A novel rotation/scale invariant template matching algorithm using weighted adaptive lifting scheme transform

M. Amiri, H.R. Rabiee *

Digital Media Laboratory, Advanced Information and Communication Technology Center, Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 6 December 2008
Received in revised form
18 September 2009
Accepted 22 December 2009

Keywords:
Adaptive lifting scheme
Object detection
Rotation/scale invariant template matching

1. Introduction

The problem of detecting rigid objects in an image by template matching is considered in this paper. Template matching in grayscale images is one of the main building blocks in various image processing and machine vision systems. The general approach for solving feature based template matching problems is as follows. First, some reliable features from the object of interest are extracted and stored as an object model. Then, the same feature extraction method is applied to a given test image. These features are compared with the stored object features for finding the best match.

In the past decade, wavelet-based methods for detection and enhancement tasks have received considerable attention within the image understanding community. In particular, they have shown to be effective in locating objects of interest in various kinds of digital images such as mammograms, SAR, and hyperspectral images [1–3]. The discrete wavelet transform (DWT) has properties that make it an ideal transform for image understanding applications. These properties include efficient representation of abrupt changes and precise spatial information, existence of the fast processing algorithms, ability to adapt to high background noise, and robustness against uncertainty and changing local image statistics.

Inherent ability for the efficient approximation of smooth signals is one of the prominent reasons for the success of wavelets in various image processing applications. However, real-world signals are not necessarily smooth, as required by the classical wavelet transform algorithms. Therefore, adaptive approaches are required to overcome discontinuities encountered in the real-world signals.

For the smooth input signals, most of the coefficients in the high-pass component of the wavelet transform are equal to zero. Therefore, one may conclude that the remaining coefficients in the high-pass channel, which have large magnitude and correspond to the details and edges, may be considered as the features of the input signal. This fact in the context of “wavelet transform modulus maxima representation” was first presented by Mallat and Zhong [4] and has been widely used by various researchers in detection applications [5–7,1].

In this paper, we present a novel template matching algorithm based on adaptive wavelet transform. In our algorithm, for a given object of interest, we design an adaptive lifted wavelet transform such that the desired coefficients in the high-pass component of the non-adaptive transform, vanish in the high-pass component of the adaptive transform. Moreover, the vanishing percentage of each coefficient is proportional to its value, such that the larger coefficients are vanished more often than the smaller coefficients. Both of the non-adaptive and adaptive transforms, are applied to a given test image. An algorithm is presented for detecting the object of interest by comparing the high-pass component coefficients of the non-adaptive and the adaptive wavelet transforms. In addition, the proposed detection algorithm is combined with the proper log-polar mapping model in the parametric template space to attain rotation/scale invariance property. Finally, we have verified the properties of our proposed algorithm with experimental results.

© 2009 Elsevier Ltd. All rights reserved.
2. Adaptive lifting related work

Many adaptive approaches for object feature extraction have been developed by various researchers, in recent years. The best basis algorithm [8] is a good example of a common adaptive approach where one chooses a wavelet basis which depends on the input signal. In this type of approach, the basis is selected by minimizing a cost function such as entropy in the wavelet packet transform tree. However, this is a global adaptive approach and the chosen basis is fixed for the entire block of data.

Lifting scheme, presented by Sweldens [9], provides a good structure for creating adaptive wavelet transforms. Lifting scheme presents a means for decomposing wavelet transform into predict and update stages. One may adapt prediction or update stage filters to local signal properties and build the desired adaptive wavelet transform.

Claypoole et al. [10,11] proposed an adaptive lifting scheme for image compression and denoising applications. They switch between different linear predictors at the predict stage; higher order predictors where the image is locally smooth and lower order predictors near edges to avoid prediction across discontinuities. However, one would have to keep track of the chosen filters at each sample, to guarantee perfect reconstruction at the synthesis stage. In addition, they should apply the update stage first, in order to avoid sending information on a chosen predictor to the reconstructor.

An update first strategy was also utilized by Piella and Heijmans [12]. Unlike Claypoole et al. they choose a fixed predictor and move adaptiveness into the update stage in such a way that no bookkeeping is required. They use a binary decision map (BDM) to find where the local gradient is over/under a threshold. Then, they select the update filter based on the BDM output in order to regulate the homogeneous areas and ignore discontinuities. It is important to note that the approaches presented in [11,12], do not fit within the classical lifting scheme as the prediction step does require input from both channels.

Trapp and Liu [13] also adapt the predict stage. They try to minimize the predicted detail signal by designing a data-dependent prediction filter. They present two different approaches. The first one is the global adaptivity and its goal is to minimize norm of the entire detail signal. In the second approach, the coefficients of the prediction filter vary over time, based on a local optimization criterion. However, there is not any special consideration for discontinuities in contrast to the algorithms in [10,12]. In addition, similar approaches have been earlier proposed by other researchers like Gerek and Çetin [14], Bouglouris et al. [15], and Chan and Zhou [16].

The discontinuities, representing edges, play an important role as features for tracking the objects of interest. Although, our proposed method for designing the adaptive transform fits within the classical lifting scheme structures, but it offers special consideration for discontinuities of the signals. Moreover, in all of the previous approaches, the main purpose for design of adaptive wavelets is to highlight the signal features in the transform domain and to minimize norm of the detail signal. On the contrary, the main purpose of our designed adaptive wavelet is to vanish the signal features in the transform domain, which provides the essence of our new detection algorithm [17,18].

3. Rotation/scale invariant template matching related work

In conventional template matching algorithms, given a template image (reference block) and an input test image, the matching algorithm finds a block in the test image that most closely matches the template image in terms of measures of closeness, such as Euclidean distance or normalized cross correlation. The classic pixel-by-pixel block matching algorithm (BMA) is not robust to various image irregularities including illumination fluctuation, noisy environment, rotation, and scaling. When we try to search an object with unknown orientation and scaling, the straightforward way to use BMA in this case, is to rotate and scale the reference block in every possible orientation and scaling. This “brute force” solution to the problem has extremely high computation complexity and is impractical when arbitrary rotation and scaling is allowed.

Within the past few decades, various feature and model based methods have been explored by researchers to bring robustness properties to the problem of template matching. In this section, we provide a brief survey of the prominent related works.

Geometric hashing [19,20], generalized Hough transform (GHT) [21,22], graph matching [23,24], scale-invariant feature transform (SIFT) [25,26], curvature scale space (CSS) [27,28], Harris–Laplacian interest point detector [29], and speeded up robust features (SURF) [30] are among a group of approaches that achieve rotation and scale invariant template matching using a set of robust features and key-points. Geometric hashing and graph matching based methods use high-level data, while GHT, SIFT, SURF, CSS and Harris corner detection based methods use low-level data. Methods of this group do not use grayscale pixel-level data in a direct way and rely on the limited number of feature points which are extracted from those data. Therefore, their effectiveness will be affected by the matching capabilities of the extracted features. For example, when the features are based on the object’s contour, the benefit of any texture which may be present in the grayscale pixels of the object’s surface will be ignored.

Another group of approaches achieve robust template matching by limiting the field of work to the binary images. Zernike moment based method in [31] and the algorithm based on the extraction of topological object characteristics from binary templates in [32] are among the distinguished examples for this group of methods. When working on a grayscale image, they first use a thresholding algorithm to convert it to the binary image. Then they extract rotation and scale invariant features for the connected components of the binary image that do not belong to the background. For template matching, features of these connected components will be compared with the features of the binary template. Hu's seven moments [33] and Zernike moments [34,35] are among the most commonly used features in this group. Some of the recent features include fractal geometry based feature in [36], polygonal approximation based feature in [37], 1D projection based feature in [38], and complex moments for symmetric objects in [39]. However, all these features are only rotation invariant. To make them scale invariant, each component of the binary image is isolated and the component area is normalized to one. The requirement to isolate individual shapes and connected components of the grayscale image, is the main disadvantage of this group of methods.

There is a group of approaches that do not require isolation of the connected components and use different forms of projective mappings to extract robust features. Ring projection in [40–42] is a mapping used to project the 2D image pixels under concentric circles into a 1D signal as a function of radius. This way, the resulted features will be rotation invariant and the computation
complexity of matching process will be reduced. Similarly, radial projection is a mapping used to project the 2D image pixels under radial lines into a 1D signal as a function of radial line angle, to achieve scale invariant features. A cascade of ring projection and radial projection based filters used in [43] to achieve both rotation and scale invariant properties. However, log-polar mapping model, which is utilized in our paper, inherently includes both ring and radial projections [44]. Another prominent work of this model, which is utilized in our paper, inherently includes both and scale invariant properties. However, log-polar mapping achieve scale invariant features. A cascade of ring projection and radial lines into a 1D signal as a function of radial line angle, to projection is a mapping used to project the 2D image pixels under complex of matching process will be reduced. Similarly, radial projection is a mapping used to project the 2D image pixels under radial lines into a 1D signal as a function of radial line angle, to achieve scale invariant features. A cascade of ring projection and radial projection based filters used in [43] to achieve both rotation and scale invariant properties. However, log-polar mapping model, which is utilized in our paper, inherently includes both ring and radial projections [44]. Another prominent work of this group utilize gradient information in the form of orientation codes in [45]. Orientation code based features are robust in cluttered environments and perform well in the cases of illumination fluctuations resulting from shadowing or highlighting. For rotation invariant matching in [45], they first construct the histograms of orientation codes for the template and a subimage of the same size. Then, compute the similarity between the two histograms for all the possible orientations by shifting the subimage histogram bins relative to the template histogram. This process is to approximate the rotation angle of the template in the subimage. On the other hand, orientation code based features are not scale invariant in nature.

In 1991, Ullman and Basri, show that the variety of views depicting the same object under different transformations can often be expressed as the linear combinations of a small number of views [46]. To use this idea for the matching process, one may consider the template as a linear combination of models and determine the coefficients of this linear combination to minimize the difference between the subimage and the template. They show that how a small number of feature-based corresponding points can be used to determine the linear parameters of the model. The linear combination of features and alternative methods of determining the coefficients for minimizing the difference between the subimage and the template, have been the basis for several matching algorithms in the past decade [47,48]. On the other hand, when some prior information of the template's geometrical shape is available, probabilistic models based on a parametric deformable template can be used. For example, Yuille et al. [49] show how one can draw eye and mouth templates using circles and parabolic curves. The parameters which control the shape of a template are the center and the radius of the circle, and the characteristic parameters of the parabola. They were able to accurately locate eyes and mouths in real images, when the initial positions of the templates are close enough to the desired objects. Also, in [50], an edge-based parametric model is constructed, and matching between the deformable template model and a reference image is done by an iterative maximal likelihood estimation based on Bayesian optimization. A deformable template is able to "deform" itself to fit the data, by transformations that are possibly more complex than translation, rotation, and scaling. But, the applicability of parametric deformable model is limited because the shapes under investigation have to be well defined so that they can be represented by a set of curves with preferably a small number of parameters.

In this paper, for the purpose of rotation and scale invariant matching, we employ log-polar mapping and parametric template space [51] along with our proposed method of adaptive transform design method. Using proper log-polar mapping model will make us free from cascaded ring and radial projection filters. Also, parametric template based approach will not require feature-based corresponding points and in contrast to the parametric model methods, allows estimation of the model parameters by a direct linear calculation rather than an iterative calculation.

It is worth to mention that, we have excluded the survey of those methods that achieve fast template matching using various rejection related schemes. A good survey of such methods can be found in [52].

4. The proposed detection algorithm

The block diagram of our novel object detection system is shown in Fig. 1 (offline step) and Fig. 2 (online step). In this section, we first describe the main building blocks of each block diagram, individually. Then the role of each building block in the whole detection system will be described.

4.1. Problem definition

Consider a $m \times n$ reference template $g$ and an $M \times N$ input test image $t$. The problem of template matching can be stated as follows: To find the rotation angle $\theta$, the scale factor $\alpha$ and

![Fig. 1. Block diagram of our proposed object detection system (offline step).](image-url)
location \((u, v)\) in the image \(t\) that minimizes the objective function,
\[
\arg\max_{u, v, \theta} f(u, v, \sigma(\theta \odot t))
\]
(1)

where \(\odot\) denotes the rotation operator and \(\sigma\) denotes the scale operator. The function \(f\) is a measure of similarity between the rotated and scaled template \(g' = \sigma(\theta \odot t)\) and the image patch \(t'\) from image \(t\) centered at \((u, v)\). Different matching methods take different forms of function \(f\). As a classic example, \(f\) may represent the normalized cross correlation.

### 4.2. Designing the adaptive wavelet transform

Designing an adaptive wavelet transform for a given object of interest based on the lifting scheme structure is the main building block of our detection system. First the classical wavelet transform is applied to the given template of the object of interest and large values in the high-pass component of this non-adaptive wavelet transform are considered to be the object features. The main idea of the proposed detection algorithm is the design of an adaptive transform based on these features. Therefore, our new adaptive wavelet transform is designed such that the desired large coefficients in the high-pass component of the non-adaptive transform vanish in the high-pass component of the adaptive transform.

The process of vanishing object features in the transform domain will help us in two ways. First, it will assist us to construct a system of linear equations for designing the desired adaptive wavelet transform as described in Section 4.2.2. Second, it will provide a basis for comparing the outcome of the adaptive and non-adaptive filters which is the essence of similarity measurement in our detection system. In the online step, we apply both non-adaptive and adaptive transforms to a given test subimage. Then, the high-pass component coefficients of the non-adaptive and the adaptive wavelet transforms are compared for detecting the object of interest.

In the following subsections, we first introduce the concept of dual lifting step in the lifting scheme. Then, it will be shown that how we could design the desired adaptive wavelet transform for 1D signals using the lifting scheme structure. This is followed by introduction of the detection algorithm for finding the 1D objects. In addition, we extend the algorithm to the 2D case for detecting an object of interest in a given test image. Finally, we provide a discussion on selection of the required parameters in our algorithm and its efficiency. Moreover, a weighted adaptive version of the proposed method will be presented that enhances the detection power of our algorithm [53].

#### 4.2.1. The dual lifting step

The fast lifted wavelet transform block diagram using a dual lifting step [54] is shown in Fig. 3. Here, \(\hat{h}^{\text{old}}\) and \(\hat{g}^{\text{old}}\) are the low-pass and high-pass analysis filters of the non-adaptive wavelet transform that are applied to the input signal \(x\) respectively. The prediction filter \(\tilde{f}\) is applied to the low-pass component \(\lambda\) and the output \(\gamma^{\text{old}}\) is subtracted from the old high-pass component, \(\gamma^{\text{old}}\), in order to produce the new high-pass component \(\gamma\) as follows:

\[
\omega = (x \ast \hat{h}^{\text{old}}) \ast \tilde{f} \tag{2}
\]

\[
\gamma^{\text{old}} = x \ast \hat{g}^{\text{old}} \tag{3}
\]

\[
\gamma = \gamma^{\text{old}} - \omega \tag{4}
\]

where \(\ast\) denotes the convolution operator.

#### 4.2.2. The prediction filter in 1D case

In this subsection, we show how to find the coefficients of the prediction filter \(\tilde{f}\), such that the coefficients of the non-adaptive wavelet transform’s high-pass component, vanish in the high-pass component of the adaptive lifted wavelet transform. Let \(s\) be the signal of interest. Applying the non-adaptive wavelet transform to this signal will produce the following low-pass \((\lambda)\) and
high-pass ($\gamma^{old}$) components.

$$\lambda = s * h^{old} \Rightarrow \tilde{\lambda}_k = \sum_j S_{h^{old}(k+1-j)}$$

$$\gamma^{old} = s * g^{old} \Rightarrow \gamma^{old}_k = \sum_j S_{g^{old}(k+1-j)}$$

where $k$ is the index of elements belonging to low-pass and high-pass components and $j$ is the index of elements belonging to the source signal $s$.

Given the prediction filter $\tilde{\gamma}$, the high-pass component of the adaptive lifted wavelet transform ($\gamma$) is obtained as follows:

$$\gamma^\prime = \frac{\gamma^{old}}{\tilde{\gamma}}$$

where $\gamma$ is obtained from the dual lifting step structure of Fig. 3 will form our adaptive filter. This new transform could be used in the following algorithm for detecting the 1D signal of interest, in a given test signal.

1. The signal of interest $s$ and the test signal $x$ are assumed to be the input arguments.
2. Select a non-adaptive wavelet transform, and values of the parameters $p$ and $q\ll p-1$.
3. Find the desired prediction filter $\tilde{\gamma}$, as described in Section 4.2.2.
4. Apply the non-adaptive and the adaptive lifted wavelet transforms to the test signal $x$ and find the high-pass components $\gamma^{old}$ and $\gamma$.
5. Construct an empty vector $D$ with the same length as $\gamma^{old}$ and $\gamma$.
6. Compare each coefficient of $\gamma^{old}$ with the corresponding coefficient in $\gamma$ and if it is decreased, find the vanishing percentage (VP), and save it in vector $D$.

$$D_k = \begin{cases} 0 & |\gamma^{old}_k| \leq |\gamma^{old}_k| \\ 100 - (100|\gamma^{old}_k/\gamma^{old}_k|) & |\gamma^{old}_k| < |\gamma^{old}_k| \end{cases} \forall k$$

7. Sweep vector $D$ with a window of the same length as signal $s$, and find sum of the VPs for each windowed location, $\tau$. The location of the maximum value for this sum, could be considered as the location of the signal $s$, in the test signal $x$.

$$\tau_k = \sum_{i=k-L}^{k+L} D_i \forall k$$

where $L$ is half of total support width of reference signal $s$ and $k$ is the index of elements belonging to the high-pass component.

4.2.4. The detection algorithm, 2D case, basic form

The presented detection algorithm for 1D signals, could be expanded to the 2D case for detecting an object of interest in a given test image. We may consider the 2D object as a set of separable 1D signals corresponding to rows and columns of the 2D object. The algorithm for 2D case is as follows.

1. Choose a reference block $O_{m \times n}$, which encompasses the object of interest and test image $T_{M \times N}$ as the input arguments.
2. Consider row $i_k$ of the object $O$ as the 1D “signal of interest” and find prediction filter $\tilde{\gamma}_k$ as described in Section 4.2.2. Repeat this for $i_k = 1, \ldots , n$.
3. Consider column $j_k$ of the object $O$ as the 1D “signal of interest” and find prediction filter $\tilde{\gamma}_k$ as described in Section 4.2.2. Repeat this for $j_k = 1, \ldots , m$.
4. Sweep test image $T$ with a 2D window of the same size as object $O$.
5. Apply non-adaptive and adaptive lifted wavelet transforms to the rows and columns of the windowed image. Compare corresponding coefficients similar to the 1D case and find sum of the VPs. The location of the maximum value for this sum, could be considered as the location of the reference block $O$ in the test image $T$.

4.2.5. The weighted adaptation in basic form

Our experimental results have shown that our new algorithm is an effective detection algorithm [18]. However, in presence of noise, when the image quality is low (i.e. low signal to noise ratio), the performance of the algorithm would deteriorate. Therefore, a modified version of the prediction filter could help us to boost the detection power in presence of noise by introducing a new weighting parameter in the adaptation process.
If we represent each matrix in Eq. (15) by a single letter, we obtain
\[ A\mathbf{T} = \Gamma \]
(18)
where \( A \) is an \((v+1)\)-by-\( p \) matrix, \( \mathbf{T} \) is a column vector with \( p \) entries and \( \Gamma \) is a column vector with \((v+1)\) entries.

We desire the vanishing percentage of each coefficient to be proportional to its value, in a way that the larger coefficients vanish more than the smaller ones. Therefore, a weight vector, \( \mathbf{w} \), is added to both sides of Eq. (18), resulting in the following weighted OLSE:
\[ (\mathbf{w} \cdot \times A)\mathbf{T} = \mathbf{w} \cdot \times \Gamma \]
(19)
in which
\[ \mathbf{w} = [w_1, w_2, \ldots, w_v, 1]^T \]
(20)
where \( \cdot \times \) denotes element-by-element product and
\[ w_i = 1 + B \left( \frac{\gamma_i - \gamma_k}{\gamma_k - \gamma_k} \right) \quad i = 1, \ldots, v \]
(21)
where \( \gamma_k \) and \( \gamma_i \) are the largest and the smallest among the selected coefficients, respectively. Therefore, according to Eq. (21), weight values will be between one and \((B+1)\). Parameter \( B \) is a constant vanishing booster coefficient and its value may vary from one to infinity. The significance of the parameter \( B \) is explained in the following section. It is important to note that the Gauss–Newton method may be used to solve Eq. (19) in order to obtain the coefficients of the prediction filter \( \hat{\mathbf{f}} \).

4.2.6. Prediction filter design discussion
Finding the prediction filter for each row and column of the reference block could be a time consuming task. But in many applications, like image retrieval, we only need to compute the prediction filters once, and use the same filters for detecting object of interest in any chosen test image from the database.

Moreover, due to the following reasons, noise or slight deformations in the object of interest, would not have considerable impact on the resulted VPs.

- Most of the large values in the high-pass component remain among large values in the noisy signals as well.
- Both the non-adaptive and the adaptive transforms are applied to the same noisy signal; therefore the vanishing percentage values will not experience a considerable change.

There is a trade-off in choosing the value of parameter \( B \). Greater values for \( B \) will result in greater vanishing percentage for larger old high-pass component coefficients, and as a result, the detection algorithm will be less sensitive to the noise, because the large high-pass component coefficients represent edges where the noise has less impact on their values. On the other hand, greater values for \( B \) will make it difficult to detect objects that have a large number of small edges.

We conclude that in the noisy images, one may choose larger values for the parameter \( B \) to achieve better detection results. Moreover, when a blurred version of the object is expected in the test image, one may choose smaller values for the parameter \( B \). Albeit, even when parameter \( B \) is one, the weight vector will keep the vanishing percentage of each coefficient proportional to its value, leading to better detection results in the weighted form of our algorithm.

4.3. Log-polar mapping model selection
There is a vast amount of literature related to the fields of template matching, image registration and motion estimation, that make use of log-polar mapping based methods. The motivations in this regard, stem from its scale and rotation invariant properties and the existence of a biological foundation.

The main geometric properties and a well-accepted mathematical definition for the projection of the retina onto the human visual cortex are presented by Schwartz [55]. His model in only of theoretical significance and due to singularity of the logarithmic function in the origin (fovea), it cannot directly be used for a computer implementation. To avoid the problem of having infinite mapping in the fovea, different solutions have been proposed. A good study of the existing log-polar mapping models and their properties may be found in [56].

We have summarized the characteristics of elaborated log-polar mapping models in Table 1. It could be seen that only the model of Tistarelli–Sandini [44] presents both rotation and scale invariant properties. Although this model will include blindness in fovea, but the blind area is not extended more than a few pixels and could be ignored. We perform log-polar mapping in software by the lookup table (LUT) method. Every log-polar pixel \((\xi, \eta)\) will be calculated by the normalized weighted sum of a list of Cartesian pixels \((x, y)\). At the offline step we initialize such a list for each log-polar pixel in the LUT. This LUT will only be designed once for a subimage which could enclose the largest scaled form of the desired object. Therefore, it will be used wherever in the object detection system that the log-polar mapping is required.

### 4.4. Parametric template space

We are using parametric template matching method [51] in log-polar domain to estimate rotation angle and scaling factor of the reference template in a given subimage. Suppose we have \( M \) vertex templates in Cartesian domain \((V_{11}, V_{12}, \ldots, V_{1M})\) which are rotated and scaled versions of the reference template \( O_t \). Rotation angle and scaling factor of each vertex template are known parameters \((\theta_1, x_1)\), \((\theta_2, x_2)\), \ldots, \((\theta_M, x_M)\). After log-polar mapping at the offline step (Fig. 1), the vertex templates \((V_{op1}, V_{op2}, \ldots, V_{opM})\) will be used to create our parametric template space. Therefore, every “parametric template” in this space, will be obtained by
\[ t_{op} = \frac{\sum_{i=1}^{M} \omega_i V_{op}}{\sum_{i=1}^{M} \omega_i} \]
(22)

At the online step, we take a subimage \( O_t \) from the given test image and apply log-polar mapping to get \( O_{op} \). Parameters \( \omega_i \) are calculated by maximizing the normalized correlation of \( O_{op} \) and

| Table 1 Characteristics of renowned log-polar mapping models. |
|-----------------|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Model reference | Blindness in fovea | Computation complexity | Rotation invariance | Scaling invariance |
| Sch. [55]       | No             | Infinite        | Yes             | Yes             |
| Jur. [57]       | Yes            | Moderate        | Yes             | No              |
| Tiss. [44]      | Yes            | Low            | Yes             | Yes             |
| WalOBS. [58]    | No             | Very high      | No              | No              |
| PetBR. [59]     | No             | High           | Yes             | No              |
parametric template $\tau_o$ [51] as stated in

\[
\text{Normalized correlation} (O_{\text{lp}}, \tau_o) \rightarrow \max, \left( \sum_{i=1}^{M} \omega_i \right)
\]  

(23)

Eq. (23) is a maximization problem with a constraint. It could be solved by the Lagrange multiplier method, then the parameters $\omega_i$ will be obtained by

\[
\hat{\omega} = \frac{H_{ct}^T G_{\text{pmt}}}{\bar{n} H_{ct}^T G_{\text{pmt}}}
\]

(24)

where

\[
\hat{\omega} = [\omega_1, \ldots, \omega_M]^T, \quad \bar{n} = [1, \ldots, 1]^T
\]

(25)

\[
H_{ct} = \begin{bmatrix}
(V_{\text{lp}_1}, V_{\text{lp}_2}) & \cdots & (V_{\text{lp}_1}, V_{\text{lp}_M}) \\
V_{\text{lp}_1} & \cdots & V_{\text{lp}_M}
\end{bmatrix}
\]

(26)

\[
G_{\text{pmt}} = \begin{bmatrix}
(O_{\text{lp}_1}, V_{\text{lp}_1}) \\
\vdots \\
(O_{\text{lp}_1}, V_{\text{lp}_M})
\end{bmatrix}
\]

(27)

The correlation matrix $H_{ct}$ can be calculated at the offline step as depicted in Fig. 1. Based on the parametric template matching method, after finding parameters $\omega_i$, the rotation angle and scaling factor $\alpha$ of the reference template $O_t$ in the given subimage $O_t$ are estimated from rotation angle and scaling factor of vertex templates by the following equations, respectively

\[
\theta = \sum_{i=1}^{M} \omega_i \theta_i
\]

(28)

\[
\alpha = \sum_{i=1}^{M} \omega_i \alpha_i
\]

(29)

4.5. The complete object detection system

We are using our core detection algorithm (which was described in Section 4.2) along with building a parametric template space in log-polar domain to construct a rotation and scale invariant object detection system. Proper log-polar mapping, as described in Section 4.3, is used to convert the rotation and scale effects in the Cartesian domain to translations in the log-polar domain. We design the adaptive lifting scheme transform based on rows and columns of the reference template in the log-polar domain, for our core detection algorithm. A parametric template space constructed by several vertex templates in the log-polar domain, as described in Section 4.4, is used to estimate rotation angle and scaling factor of a given subimage. Then, the rows and columns of this subimage in the log-polar domain are rearranged to eliminate the rotation and scale effects. Afterwards, non-adaptive and the designed adaptive wavelet transforms are applied to the fitted subimage and the similarity value is calculated by the comparison of wavelet domain coefficients as described in Section 4.2.5. A detailed step-by-step description of the complete object detection system is presented in this section.

4.5.1. The detection algorithm, offline step

The main building blocks of the offline step are depicted in Fig. 1.

1. The size of an image patch which could enclose the largest scaled form of the desired reference template is selected and a look-up table (LUT) for the described log-polar mapping model is initialized. We will use this LUT wherever log-polar mapping is required. The number of concentric circles $n_c$ and the number of samples in angular direction $n_t$ will represent the log-polar image size as $n_c \times n_t$. These are two main fixed parameters in this step.

2. Log-polar mapping is applied to the given reference template $O_t$ to obtain $O_{\text{lp}}$.

3. The adaptive wavelet transform is designed based on the rows and columns of the reference template in log-polar domain ($O_{\text{lp}}$), as described in Section 4.2. The length of lifting step filter ($p$) and the number of large values in the high-pass component ($\alpha$) are two main fixed parameters in this step.

4. Several vertex templates, $(V_{\text{lp}_1}, V_{\text{lp}_2}, \ldots, V_{\text{lp}_M})$, are created from the given reference template with different rotation angle and scaling factors. The number of vertex templates and their rotation and scale effects are fixed parameters in this step and may vary based on the requirements of the desired application.

5. Log-polar mapping is applied to the vertex templates $(V_{\text{lp}_1}, V_{\text{lp}_2}, \ldots, V_{\text{lp}_M})$ to obtain $(V_{\text{lp}_1}, V_{\text{lp}_2}, \ldots, V_{\text{lp}_M})$.

6. Using Eq. (26), the correlation matrix $H_{ct}$ is calculated for the parametric template space constructed by the vertex templates $(V_{\text{lp}_1}, V_{\text{lp}_2}, \ldots, V_{\text{lp}_M})$.

4.5.2. The detection algorithm, online step

The main building blocks of the online step are depicted in Fig. 2. The process which is shown in this figure, shows how we find the similarity value for any desired location of the given test image.

1. We consider $O_t$ to be the image patch which is taken from the given test image and its center matches with the location $(u, v)$. Log-polar mapping is applied to the subimage $O_t$ to obtain $O_{\text{lp}}$.

2. Using Eq. (27), the correlation matrix $G_{\text{pmt}}$ is calculated between $O_{\text{lp}}$ and the vertex templates $(V_{\text{lp}_1}, V_{\text{lp}_2}, \ldots, V_{\text{lp}_M})$.

3. Calling the correlation matrix $H_{ct}$ from the offline step and having the correlation matrix $G_{\text{pmt}}$, we use Eqs. (24), (28), and (29) to estimate rotation angle and scaling factor of the reference template $O_t$ in the image patch $O_t$.

4. Based on the selected log-polar mapping model, the rotation and scale effects in the Cartesian domain are converted to translations in the log-polar domain. So, we can rearrange the rows and columns of $O_{\text{lp}}$ based on the estimated rotation angle and scaling factor to obtain the fitted subimage $O_{\text{lp}_{\text{fit}}}$.

5. The fitted subimage $O_{\text{lp}_{\text{fit}}}$ will be considered as the input test image for our core detection algorithm based on the designed adaptive lifting scheme transform in the offline step. Both of the non-adaptive wavelet transform and the designed adaptive transform are applied to the fitted subimage and wavelet domain coefficients are compared as described in Section 4.2 to find the similarity value between the reference template $O_t$ and the image patch $O_t$ at location $(u, v)$.

5. Implementation notes and experimental results

5.1. Discussion and implementation notes

The first advantage of using the proper log-polar mapping as a preprocessing module is because the rotation and scale effects in the Cartesian domain are converted to translations in the log-polar domain. Another advantage of this domain for object detection applications comes from its space variant sampling property. It provides means to have high resolution in the areas of interest and also to control the number of information bearing pixels. The log-polar image size is controlled by the number of
concentric circles \( n_h \) and the number of samples in angular direction \( n_u \). The sampling resolution of log-polar mapping is higher around fovea and decreases exponentially as it gets further from that, so the area near the center point automatically becomes more important than the surrounding areas which are more likely to be the background of the target image.

When there is a priori information about orientation and scale of the object of interest in the given test image, one can choose proper set of vertex templates. For example, this set may only include several rotated forms of the reference object with a fixed scale, if we know that the specified object is not scaled in the test image.

Most of the time-consuming tasks in our detection system can be done in the offline step. The LUT initialization for log-polar mapping, designing lifting step filter for adaptive wavelet transform, calculating correlation matrix of vertex templates in log-polar domain for parametric template matching, are all required to be done once for the given reference template in the offline step. Moreover, at the online step, there is not any iterative algorithm or any loop for covering all possible rotation angles and scaling factors. The similarity value for each location is obtained using direct form calculations.

Computational time of the proposed object detection system on the Pentium 2.8 GHz personal computer is 4.5 s at the offline step and 52 s at the online step (0.8 ms for calculating similarity value of each desired location within the whole test image). This is calculated based on the average runtime of our experiments for the typical settings presented in Table 2.

Considering a fixed scale and orientation for the object of interest, the computational time of the conventional normalized cross-correlation (NCC) template matching method is 8 s for the typical settings of Table 2. If the search for the object rotation range from −60 to 60 with 1 resolution, and search for the scaling factor range from 0.5 to 1.8 with 0.1 steps, computational time of the NCC method will dramatically increase to 6344 s.

The average percentage of time that each main building block at the online step. The average percentage of time that is taken by each main building block at the online step.

### Table 3

<table>
<thead>
<tr>
<th>Task</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initializing the look-up table (LUT) for log-polar mapping</td>
<td>11.6%</td>
</tr>
<tr>
<td>Designing the adaptive lifting scheme wavelet transform</td>
<td>80.3%</td>
</tr>
<tr>
<td>Calculation of correlation matrix ( \mathbf{C}_{uv} ) and estimating ( x, \theta )</td>
<td>0.1%</td>
</tr>
<tr>
<td>Log-polar mapping of vertex templates and reference template</td>
<td>8%</td>
</tr>
</tbody>
</table>

5.2. Object detection system example experiment

The following example is provided as a simple experiment to show the result of proposed object detection system. In this example, we have chosen toy letter “L” shown in Fig. 4a as the reference template. The rotation angle and scaling factor parameters used for the 5 vertex templates are \((-40, 0.80), (-10, 0.75), (0, 0.90), (15, 0.70)\). The Cartesian domain image of these vertex templates are shown in Fig. 4b. The input test image shown in Fig. 4c contains two instances of the toy letter “L”. The rotation angle and scaling factor parameters for the one at the top-right is \((30, 0.85)\) and for the one at the bottom-left is \((-20, 0.75)\). We have obtained the similarity value at each location of the test image. The resulted values are depicted as an intensity map in Fig. 4e for the normalized cross correlation (NCC) based method and Fig. 4f shows the intensity map of the similarity values calculated by the proposed object detection system. The typical values presented in Table 2 have been used as the main parameters of the detection system. The blocks corresponding with highest peaks of the intensity map for the proposed method (Fig. 4f) are illustrated over the test image in Fig. 4d. The calculated value for the rotation angle and scaling factor are also applied to the illustrated bounding boxes. The error values of the estimated rotation angles were less than 5° and for the scaling factors were less than 0.05. There is no such discrimination among peak points of the intensity map for the NCC based method and its highest peaks are not matched to the true location of the reference template.

5.3. Object detection system evaluation

Due to page limit, we have excluded the experimental results of the core detection algorithm (Section 4.2) on 1D signals and images. In addition, some experimental results on the core detection algorithm may be found in our previously published papers [18,53].

We have not compared the accuracy of our object detection system with other existing methods, because as we have elaborated in Section 3, apparently, there is no simultaneous rotation and scale invariant method with similar input parameters in the literature for having a fair comparison. However, we have conducted a comprehensive set of experiments to obtain a detailed evaluation for the results of the proposed object detection system.
We have selected 10 different test images similar to the one presented in Fig. 5c. Each test image have different configuration of the toy letters. Toy letter “O” which is shown in Fig. 5a, is selected as the reference template. A sample test image is created by placing rotated and scaled form of the reference template instead of the empty black box in Fig. 5c and adding one of the three different background images (texture1, marble, and trees) shown in Fig. 5d.

Fig. 4. Template matching example. (a) Reference template, (b) vertex templates in the Cartesian domain, (c) test image, (d) result of the proposed object detection system illustrated on the input test image, (e) intensity map of the similarity values by the NCC method, (f) intensity map of the similarity values by the proposed method.

Fig. 5. Template matching evaluation example. (a) Reference template, (b) vertex templates in the Cartesian domain, (c) sample test image configuration, (d) three different background images (texture1, marble, and trees), (e) sample test image after adding marble background and toy letter “O”, (f) intensity map of the similarity values by the proposed method for the illustrated sample test image (e).
Rotation angle range of the reference template instance in the test image was chosen from $-90^\circ$ to $+90^\circ$ with $15^\circ$ steps (13 different rotation angles). Scaling factor range of the reference template instance in the test image was from 0.5 to 1.5 with 0.1 steps (11 different scaling factors). A sample test image which was generated by using $-30^\circ$ for rotated and 0.8 for scaled version of the reference template and adding “marble” background, is shown in Fig. 5e. Considering 10 different configuration for basic toy letter test images, 3 different background images, 13 different rotation angles and 11 different scaling factors for the reference template instance, this experiment consists of 4290 different test images. From another point of view, for each combination of values in the chosen range of rotation angles and scaling factors, we conducted 30 different tests. The proposed detection system is used to locate reference template in each test image. The typical settings presented in Table 2 were used in this experiment. The 15 vertex templates shown in Fig. 5b were obtained by the combination of five rotation angles ($-75^\circ$, $-30^\circ$, $0^\circ$, $+30^\circ$, $+75^\circ$) and three scaling factors (1.3, 1.0, 0.7).

Correct detection percentages obtained by a total of 4290 tests are illustrated in Fig. 6a for the proposed object detection system and in Fig. 6b for SIFT. Each point in this figure represents the correct detection percentage obtained by 30 different tests for a fixed rotation angle from $-90^\circ$ to $+90^\circ$ and scaling factor from 0.5 to 1.5. Excluding the boundary results obtained in scales 0.5 and 1.5 and rotated angles of $-90^\circ$ and $+90^\circ$, the correct detection rate is above 85%. It was also over 96% when the template instance in the test image have rotation angle and scaling factor parameter values equal or near to the parameter values of a vertex template. Here, the proposed detection system is more accurate than SIFT due to proper selection of vertex templates. We have conducted this comparison, as SIFT is the widely accepted benchmark method for object detection in the literature. The input parameters of our proposed system and SIFT are not similar and it is difficult to have a fair comparison, therefore we...
The overall performance is rather good as in the majority of tests the mean error of the estimated rotation angle is less than 15° and the mean error of the estimated scaling factor is less than 0.15.

6. Conclusions and future works

In this paper, we have presented a novel adaptive lift wavelet transform algorithm for detecting objects of interest in a given test image. Many variations of the proposed detection algorithm could be designed to improve its performance for detecting noisy and degraded forms of the desired objects. For example, one may consider deeper levels of the wavelet packet transform tree for detecting dilated and condensed objects. Moreover, one may use several different instances of the desired object for designing the prediction filter in the adaptive transform. This would make the algorithm more robust to slight object deformations in the test image.

It was also shown that how the proposed detection algorithm can be combined with the proper log-polar mapping model and parametric template space to poses rotation/scale invariant property. The resulting object detection system avoids any iterative calculations. The similarity measure at each location of the given test image can be obtained through direct form calculations. The computation time can be improved using hardware implementation for the log-polar mapping and adopting some of the rejection related schemes [52]. Furthermore, in some of the applications one may avoid calculating similarity measure for all of the image pixels by using a search window or image pyramid approximation techniques. Moreover, color components may be considered in the future developments. Currently, we are developing a new model for matching image interest points by designing adaptive lifting scheme transform for the specific spatially sampled image patches around each interest point. The new model will provide resistance to partial occlusion and is relatively insensitive to changes in the viewpoints.

References


About the Author—MAHDI AMIRI received his M.Sc. in Computer Architecture Engineering from Sharif University of Technology, Tehran, Iran in 2003 and a B.Sc. in Electrical Engineering from University of Tabriz, Tabriz, Iran. He is currently a Ph.D. student at the Department of Computer Engineering, Sharif University of Technology. His current research interests include adaptive signal processing, wavelet theory, lifting scheme and object detection.

About the Author—HAMID R. RABIEE (SM) received his B.S. and M.S. degrees (with great distinction) in Electrical Engineering from CSULB, USA, his EEE in Electrical and Computer Engineering from USC, USA and his Ph.D. in Electrical and Computer Engineering from Purdue University, West Lafayette, USA in 1996. From 1993 to 1996 he was a Member of Technical Staff at AT&T Bell Laboratories. From 1996 to 1999 he worked as a Senior Software Engineer at Intel Corporation. He was also with PSU, OGI and OSU universities as an adjunct professor of Electrical and Computer Engineering from 1996 to 2000. Since September 2000, he has joined Sharif University of Technology (SUT), Tehran, Iran. He is the founder of Sharif University’s Advanced Information and Communication Technology Research Center (AICTC). Sharif University Advanced Technologies Incubator (SATI), Sharif Digital Media Laboratory (DML) and Mobile Value Added Services (MVAS) laboratories. He is currently an Associate Professor of the Computer Engineering at Sharif University of Technology, an Adjunct Professor of Computer Science at UNB, Canada, and the Director of AICTC, DML and MVAS. He has been the initiator and director of national and international level projects in the context of UNDP International Open Source Network (IOSN) and Iran’s National ICT Development Plan. Dr. Rabiee has received numerous awards and honors for his Industrial, scientific and academic contributions. He has acted as chairman in a number of national and international conferences, and holds three patents. He is also a Senior Member of IEEE.