Chapter 7

DWT and LDWT for Tracking via
WSD Descriptors\(^1\)

7.1 Introduction

Object tracking is a practice devoted to trace the moving object in the series of frames. It is found that the object tracking is one of the high end applications in the field of CV. The exact detection of moving objects and tracking in sequence of frames is extremely significant to offer the focus of attention and evaluate the motion.

In the previous chapter, the moving regions are extracted through gray optical flow of Horn-Schunck. The edges of the template as well as candidate templates are obtained to estimate the FD. The procedure developed tries to establish the Euclidean distances between the FD of template and the candidate templates to declare the best match to update.

In this chapter, Discrete Wavelet Transforms (DWT) is considered to track the object. In the proposed methodologies, third level DWT or Lifting based DWT (LDWT) of three consecutive frames is computed. Consequently, moving objects are erected by way of Double Change Detection (DCD). Further, WSD

\(^1\)Some parts of the material of this chapter appear in the research paper “An Exploration with Classical DWT and Lifting based DWT for Tracking the Object and employment of Weighted Standard Deviation (WSD) Descriptor for Template Updating”, Eighth International Conference on Image and Signal Processing (ICISP-2014), Proceedings Published by UVCE, Bangalore, India, ELSEVIER, DBLP, indexing, pp. 60-64.
Descriptors for each of the template and candidate templates are determined. In order to update the template, judicious decision for the best match by means of Euclidean distance between the WSD descriptors of the template and the candidate templates is necessary.

The literature supports a few milestone representative approaches like the background subtraction, temporal differencing and optical flow for acquiring the motion regions [107, 108, 109]. The comparative analysis with the Cheng’s work [110] has been carried out to reveal the performance of the proposed tracking criteria.

The extraction of moving regions by way of background subtraction yields better motion blobs but it is affected by the change in illumination. Therefore, the background reference must be updated regularly to overcome the issues [110, 111, 112]. In frame differencing technique, for dynamic background, consecutive frames are used to get the moving regions. Moreover, the moving object information is scanty and requires judicious morphological operations to fill the gaps. Further, the motion regions are extracted using the optical flow approaches such as Horn-Schunck or Lukas-Kanade. These optical flow methods are normally computationally expensive. So, background subtraction, frame differencing and optical flow methods are susceptible to changes in illumination, fake motion and noises etc [113, 114, 115, 116, 117]. On other hand, researchers have suggested different methods to overcome such difficulties in acquiring moving objects [118, 119, 120, 121, 122].

In order to extract the moving regions, the 2-D DWT has been employed which decomposes an image into four sub-band images (LL, LH, HL and HH). The frame difference between the low frequency bands LL of current image and of the previous frame is estimated. Similarly, the difference image between the current frame and the next frame is computed. Both difference images are merged with the help of union operation as a task of Double Change Detection (DCD).

The features of moving regions are structured with the help of WSD descriptors to update the template. The measure of Euclidean distance has been exploited to establish the matching between the WSD descriptors and of the candidate templates. This has explored the disadvantage of high computational cost for decomposition, but it has attempted to address the issues concerned to
noise removal, slow motion discarding and reduction in the computation of post-processing.

Here, the exhaustive experiments with third level decomposition with DCD have confirmed the improvement in the tracking performance. Further, experimental investigation with LDWT has shown 50% reduction in the computational time [113].

The authors Al-Berry et al [123] have contributed the technique based on the proposal of accumulative frame differencing (AFD) employing 2-D DWT. Here, $L-1$ difference images are created by way of subtracting each of the $L$ images of buffer from the reference image. The AFD is figured by carrying out a positive addition for each pixel location in the image frame whenever a difference happens at that pixel location between the reference and the image in the buffer.

The research team, Shyang-Li Chang et al [124], has proposed the system of human body tracking. Here, CCD camera is mounted on a rotary platform and position information of the interested object has been used to control the movement of the platform to lock object around the central area of the frame. In this effort, the image is converted into YIQ color scheme and the resulted image is decomposed with 2 levels. The luminance parts (Y), each of 8 subsequent frames are averaged to form the dynamic background image. Further, difference image is established by subtracting dynamic background frame from the current frame. The spatial information of the moving objects is stored using the connected component. In order to track the object, color and spatial information are used to detect the correct target by template matching.

In the work of Seema Rajput and S. D. Oza [125], image is decomposed by 2 levels. LL sub-band of foreground and also background are normalized. The frame differencing of these normalized images is performed to extract the moving objects. The spatial information is made use to track the moving objects.

In the work of Ganesh Rakate [126], human body tracking with ARM-Linux platform has been proposed. Here, image is decomposed into four sub-bands. Likewise, third level decomposition is performed. The low frequency band of LL-3 of current frame and previous one is estimated. The frame difference between these two images will give incomplete moving objects. Therefore, image restoration is employed to get adequate moving regions.
In the effort of Cheng et al [110], the 2-D DWT was exploited to decompose an image into four sub-band images (LL, LH, HL and HH) to detect and track moving objects. The LL sub-band is further decomposed into second level and finally into third level sub-band (LL-3). The LL-3 of each frames are used for frame differencing with DCD to get the moving objects.

In this work, contemporary literature has dragged the attention to contribute in similar line which is able to bear less computational intensive criterion such as Cheng et al [110]. Further, literature directed to carry out the research on tracking through DWT, LDWT for acquiring moving objects and WSD for template updating.

In Section 7.2, DWT and LDWT models have been described and Section 7.3 exhibits the Weighted Standard Deviation (WSD) Descriptor Model. Proposed algorithm is described in Section 7.4 and experimental results are detailed in Section 7.5. Conclusion is given in Section 7.6.

### 7.2 DWT and LDWT Models

The mathematical basis of Discrete Wavelet Transforms (DWT) and Lifting Discrete Wavelet Transforms (LDWT) is utilized to extract the moving regions. In order to update the template, the feature extraction of moving objects is achieved by means of WSD descriptors and Euclidean distance is exploited to establish the matching.

#### Discrete Wavelet Transforms (DWT)

The DWT has been used since 1985 in various fields of signal processing, seismology and image processing etc. It has superior advantage over the Fourier transforms because of its localization in frequency as well as in spatial domains. For decomposition, Haar wavelet is used because of speedy operation and simplicity. The decomposed four sub-bands are marked with LL, HL, LH and HH. This has been shown in Figure 7.1. Each sub-band consists of various frequency contents. HH band consists of high frequency values and LL band provides low frequency information giving most energy, noise free and useful stuff. In order to extract the moving objects, the DWT of current, previous and
next image are prepared. Difference Image (DI1) between the current and previous image is estimated and current image and next frame are used to compute the Difference Image (DI2) using frame difference technique. This is called double change detection (DCD) as mentioned earlier. Later, two motion regions are merged by way of union operation.

![Figure 7.1: (a) Original image (b) first-level DWT (c) second-level DWT (d) sub-bands of second-level DWT.](image)

**Lifting based Discrete Wavelet Transforms (LDWT)**

The lifting scheme is the emerging strategy [127, 128, 129] with advantages of less computational complexity and simplicity over the conventional DWT. Lifting is employed to propose the second generation wavelets which do not use the same function at different levels. The process of LDWT is supported in three stages such as split, predict and update phases as portrayed in Figure 7.2 and Eqs. (7.1) to (7.5).

![Figure 7.2: Lifting-based Discrete Wavelet Transform implementation (LDWT).](image)

**a) Split Phase:** Here, the original image is separated into two sub-sequences yielding odd indexed samples \((X_o)\) and even indexed samples \((X_e)\) as given in

\[
\begin{align*}
    \text{Odd} & \quad \text{Split} \quad \text{Predict} \quad \text{Update} \\
    \text{Even} & \quad \text{Smooth} \quad S_j \quad \text{Detail} \quad D_j
\end{align*}
\]
Eqs. (7.1) and (7.2). This sub-sampling is termed as lazy wavelet transforms.

\[
X_o \quad d_i \leftarrow X_{2i+1} \quad (7.1)
\]

\[
X_e \quad S_i \leftarrow X_{2i} \quad (7.2)
\]

(b) **Prediction Phase or Dual Lifting:** This stage will produce the detailed coefficients \(d_j\) and neighboring even samples are added, multiplied by prediction factor of half. The end result is added to the odd sample to produce the high pass coefficient as given in Eq. (7.3)

\[
HP(x_{2i-1}) = x_{2i-1} - \frac{1}{2} (x_{2i} + x_{2i+1}) \quad (7.3)
\]

(c) **Update Phase:** Now, low pass filtered output is generated by multiplying a factor of one fourth to the summation of the detailed coefficients produced by predict stage and the result is added with the even sample to produce the low pass filtered coefficient or smooth value \(S_j\) as per the Eqs. (7.4) and (7.5)

\[
S_i = S_i + U(d_j) \quad (7.4)
\]

\[
LP(x_{2i}) = x_{2i} + \frac{1}{4} [HP(x_{2i-1}) + HP(x_{2i+1}) + 2] \quad (7.5)
\]

### 7.3 Weighted Standard Deviation (WSD) Descriptors Model

WSD is a texture descriptor of the gray scale image. Here, gray scale image is decomposed into \(L\) levels using Haar wavelet. For each level of decomposition, there are three detailed bands (LH, HL and HH). Each approximate image is further decomposed to get three detailed images. The standard deviation of each detailed bands are computed. Finally, standard deviation of approximate image and its mean are estimated [130]. All the calculated coefficients are used to construct the WSD descriptors as per the Eq. (7.6).

\[
f = \{ \sigma_1^{LH}, \sigma_1^{HL}, \sigma_1^{HH}, \frac{1}{2} \sigma_2^{LH}, \frac{1}{2} \sigma_2^{HL}, \frac{1}{2} \sigma_2^{HH}, \ldots, \frac{1}{2^{L-1}} \sigma_L^{LH}, \frac{1}{2^{L-1}} \sigma_L^{HL}, \frac{1}{2^{L-1}} \sigma_L^{HH}, \frac{1}{2^{L-1}} \sigma^A, \mu^A \} \quad (7.6)
\]
where $\sigma_{i}^{MM}$ denotes as standard deviation of the MM detailed image such as LH, HL or HH of $i$th level decomposition. $\mu^{A}$ is mean and $\sigma^{A}$ is the standard deviation of the approximate image. The decomposed sub-band image at level $i$ has been weighted by the factor of $1/2^{i-1}$. The standard deviation at each level of sub-band image provides quantification of amount of detail. Higher weights are attached to lower levels of decomposition because more texture information is contained. The intensity information of the scene is available in the mean of approximate image. The size of the feature vector is $3L+2$ for decomposition of $L$-level.

### 7.4 Proposed Algorithm

**Third Level DWT for Tracking via WSD**

The proposed strategy of third level DWT for extracting moving objects with the help of DCD through the frame differencing for object tracking and updating through the WSD is depicted in Figure 7.3.
Figure 7.3: Flow chart for tracking via WSD.

**LDWT for Tracking via WSD**

Steps for this are given below:

1. Convert the video into frames.
2. Read the previous frame $f_{n-1}$, current frame $f_n$ and next frame $f_{n+1}$.
3. Estimate the LDWT of $f_{n-1}, f_n$ and $f_{n+1}$.
4. Exercise the absolute difference (DI1) between $LL_{n-1}$, $LL_n$, and DI2 between $LL_n$ and $LL_{n+1}$.

5. Morphological operations used to get binary masks as $B_1$ and $B_2$ respectively.

6. Merge both $B_1$ and $B_2$ employing union operation.

7. Connected components theory is utilized to structure the knowledge base with spatial information of all moving objects.

8. Compute the WSD descriptors for template and the candidate templates.

9. Establish the matching between WSD descriptors of the template and the candidate templates by Euclidean distance.

10. Minimum matched moving object is replaced as new template and it’s related spatial information passed on to put the bounding box on the object.

11. If video is not ended go to Steps 2 to 10.

7.5 Experimental Analysis

Exhaustive experiments of the proposed approaches have been conducted on PETS2001(3), PETS2000 and PETS2001(2) to reveal the outcome. The performance and quality of tracking the object is exhibited through the qualitative and quantitative analysis.

Feature extraction technique of WSD descriptor is considered for all the experiments on different dataset. Results of experiments are portrayed in the Table 7.1 and Figure 7.4. The suggested methodologies of DWT or LDWT with WSD as feature extraction has been compared with the work of Cheng et al [110].
# Third Level DWT (Cheng’s Work)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Threshold</th>
<th>Dist. Meas.</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>P</th>
<th>R</th>
<th>FAR</th>
<th>CR</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20</td>
<td>CSD</td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>98</td>
<td>100</td>
<td>2</td>
<td>98</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>CSD</td>
<td>60</td>
<td>8</td>
<td>2</td>
<td>88</td>
<td>97</td>
<td>12</td>
<td>89</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>C</td>
<td>30</td>
<td>CSD</td>
<td>83</td>
<td>11</td>
<td>6</td>
<td>88</td>
<td>93</td>
<td>12</td>
<td>89</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Third Level DWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LDWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
</tbody>
</table>

A – PETS2001(3); B – PETS2000; C – PETS2001(2)

Table 7.1: Results of DWT, LDWT with WSD.

![Combined plot of F with proposed strategies](image)

Figure 7.4: Combined plot of F with proposed strategies.
Comparative analysis of proposed third level DWT with Cheng’s work shows no significant improvement in the tracking performance for PETS2001(3) but there is improvement in tracking performance with PETS2000 dataset. A similar comparative analysis of the LDWT shows considerable improvement in the tracking performance with different dataset. The tracked frames for third level DWT with WSD have been presented for PETS2001(3), PETS2000 and PETS2001(2) dataset in Figures 7.5 to 7.6.

Figure 7.5: Tracking PETS2001(3).
Figure 7.6: Tracking PETS2000.

Figure 7.7: Tracking PETS2001(2).
7.6 Conclusion

The outdoor surveillance benchmark dataset have been employed to carry out the experimentation. The wavelets of third level DWT and LDWT have been employed to reduce the noise. In order to extract the moving regions, the frame difference technique of DCD is employed with suitably opted threshold values for each proposed strategy. Subsequently, spatial information of moving objects are stored.

When compared with Cheng’s work, it is found that the tracking performance of LDWT with WSD is superior for different dataset than the performance obtained with third level DWT. The computational aspects will reduce to 50% with LDWT compared with classical DWT approaches. Hence, both the proposed algorithms have enhanced the robustness against the change in the background illumination, movements and noise etc.