ball location is shown at the bottom of the figure.

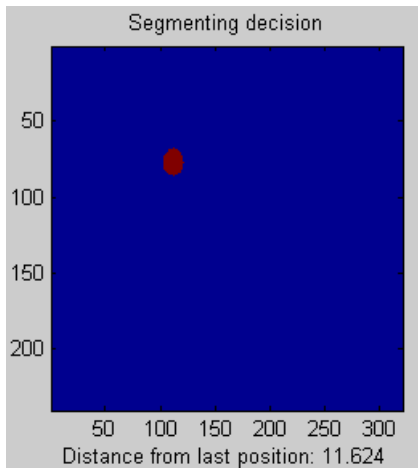


Figure 5. Segmenting Decision with nearness constraint

### 2.5 Contour Tracking

After the ball object location are identify, a re-sampling ball pixel process is consider. The centroid location from the previous step did not fully reflect the precise location of the ball when there are more than one candidate objects. This challenge is due to the histogram thresholding method as it acquires the number of high response pixels that equals to the number of pixels of the ball kernel. Consequently, if there are more than one candidate object, the true ball region will not get as many pixels.

Based on the centroid location of the ball, the processing frame is cropped into a smaller image which has the centroid location as the center. The ball region is re-acquired using the histogram thresholding on that cropped image. Now I have the full size of the ball region. Then, I re-calculate the centroid of this new region. This centroid will be more accurate to the true ball location. Although this process takes as much time as acquiring the ball region in the first time, the algorithm become more robust to image with a lot of background objects that are similar to the ball image.

The outline of the tracked soccer is draw as a circle with the same radius as the ball kernel with thickness of 5 pixels. The center of this circle is the new centroid location. Figure 6 is the original frame with the tracking contour.



Figure 6. Original frame with tracking contour

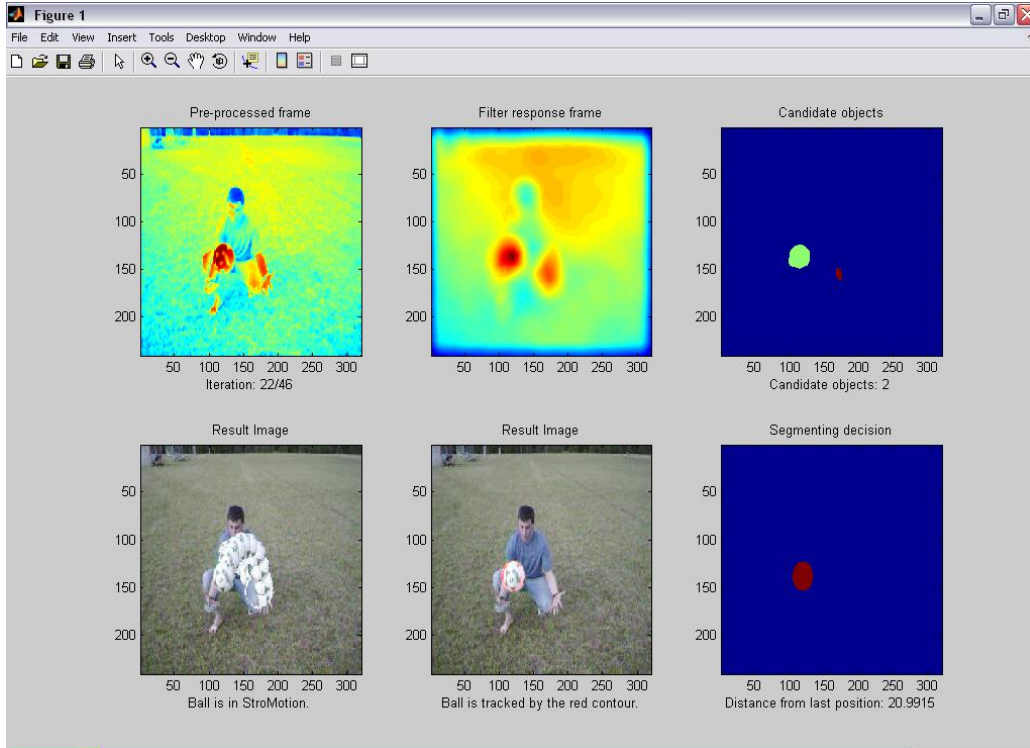
### 2.6 StroMotion Displaying

StroMotion displays static images of the ball in time and space up to the current frame. In order to keep with all the static images, a mask image, called "clipboard", is developed. The pixels from the previous frame which corresponding to ball pixel in the clipboard will be pasted into each frame as the frame is processing. Acquiring the new location of the ball, the algorithm makes sure that the previous ball images are not overlapped the current ball image. For updating the clipboard, each time the ball is tracked, its image, bounded by the contour, is added into the existing clipboard.

Figure 7 is the resulting of StroMotion technique.



Figure 7. The StroMotion image of the 2<sup>nd</sup> frame



**Figure 8. The visualization component of the algorithm at the 22<sup>nd</sup> frame: (1) Pre-processed frame after applying basic image processing techniques. (2) Filter response frame after template matching. (3) Candidate objects after Connected Component Labeling. (4) Segmenting Decision with nearness constraint. (5) Result Image with red tracking contour. (6) Result Image with StroMotion technique.**

### 3. EXPERIMENT

#### 3.1 Experiment properties

Two experiments are carried out, first is the ball juggling by hands and second is the ball juggling by feet. The ball diameters are 32 and 18 pixels for each experiment respectively. A few conditions have to be satisfied including sunny weather, less traffic movements on the road and the camera is hold static. Video sequences are taken using KODAK CX 7200 3MP digital camera. The sequences have 46 and 81 frames respectively. The frame dimension is 320 X 240 pixels.

The experiment took into account some background noise challenges such as white soccer shoes, white stripe on the shirt, player’s arms, and white collar. On the second experiment, I had to add the nearness constraint because the largest area constraint had failed in frame 16 to 26 where the shirt collar actually responded a larger area than the soccer ball.

#### 3.2 Visualization and ground truth collection

A visualization component is implemented to show the users how the algorithm is progressed through each step. I would like to give the user more details of how the algorithm solves this problem. Figure 8 shows the progress of the first experiment.

Ground truth (GT) for this algorithm was painful at first. For each frame from the sequence, I opened the image; located the center of the ball, recorded the x, y coordinate of that location. However, I discovered a ground truth collection method. This method reads in each frame, and displays it. Then, I am only responsible for

clicking at the ball location. As soon as I click on the ball center, the method saves that coordinate into a GT matrix and gets the next frame. The GT matrix is used later to compute the accuracy of the algorithm.

#### 3.3 The results

The evaluation result of the algorithm is done by compute the accuracy of the predicted center of the ball in comparison with the ground truth coordinates. For each pair of points, the Euclidean distance is measured. Then, these distances are compared with a given threshold distance to be classified as correctly detected center. The accuracy is measure by the ratio of correctly detected center and the total number of detection point. The following tables describe how accurate the algorithm in the experiment image sequences. The first column is the threshold distance between the predicted and ground truth. The second column is the corresponding accuracy of the algorithm (1 = 100%).

**Table 1. Evaluation result for the first experiment**

Distance	Accuracy
10	0.8913
9	0.78261
8	0.71739
7	0.58696
6	0.30435
5	0.15217
4	0.1087
3	0.086957
2	0.043478
1	0

**Table 2. Evaluation result for the second experiment**

Distance	Accuracy
7	1
6	1
5	1
4	0.85393
3	0.51685
2	0.13483
1	0.022472

## 4. DISCUSSION AND CONCLUSION

### 4.1 Discussion

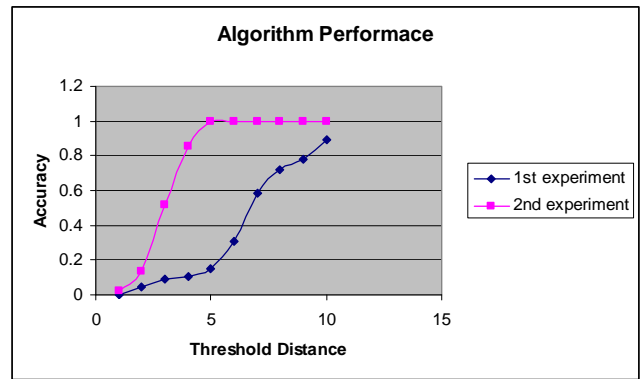
As presented in Table 1, within 10 pixels radius, the 32 pixels wide soccer ball in track 89% of the time. The accuracy is decrease as the precision increase (or smaller threshold distance). Table 2 gives the result of the second experiment. The 20 pixels wide soccer ball is tracked 100% of the time with 5 pixels precision. Figure 9 shows the algorithm performance with different precision scales. The second experiment clearly has much better performance than the first one.

The promising result of the second experiment gives a potential exploration to more precise method in the future. Since the StroMotion is dependent on the tracking result, tracking should be addressed and improved first in order to make StroMotion better.

### 4.2 Conclusion

Soccer ball movements has been tracked and displayed as StroMotion in this paper. Template matching proves to be an effective method to capture the ball image. StroMotion, a decomposition of moving object, is implemented successfully in two short video sequences.

Although the algorithm is able to track the soccer ball, the precision remains a challenge. At high precision, such as threshold distance of 3 pixels or smaller, the algorithm had a low accuracy rate. Moreover, the video sequences are taken relatively close to the ball, so the ball appears in the video as one of the major moving object. In real application where the camera is located further away, and acquires many more background objects, this algorithm might have some difficulties.



**Figure 9. The Algorithm Performance in 2 experiments**

For future direction, more sequences can be analyzed to help improving the algorithm accuracy and robustness. Furthermore, there are many tracking algorithms [6] in the literature that can be used to solve similar problems. My method, however, has combined what I learned throughout the course of Computer Vision to solve an interesting problem of my favorite sport.

## 5. REFERENCES

- [1] Cole, Luke, Austin, David, and Cole, Lance. *Visual Object Recognition using Template Matching*. Australian Robotics & Automation Association Conference, 2004.
- [2] Dartfish Ltd. *Discovering DartCaster- SimulCam & StroMotion in Action*. Technical Report. 2002  
<[www.dartfish.com/data/document/document/506.pdf&name=StroMotion%20&%20SimulCam](http://www.dartfish.com/data/document/document/506.pdf&name=StroMotion%20&%20SimulCam)>
- [3] Fisher, R., Perkins, S., Walker, A., and Wolfart, E., *Connected Component Labeling*, 2003.  
<<http://homepages.inf.ed.ac.uk/rbf/HIPR2/label.htm>>
- [4] Gonzalez, R., Woods, R., and Eddins, S., "Intensity Transformations and Spatial Filtering." *Digital Image Processing using Matlab*. Pearson Prentice Hall: New Jersey 2004.
- [5] Souvenir, Richard, *Computer Vision Applications*, UNC-Charlotte, Charlotte, NC. 13 September, 2007.
- [6] Yilmaz, A., Javed, O., Shah, M., *Object Tracking: A survey*. ACM Computing Survey 38, 4, Article 13, 2006.