EURASIP Journal on Applied Signal Processing 2005:13, 1956–1968

An Application of MAP-MRF to Change Detection in Image Sequence Based on Mean Field Theory

Qiang Liu
Laboratory for Computational Neuroscience, Departments of Neurological Surgery and Electrical Engineering,
University of Pittsburgh, Pittsburgh, PA 15213, USA
Email: qliu@neuronet.pitt.edu

Robert J. Sclabassi
Laboratory for Computational Neuroscience, Departments of Neurological Surgery and Electrical Engineering,
University of Pittsburgh, Pittsburgh, PA 15213, USA
Email: bobs@neuronet.pitt.edu

Ching-Chung Li
Laboratory for Computational Neuroscience, Departments of Neurological Surgery and Electrical Engineering,
University of Pittsburgh, Pittsburgh, PA 15213, USA
Email: ccl@engr.pitt.edu

Mingui Sun
Laboratory for Computational Neuroscience, Departments of Neurological Surgery and Electrical Engineering,
University of Pittsburgh, Pittsburgh, PA 15213, USA
Email: mrsun@neuronet.pitt.edu

Received 24 December 2003; Revised 22 October 2004

Change detection is one of the most important problems in video segmentation. In conventional methods, predetermined thresholds are utilized to test the variation between frames. Although certain reasonings about the thresholds are provided, appropriate determination of these parameters is still problematic. We present a new approach to change detection from an optimization point of view. We model the video frames and the change detection map (CDM) as Markov random fields (MRFs), and formulate change detection into a problem of seeking the optimal configuration of the CDM. Under the MRF assumption, the optimal solution, in the sense of maximum a posteriori (MAP), is obtained by minimizing the energy function associated with the MRF which is designed by utilizing the prior knowledge of noise and contextual constraints on the video frames. An algorithm that computes the potentials and optimizes the solution is constructed by applying the mean field theory (MFT). The experimental results show that the new method detects changes accurately and is robust to noise.

Keywords and phrases: change detection, image processing, Markov random field, mean field theory, video segmentation.

1. INTRODUCTION

Content-based video processing has been widely studied and is supported by a number of standards, such as MPEG-4 for video object-based compression and MPEG-7 for video content description [1, 2]. These standards involve functionalities that rely on segmenting video sequences into semantic regions or video objects. Change detection, which generates an initial segmentation mask, usually constitutes the first step of video segmentation [3, 4].

Much research effort has been devoted to change detection in recent years [5, 6, 7, 8]. Most existing approaches focus on thresholding which contains two essential steps of defining a metric function of intensity variation and choosing a proper threshold to be applied to the metric function. The key issue of these methods is to determine the threshold. However, it is often problematic choosing the threshold, since a large threshold removes noise as well as the signal (change caused by motion), while a small threshold makes the detection sensitive to noise. One way to determine the threshold is to introduce contextual constraints. Aach proposed a multiple-threshold approach from a framework of maximum a posteriori (MAP) estimation [9]. Unfortunately, this method in general does not provide a MAP solution, because the thresholds are chosen in a deterministic fashion.

In this work, we present a new approach from a strict optimization point of view. We consider the change detection problem in a global perspective. If we take the change detection map (CDM) as a 2D binary random field, then labeling each pixel as "changed" or "unchanged" becomes a problem of finding an appropriate configuration of this random field. This concept may be implemented by employing Markov random fields (MRF) theory, which is a well-known model in describing image characteristics [10]. In general, this theory says that the value of a random variable at one site in a MRF is only affected by the values of variables at its neighboring sites (determined by a neighborhood system that is defined). If one can describe the interactions between the neighboring sites, then it is possible to obtain a global description of the whole field. Moreover, if an a priori of image characteristics is applied in describing the interaction between neighboring sites (pixels), then one may obtain an optimal solution associated with the MRF in MAP sense [11]. For the change detection problem, the interaction between neighboring sites can be translated into contextual constraints between neighboring pixels. A simple example would be the constraint of smoothness, which means that the neighboring pixels of a changed/unchanged pixel are likely to be changed/unchanged too. These contextual constraints come from prior knowledge of our assumption on the image sequence being analyzed. Based upon this knowledge, the CDM from a pair of frames can be appropriately modeled as MRFs, by choosing the neighborhood system and formulating the impact between the neighboring sites. The rest of the task is then to search for a configuration of the CDM that satisfies the MAP criterion. In the literature, there are several methods to perform this search using, for example, simulated annealing and iterative conditional mode algorithms [12, 13]. The former aims at providing the global extremum, but requires extensive computation; the latter reduces the computational cost, but may converge to a local extremum. We adopt the mean field theory (MFT) approach as studied recently in [14, 15], which trades off between these two approaches.

This paper is organized as follows: Section 2 gives a brief review of the related theories; Section 3 describes the proposed methods and algorithms of change detection; Section 4 presents the experimental results based on the proposed method; and Section 5 provides a conclusion.

2. BACKGROUND THEORIES

Fundamentals of the MRF and the MFT are briefly introduced in this section.

2.1. Markov random field theory in change detection

Let $\bar{F} = \{F_{1,2}, \dots, F_{i,j}, \dots, F_{m,n}\}$ be a 2D random array, where $F_{i,j}$, $1 \le i \le m$, $1 \le j \le n$, is a random variable at site (i, j). Let $S = \{(i, j) | 1 \le i \le m, 1 \le j \le n\}$ be the set of all sites. Frame $\bar{f} = \{f_{i,j}, (i, j) \in S\}$ is a realization of $\bar{F}$. Let $p(\bar{f})$ denote the joint probability density function (pdf) of $\bar{F} = \bar{f}$, where $p(\bar{f}) = p\{\bar{F} = \bar{f}\} = p\{F_{i,j} = f_{i,j}, (i, j) \in S\}$. Then, with the same notation, $\bar{F}$ is a MRF if (1) $p(\bar{f}) > 0$, for all $\bar{f} \in \bar{F}$, and (2) $p(f_{i,j} | f_{S'}) = p(f_{i,j} | f_{N_{i,j}})$, where $S' = S - (i, j)$, with symbol "−" denoting exclusion, and $N_{i,j} = \{(i', j') | (i - i')^2 + (j - j')^2 \le k, (i', j') \in S'\}$, with k being a positive integer. $N_{i,j}$ defines the set of the kth order neighboring sites of (i, j). With the definition of $N_{i,j}$, a clique, denoted by c, is defined as a set containing single or multiple sites that are connected within $N_{i,j}$, $(i, j) \in S$. Figure 1 illustrates an example of cliques of a first-order neighborhood, where c may be a collection of single sites or double sites. It was introduced in [16] that the joint pdf $p(\bar{f})$ may be approximated by the Gibbs distribution:

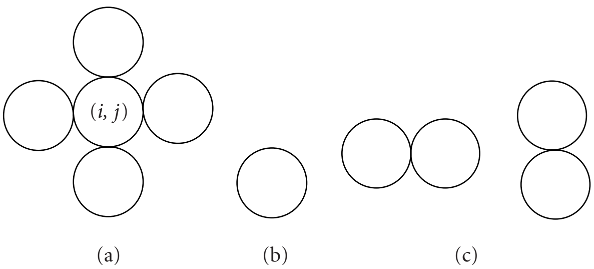


FIGURE 1: (a) A first-order neighborhood system, (b) single-site clique, and (c) double-site cliques.

$$p(\bar{f}) = \frac{e^{-(1/T)U(\bar{f})}}{\sum_{\bar{f}} e^{-(1/T)U(\bar{f})}}, \tag{1}$$

where T is a constant and U is an energy function of the MRF given by

$$U(\bar{f}) = \sum_{c} V_c(\bar{f}) \tag{2}$$

with V_c's being clique potentials or clique functions. The V_c functions represent contributions to the total energy from single-site cliques, double-site cliques, and so forth. Note that (1) and (2) reflect the fact that the global identity $p(\bar{f})$ is determined by the local activities, namely, the clique potentials. Considering the first-order neighborhood, we may rewrite (2) into the following form [11]:

$$\begin{aligned} U(\bar{f}) = \sum_{(i,j)} \{ & V_{(i,j)}(f_{i,j}) + V_{\{(i,j),(i+1,j)\}}(f_{i,j}, f_{i+1,j}) \\ & + V_{\{(i,j),(i,j+1)\}}(f_{i,j}, f_{i,j+1}) \}, \end{aligned} \tag{3}$$

where the first, second, and third term are single-site, horizontal double-site, and vertical double-site clique potentials, respectively. Notice that for a double-site clique $\{(i, j), (i', j')\}$, the associated clique potentials $V_{\{(i,j),(i',j')\}}(f_{i,j}, f_{i',j'})$ and $V_{\{(i',j'),(i,j)\}}(f_{i',j'}, f_{i,j})$ are equal. Therefore, (3) may be rearranged into

$$\begin{aligned} U(\bar{f}) &= \sum_{(i,j)} \left\{ V_{c_1}(f_{i,j}) + \frac{1}{2} \sum_{(i',j') \in N_{i,j}} V_{c_2}(f_{i,j}, f_{i',j'}) \right\} \\ &= \sum_{(i,j)} U_{i,j}(f_{i,j}), \end{aligned} \tag{4}$$

where c_1 and c_2 are single-site and double-site cliques in the defined neighborhood, and $U_{i,j}(f_{i,j})$ is the energy function associated with site (i, j). As pointed out in [11], if $p(\bar{f})$ is a posterior distribution, minimizing the energy function $U(\bar{f})$ yields a MAP estimate of the joint pdf $p(\bar{f})$.

2.2. Mean field theory

To make the MRF theory more practical, we need to introduce the MFT. From the description of the MRF, we know that the value assigned to a random variable in the MRF is affected by the values at its neighboring sites, which are further dependent on their neighbors. One way to calculate the interaction between one site and its neighbors is to apply the MFT [14, 17], which assumes that the impacts from the neighbors can be approximated by an average field. We denote the mean field for site (i, j) by $f_{i,j}^{\mathrm{mf}}$. As a result, if the first-order neighborhood is considered, one may write the energy function related to site (i, j) in the following form [14]:

$$U_{i,j}^{\mathrm{mf}}(f_{i,j}) = V_{c_1}(f_{i,j}) + \sum_{(i',j')\in N_{i,j}} V_{c_2}(f_{i,j}, f_{i',j'}^{\mathrm{mf}}), \tag{5}$$

where $V_{c_1}(\cdot)$ and $V_{c_2}(\cdot,\cdot)$ are potential functions of single-site and double-site cliques, respectively; and $f_{i',j'}^{\mathrm{mf}}$ is the mean field for $f_{i',j'}$. Then, the marginal distribution of the MRF at site (i, j) may be approximated by [14]

$$p(f_{i,j}) = \frac{1}{\sum_{f_{i,j}} e^{-(1/T)U_{i,j}^{\mathrm{mf}}(f_{i,j})}} e^{-(1/T)U_{i,j}^{\mathrm{mf}}(f_{i,j})}. \tag{6}$$

As seen from (4) and (5), the energy function is decomposed into local computations, where each site is treated independently. Therefore, the joint pdf $p(\bar{f})$ can be approximated by

$$p(\bar{f}) \approx \prod_{i,j} p(f_{i,j}). \tag{7}$$

Then, maximizing $p(\bar{f})$ is equivalent to maximizing each $p(f_{i,j})$, or, to minimizing the corresponding $U_{i,j}^{\mathrm{mf}}(f_{i,j})$.

In order to evaluate $U_{i,j}^{\mathrm{mf}}(f_{i,j})$, the mean field values $f_{i',j'}^{\mathrm{mf}}$ at the neighboring sites (i', j') within $N_{i,j}$ must be computed. The general way to calculate a mean field value is by the following form:

$$f_{i,j}^{\mathrm{mf}} = \sum_{f_{i,j}} f_{i,j} \cdot p(f_{i,j}). \tag{8}$$

Note that (8) requires the evaluation of $p(f_{i,j})$, henceforth, $U_{i,j}^{\mathrm{mf}}(f_{i,j})$. Therefore, the computation of the mean field value is usually carried out by iteration that stops when the change of the results from two consecutive iterations is sufficiently small.

3. MRF CHANGE DETECTION METHOD

3.1. MAP-MRF in change detection

We denote the CDM by $\bar{H} = \{H_{1,2}, \dots, H_{i,j}, \dots, H_{m,n}\}$, and $\bar{h} = \{h_{1,2}, \dots, h_{i,j}, \dots, h_{m,n}\}$ a configuration of $\bar{H}$, where $h_{i,j} \in \{-1, 1\}, (i, j) \in S$ with "-1" denoting unchanged and "1" denoting changed. Then, given two frames $\bar{f}^{(0)}$ and $\bar{f}^{(1)}$, our goal is to find the optimal $\bar{h}^*$ in the MAP sense, such that

$$\begin{aligned}\bar{h}^* &= \operatorname{argmax}_{\bar{h}}\, p(\bar{h} \mid \bar{f}^{(0)}, \bar{f}^{(1)}) \\ &= \operatorname{argmax}_{\bar{h}} \frac{p(\bar{f}^{(1)} \mid \bar{f}^{(0)}, \bar{h}) \cdot p(\bar{h} \mid \bar{f}^{(0)})}{p(\bar{f}^{(1)} \mid \bar{f}^{(0)})} \\ &= \operatorname{argmax}_{\bar{h}}\, p(\bar{f}^{(1)} \mid \bar{f}^{(0)}, \bar{h}) \cdot p(\bar{h} \mid \bar{f}^{(0)}).\end{aligned} \tag{9}$$

Applying MRF assumption on both $\bar{F}$ and $\bar{H}$, maximizing $p(\bar{h} \mid \bar{f}^{(0)}, \bar{f}^{(1)})$ with respect to $\bar{h}$ is equivalent to minimizing its energy function $U(\bar{h} \mid \bar{f}^{(0)}, \bar{f}^{(1)})$. This, as suggested by (9), can be accomplished by minimizing the energy functions $U(\bar{f}^{(1)} \mid \bar{h}, \bar{f}^{(0)})$ and $U(\bar{h} \mid \bar{f}^{(0)})$, which are associated with $p(\bar{f}^{(1)} \mid \bar{f}^{(0)})$ and $p(\bar{h} \mid \bar{f}^{(0)})$, respectively. $U(\bar{f}^{(1)} \mid \bar{h}, \bar{f}^{(0)})$ addresses the potential of the likelihood between $\bar{f}^{(1)}$ and $\bar{f}^{(0)}$ with the knowledge of $\bar{h}$, that is, whether the sites are changed. And, $U(\bar{h} \mid \bar{f}^{(0)})$ is always considered to represent the spatial domain constraints, for example, the smoothness or similarity between neighboring sites. Therefore, a general form of the prior model of these energy functions is

$$U(\bar{h} \mid \bar{f}^{(0)}, \bar{f}^{(1)}) = \gamma_f U(\bar{f}^{(1)} \mid \bar{h}, \bar{f}^{(0)}) + \gamma_h U(\bar{h} \mid \bar{f}^{(0)}), \tag{10}$$

where γ_f and γ_h are regularization parameters. The larger the regularization parameter values, the more the corresponding constraint is emphasized.

Equivalently, we can write (10) by

$$U(\bar{h} \mid \bar{f}^{(0)}, \bar{f}^{(1)}) = \gamma_f [U(\bar{f}^{(1)} \mid \bar{h}, \bar{f}^{(0)}) + \gamma U(\bar{h} \mid \bar{f}^{(0)})], \tag{11}$$

where $\gamma = \gamma_h/\gamma_f$. It is noticed that to minimize $U(\bar{h} \mid \bar{f}^{(0)}, \bar{f}^{(1)})$ with respect to $\bar{h}$ is equivalent to minimizing $U(\bar{f}^{(1)} \mid \bar{h}, \bar{f}^{(0)}) + \gamma U(\bar{h} \mid \bar{f}^{(0)})$. Therefore, we define the energy function in the following form:

$$U(\bar{h} \mid \bar{f}^{(0)}, \bar{f}^{(1)}) = U(\bar{f}^{(1)} \mid \bar{h}, \bar{f}^{(0)}) + \gamma U(\bar{h} \mid \bar{f}^{(0)}). \tag{12}$$

In order to design the above energy functions, one needs to employ the prior knowledge. In our application, the prior knowledge includes the distribution of the frame difference in the absence/presence of changes and the assumption of the similarity between immediate sites (pixels). There are no specific routines in designing potential functions. In general, as indicated in [10], the formulation of a potential function should maintain consistency with the prior knowledge: if the formulation of the regions in a clique tends to be consistent with the prior knowledge, the value of the energy function decreases; otherwise, the value increases.

In change detection, we interpret $U(\bar{f}^{(1)}|\bar{h}, \bar{f}^{(0)})$ as the sum of single-site clique potentials, which is

$$U(\bar{f}^{(1)}|\bar{h}, \bar{f}^{(0)}) = \sum_{c_1} V_{c_1}(\bar{f}^{(1)}|\bar{h}, \bar{f}^{(0)}) = \sum_{i,j} V_{c_1}(f_{i,j}^{(1)}|h_{i,j}, f_{i,j}^{(0)}), \tag{13}$$

where V_{c_1} is selected to be

$$V_{c_1}(f_{i,j}^{(1)}|h_{i,j}, f_{i,j}^{(0)}) = -\ln\left(p(d_{i,j} \mid h_{i,j})\right) \tag{14}$$

which is the negative of the natural logarithm of the pdf of the absolute frame difference $d_{i,j} = |f_{i,j}^{(1)} - f_{i,j}^{(0)}|$ at site $(i, j) \in S$, given the knowledge of $h_{i,j}$. Therefore, if $d_{i,j}$ is consistent with the prior belief, the conditional probability will be high. As a result, its logarithm value will be low, and vice versa, as required by the design rules. Choosing the natural logarithm is instinctive. First, more penalty would be assigned to smaller probability, for example, when probability is close to zero, the value of energy function would be extremely large. Second, considering $p(\bar{f}^{(1)}|\bar{f}^{(0)}, \bar{h} = -1)$, which is equivalent to the pdf of frame difference caused by noise, we may assume $p(\bar{f}^{(1)}|\bar{f}^{(0)}, \bar{h} = -1) = \prod_{i,j} p(d_{i,j} \mid h_{i,j} = -1)$, or $\prod_{i,j} Z_{i,j} \cdot e^{-(1/T)V_{c_1}(f_{i,j}^{(1)}|h_{i,j}=-1, f_{i,j}^{(0)})} = \prod_{i,j} p(d_{i,j} \mid h_{i,j} = -1)$, where $Z_{i,j}$ are normalization constants. Furthermore, if the noise distribution $p(d_{i,j} \mid h_{i,j} = -1)$ also has an exponential form, such as Gaussian and Laplacian, we may reasonably take the natural log on both sides of the above equation to get the potential function. For the case of $h_{i,j} = 1$, that is, with the presence of change, the independence assumption may not hold in general. However, this assumption can be accepted as a reasonable simplification to trade off computational complexity [18]. Therefore, the above reasoning may also apply to the case $h_{i,j} = 1$. The collection of prior knowledge will be described in Section 3.2.

The other energy function $U(\bar{h}|\bar{f}^{(0)})$ in (10) addresses the contextual constraints on the neighboring sites. This can be explained as follows: with the knowledge of $\bar{f}^{(0)}$, we want to obtain $\bar{h}$ that complies with the properties of $\bar{f}^{(0)}$, for example, the continuity of $\bar{h}$ if we assume that $\bar{f}^{(0)}$ is smooth. Based upon this reasoning, we define

$$U(\bar{h}|\bar{f}^{(0)}) = \sum_{i,j} \sum_{c_2 \subset N_{i,j}} V_{c_2}(\bar{h}|\bar{f}^{(0)}) = \sum_{i,j} \left\{ \frac{1}{2} \sum_{(i',j') \in N_{i,j}} V_{c_2}(h_{i,j}, h_{i',j'}) \right\}, \tag{15}$$

where c_2 is a double-site clique in a first-order neighborhood $N_{i,j}$ at site $(i, j) \in S$. The scaling factor 1/2 has been explained in (3) and (4). The clique potential $V_{c_2}(\cdot, \cdot)$ is defined as

$$V_{c_2}(h_{i,j}, h_{i',j'}) = -\ln\left(1 - 0.5\left|h_{i,j} - \lambda \cdot h_{i',j'}\right|\right), \tag{16}$$

where $\lambda \in (0, 1)$ is a constant representing the impact of site (i', j') on site (i, j). The reasons behind this design are (1) we want the state of site (i, j) to agree with its neighboring sites; (2) the logarithm form is consistent with that in (14). The term $1 - 0.5|h_{i,j} - \lambda \cdot h_{i',j'}|$ acts as a probability of the random variable at site (i, j) when its value agrees with those at its neighboring sites. Therefore, this definition also follows the design rules stated previously.

To minimize $U(\bar{h}|\bar{f}^{(0)}, \bar{f}^{(1)})$, we must evaluate the clique potential functions. A question now is how to calculate $V_{c_2}(h_{i,j}, h_{i',j'})$. As mentioned previously, we may apply MFT to simplify this calculation. If the first-order neighborhood system is assumed, we have the following approximation:

$$U(\bar{h}|\bar{f}^{(0)}) \approx \sum_{i,j} \sum_{(i',j') \in N_{i,j}} V_{c_2}(h_{i,j}, h_{i',j'}^{\mathrm{mf}}), \tag{17}$$

where

$$V_{c_2}(h_{i,j}, h_{i',j'}^{\mathrm{mf}}) = -\ln\left(1 - 0.5\left|h_{i,j} - \lambda \cdot h_{i',j'}^{\mathrm{mf}}\right|\right). \tag{18}$$

Combining (10)–(18), we have

$$U(\bar{h}|\bar{f}^{(0)}, \bar{f}^{(1)}) \approx \sum_{i,j} U_{i,j}^{\mathrm{mf}}(h_{i,j}|f_{i,j}^{(0)}, f_{i,j}^{(1)}), \tag{19}$$

where

$$U_{i,j}^{\mathrm{mf}}(h_{i,j}|f_{i,j}^{(0)}, f_{i,j}^{(1)}) = -\ln\left(p(d_{i,j} \mid h_{i,j})\right) - \left[\gamma \sum_{(i',j') \in N_{i,j}} \ln\left(1 - 0.5\left|h_{i,j} - \lambda \cdot h_{i',j'}^{\mathrm{mf}}\right|\right)\right]. \tag{20}$$

Essentially, to minimize $U(\bar{h}|\bar{f}^{(0)}, \bar{f}^{(1)})$, we only need to evaluate $U_{i,j}^{\mathrm{mf}}(\cdot)$ at each site (i, j), and choose $h_{i,j}$ between -1 and 1 to render a smaller value of $U_{i,j}^{\mathrm{mf}}(\cdot)$.

3.2. The MRF change detection algorithm

Equation (20) requires evaluation of $p(d_{i,j}|h_{i,j})$, $(i, j) \in S$. Instead of collecting the pdf for each site, we utilize the same pdf, denoted by $p(d|h)$, for all sites, where d and h have the same sample spaces as $d_{i,j}$ and $h_{i,j}$, respectively. This choice is motivated from a practical point of view, since it would be extremely expensive to allocate memory for $p(d_{i,j}|h_{i,j})$ for each $(i, j) \in S$. When $h(i, j) = -1$, this approximation can be justified because the value differences of unchanged sites are driven by noise, which is usually considered to be independently and identically distributed. For moving pixels, the above assumption is not true in general. However, if we assume that each pixel may experience the same or similar amounts of motion, the validity of using $p(d|1)$ for all the sites is also justifiable.

To train $p(d|-1)$, we utilize the video segments containing motionless scenes. This is relatively easy to accomplish in many applications, such as in surveillance and teleconference videos. In general, it is difficult to train $p(d|1)$; however, it is possible to train a prototype for specific applications. Practically, we adopt the following strategy to calculate $p(d|1)$: first, $p(d|1)$ is initialized to be a uniform distribution across the entire range of its sample space, that is, $p(d|1) = 1/(L+1)$, $d \in [0, L]$ for a discrete case; then, starting with the initial value, we adapt $p(d|1)$ during a detection process, using the following equation:

$$p^{(r)}(d|1) = (1 - \epsilon \cdot \rho) \cdot p^{(r-1)}(d|1) + \epsilon \cdot \rho \cdot p^{(r)}_{d|1}, \quad (21)$$

where $p^{(r)}(d|1)$ and $p^{(r-1)}(d|1)$ are the pdf $p(d|1)$ adapted from frame 1 to frames r and $r-1$, respectively, $p^{(r)}_{d|1}$ is the pdf of the changed pixels contained in frame r, ρ is the ratio of the number of changed pixels to the total number of pixels in that frame, and $\epsilon \in (0, 1)$ is a control parameter. The term ρ reflects the intuition that the more changed pixels there are, the more $p(d|1)$ should be adapted. Parameter ϵ is designed to control the rate of adaptation.

An important question now is how the mean field value $h^{\text{mf}}_{i,j}$, $(i, j) \in S$ is evaluated. As mentioned before, the mean field value is usually computed iteratively until it converges. As described in Section 2.2, with the local energy function $U^{\text{mf}}_{i,j}(h_{i,j}|f^{(0)}_{i,j}, f^{(1)}_{i,j})$, $h^{\text{mf}}_{i,j}$ can be evaluated by

$$h^{\text{mf}}_{i,j} = \sum_{h_{i,j}} h_{i,j} \cdot \frac{e^{-(1/T)U^{\text{mf}}_{i,j}(h_{i,j}|f^{(0)}_{i,j}, f^{(1)}_{i,j})}}{\sum_{h_{i,j}} e^{-(1/T)U^{\text{mf}}_{i,j}(h_{i,j}|f^{(0)}_{i,j}, f^{(1)}_{i,j})}}. \quad (22)$$

Applying (20), we have

$$e^{-(1/T)U^{\text{mf}}_{i,j}(h_{i,j}|f^{(0)}_{i,j}, f^{(1)}_{i,j})} = \left(p(d|h) \cdot \left(\prod_{(i',j') \in N_{i,j}} \left[1 - 0.5\left(h_{i,j} - \lambda h^{\text{mf}}_{i',j'}\right)\right] \right)^{\gamma} \right)^{1/T}. \quad (23)$$

Note that the computing time can be greatly reduced by using (23). The iteration continues until the following condition is satisfied:

$$\frac{1}{m \cdot n} \sum_{i,j} \left| h^{\text{mf}}_{i,j}(k+1) - h^{\text{mf}}_{i,j}(k) \right| < \theta, \quad (24)$$

where k is the index of iteration, $m \cdot n$ is the total number of pixels, and $\theta \in (0, 1)$ is a chosen threshold.

With these assumptions and simplifications, we present Algorithm 1 to implement the proposed model.

```
Step 1. Load $p(d| - 1)$ and initialize
        $p(d|1) = 1/256$, for $d = 0, 1, \ldots, 255$.
        Assign values to $\gamma$, $\lambda$, $\epsilon$, and $\theta$.
Step 2. Take two frames $\bar{f}^{(0)}$ and $\bar{f}^{(1)}$, and
        calculate $\bar{d} = |\bar{f}^{(0)} - \bar{f}^{(1)}|$; initialize
        mean field values $\bar{h}^{\text{mf}}$, where for each
        pixel $(i, j)$, $h^{\text{mf}}_{i,j} = 0$.
Step 3. For each pixel $(i, j)$, evaluate (20) with
        $h_{i,j} = -1$ and 1, and calculate the new
        mean field value by (22) and (23).
Step 4. Evaluate the difference between the new
        mean field value and the previous one
        as defined in (24); if the difference is
        less than $\theta$, then go to next step,
        otherwise go to step 3.
Step 5. For each pixel, if the local energy
        $U^{\text{mf}}_{i,j}(h_{i,j} = -1|f^{(0)}_{i,j}, f^{(1)}_{i,j}) > U^{\text{mf}}_{i,j}(h_{i,j} =$
        $1|f^{(0)}_{i,j}, f^{(1)}_{i,j})$, then label pixel $(i, j)$
        unchanged, otherwise changed.
Step 6. Update $p(d|1)$ by (21); finish if all the
        frames are done, otherwise go to step 2.
```

Algorithm 1: MRF-MFT change detection algorithm.

Table 1: Typical control parameters.

Parameter	T	γ	λ	ϵ	θ
Value	2	1	0.99	0.5	0.05

4. IMPLEMENTATION AND EXPERIMENTS

In this section, the experimental results based on the proposed method are reported. We present the results of two types of data: a synthetic image sequence generated by using Matlab (version R12, MathWorks Inc., Mass) and a selected set of the reference MPEG test sequences available in the public domain (e.g., http://sampl.eng.ohio-state.edu/~sampl/database.htm, http://www.neuronet.pitt.edu/~qliu/Links.htm). All the sequences are in the QCIF format (144 × 176 in size). Only the Y component is utilized to calculate frame differences.

As described previously, five controlling parameters T, γ, λ, ϵ, and θ are required. Table 1 lists typical values of these parameters, which were chosen experimentally and utilized for all the test sequences. In the following, we describe these parameters individually.

(i) T is called "temperature" in MRF-based methods, for example, simulated annealing algorithm [12]. This parameter determines the spread of the Gibbs distribution. The larger the T, the more it spreads. In simulated annealing, T is gradually decreased. However, as suggested by [19], a fixed T is able to render a satisfactory result while reducing the computational cost. Therefore, a constant T was utilized throughout our experiments.

(ii) γ is a regularization parameter to balance the constraints introduced by different clique potentials. In our application, a large γ value emphasizes the smoothness constraint.

(iii) λ models the impact between neighboring sites. In (16), $h_{i,j} - \lambda h_{i',j'}$ is utilized to represent the difference between neighboring sites (i, j) and (i', j'). The value of λ controls the degree of impact from (i', j').

(iv) ϵ is utilized to control the adaptation of the pdf of d in the presence of change. The larger the value of ϵ, the more the pdf adapts to each CDM, and the faster the adaption to test data. However, considering the risk of false detection, we assign ϵ a moderate value.

(v) θ provides a stop threshold in the calculation of the mean field values.

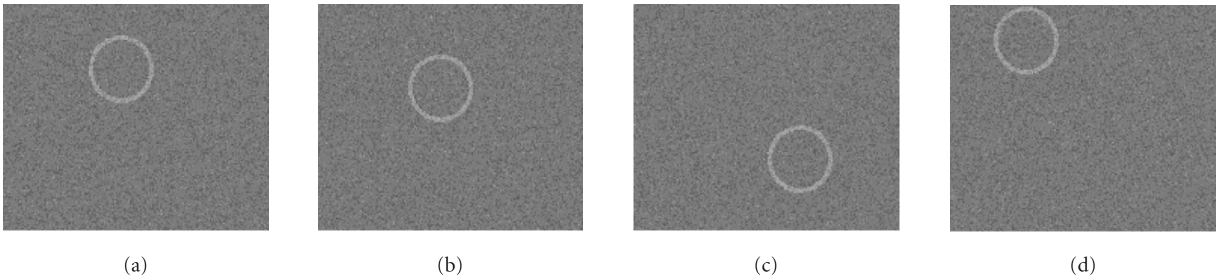


FIGURE 2: (a) Frame 1, (b) frame 2, (c) frame 35, and (d) frame 36.

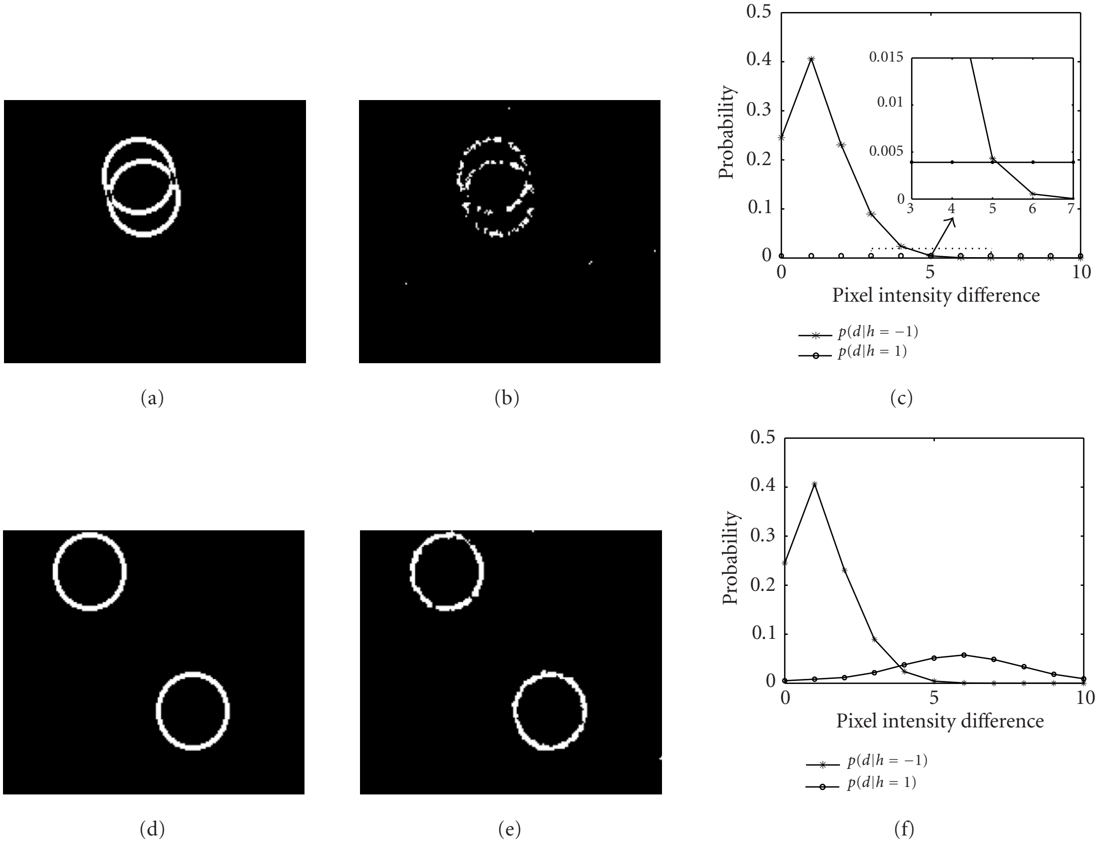


FIGURE 3: The upper plots represent the change detection results from frames 1 and 2: (a) the known CDM, (b) the detected CDM, and (c) $p(d|h=-1)$ and initial $p(d|h=1)$. The subplot embedded in (c) shows a close-look of the marked region (by the dashed line). The lower plots represent the change detection results from frames 35 and 36: (d) the known CDM, (e) the detected CDM, and (f) $p(d|h=-1)$ and $p(d|h=1)$ (adapted from frames $1 \sim 35$).

4.1. *Synthetic data*

To evaluate the new change detection method quantitatively, we generated a synthetic image sequence by using Matlab in the following way: a circle (with a radius of 20, line width of 3, both in pixels, and gray-level intensity of 5) is plotted in a frame; then, white Gaussian noise with mean 127 and standard deviation 1.6 is added to each frame. It should be noted that the signal-to-noise (SNR) ratio of the synthetic data, defined as 20 log(circle intensity/noise standard deviation), is less than 10 dB, which is much lower than the SNR in most natural videos. The coordinates of the origins were randomly generated. Two pairs of sample frames are shown in Figure 2. We denote the ground truth CDM by $\bar{h}^{(r)}$, the detected CDM by $\bar{h}^{(t)}$, and the set of sites with false labels by $S_e = \{(i,j)|h_{i,j}^{(r)} \neq h_{i,j}^{(t)},\ (i,j) \in S\}$. The error rate is then defined as

$$E_r = \frac{||S_e||}{||S||}, \tag{25}$$

where $\|S_e\|$ and $\|S\|$ denote the number of sites in S_e and S, respectively.

Figure 3 demonstrates the results of the synthetic data. The upper plots show the results obtained from frame 1 and 2. Figures 3a, 3b, and 3c show the ground truth CDM, the

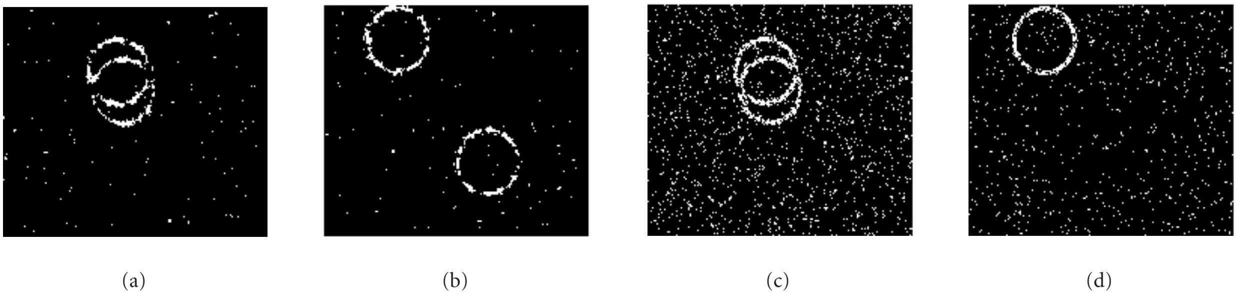


Figure 4: (a), (b) The CDMs detected by "quadratic picture function" (QPF) method: (a) CDM from frames 1 and 2; (b) CDM from frames 35 and 36. (c), (d) The CDMs detected by the method of De Geyter and Philips (M3 method): (c) CDM from frames 1 and 2; (d) CDM from frames 35 and 36.

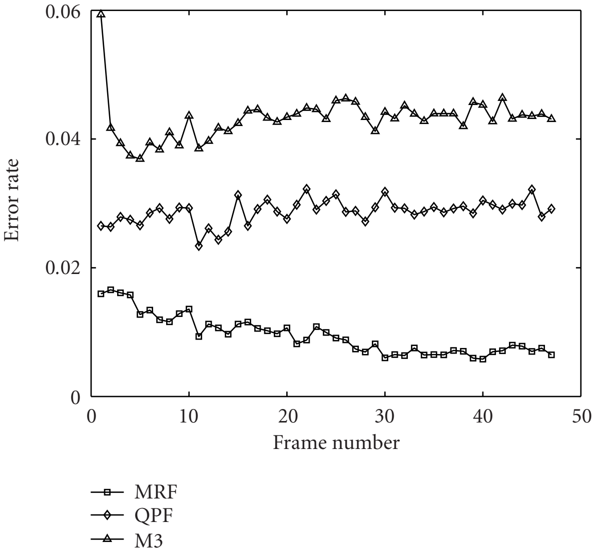


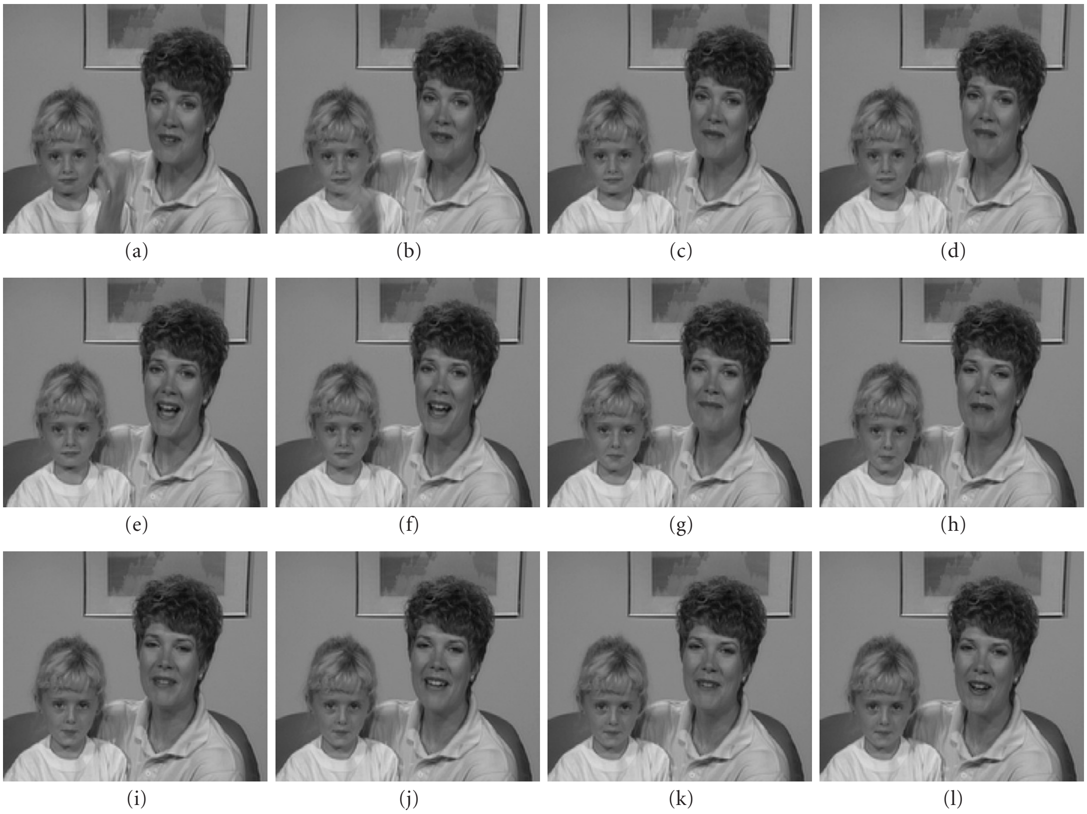
Figure 5: The error rates of our method MRF, the quadratic picture function method (QPF), and the method of De Geyter and Philips (M3).

detected CDM, and $p(d|h = -1)$ and the initial $p(d|h = 1)$, respectively. Compared with the ground truth CDM, the detected CDM has visible false detections. However, with the adaption of $p(d|h = 1)$, the false detections are reduced. As seen in the lower plots, where the results were obtained from frame 35 and 36, the detected CDM Figure 3e contains much less false detections. In Figure 3f it can be seen that $p(d|h = -1)$ was kept intact because of the assumption of stationary noise, but $p(d|h = 1)$ was adapted to a bell-shaped function according to (21).

To demonstrate the robustness of the MRF approach, we compare it with two existing methods, quadratic picture function (QPF) method developed by Hsu et al. [6], and a novel method ("Method 3," abbreviated as M3 in the following) recently presented by Geyter and Philips [20]. In the former method, the threshold value of 5.76 was selected, which corresponds to a significant level of 0.005. In the latter method, the parameters α, β, and z (see [20]) were set to 0.5, 0.9, and 3, respectively. The parameter k in M3 was tested from 2 to 5 and $k = 4$ was selected, which produced the best overall performance for the test sequences. These parameter values were utilized for all the test sequences (synthetic and natural). The results of QPF and M3 methods are illustrated in Figures 4a-4b and 4c-4d, respectively. Compared with the CDMs shown in Figure 3, these two methods appear to be more sensitive to the simulated noise. The error rates of the three methods are illustrated in Figure 5, which shows that the MRF method performed better than the two existing methods in terms of less false detection. It is seen that the error rate of the MRF method decreases as frames 1 through 30 are being processed, then becomes stable after that. The reason is that $p(d|h = 1)$ adapts gradually to the test data at the initial frames, and then becomes stationary. The adaptation speed is quite satisfactory for most common applications, as indicated by our results using other videos.

4.2. *Real-world data*

In this section, experimental results on selected MPEG test sequences are presented. Change detection was carried on these sequences at a rate of 10 frame pairs per second. First, we report the experiment on *Mother & Daughter* sequence by the proposed method. Figure 6 shows frames 58 through 91 which contain both large motions (e.g., hand movement in frame 58 and 61) and small motions (e.g., chest and shoulder movements). The detected CDMs are shown in Figure 7. It can be seen that the stationary background and the moving objects are well distinguished. The background area is quite clean, indicating that the MRF method is robust to the salt and pepper noise contained in this sequence. Figure 8 depicts the pdf's calculated from this sequence. While pdf $p(d|h = -1)$ was calculated from a background area that was manually selected, pdf $p(d|h = 1)$ was initialized and then adapted in the change detection process as described previously. Figure 8 shows the pdf's calculated progressively at frames 1,60, 300, 600, and 900.

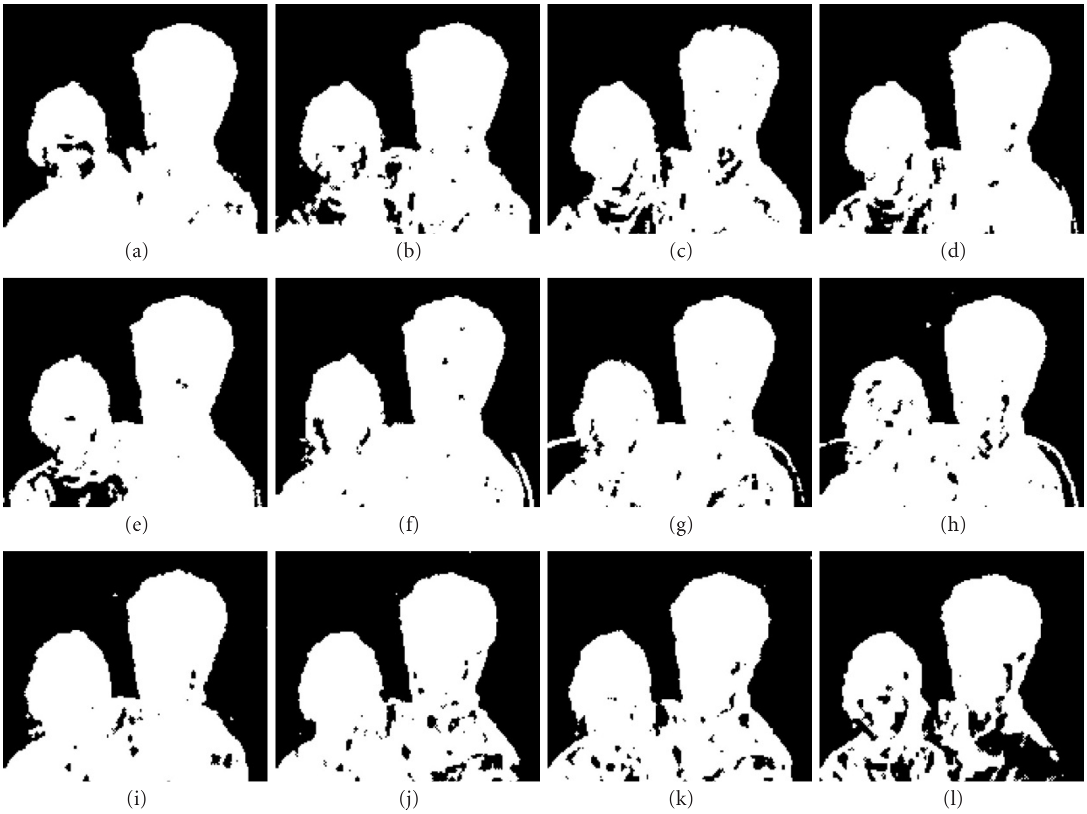


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 6: Frames of *Mother & Daughter* sequence. (a)–(l) Frames 58, 61, 64, 67, 70, 73, 76, 79, 82, 85, 88, and 91.

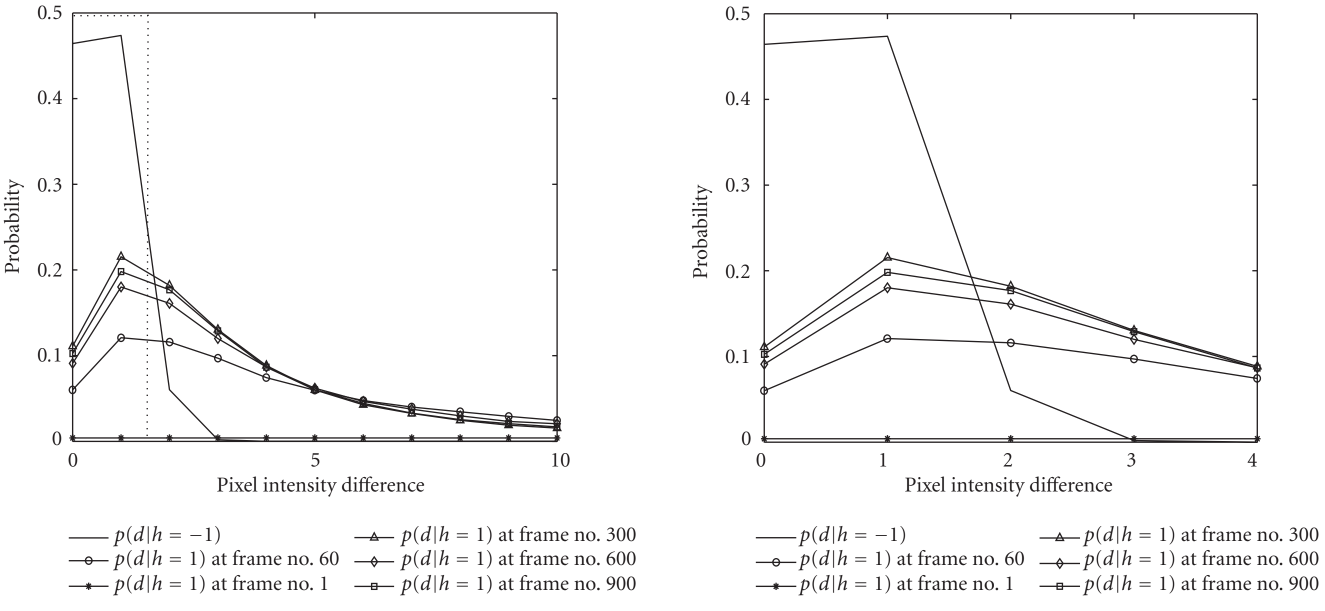


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 7: The detected CDMs from frames 58 through 91 of Mother & Daughter sequence, using the parameter values listed in Table 1. (a) (58, 61), (b) (61, 64), (c) (64, 67), (d) (67, 70), (e) (70, 73), (f) (73, 76), (g) (76, 79), (h) (79, 82), (i) (82, 85), (j) (85, 88), (k) (88, 91), and (l) (91, 94).



(a) (b)

Figure 8: The pdf's obtained from Mother & Daughter sequence. (a) $p(d|h=-1)$ and $p(d|h=1)$ at frames 1, 60, 300, 600, and 900. (b) A close look at the pdf's in the marked range (by the dashed line) in (a).

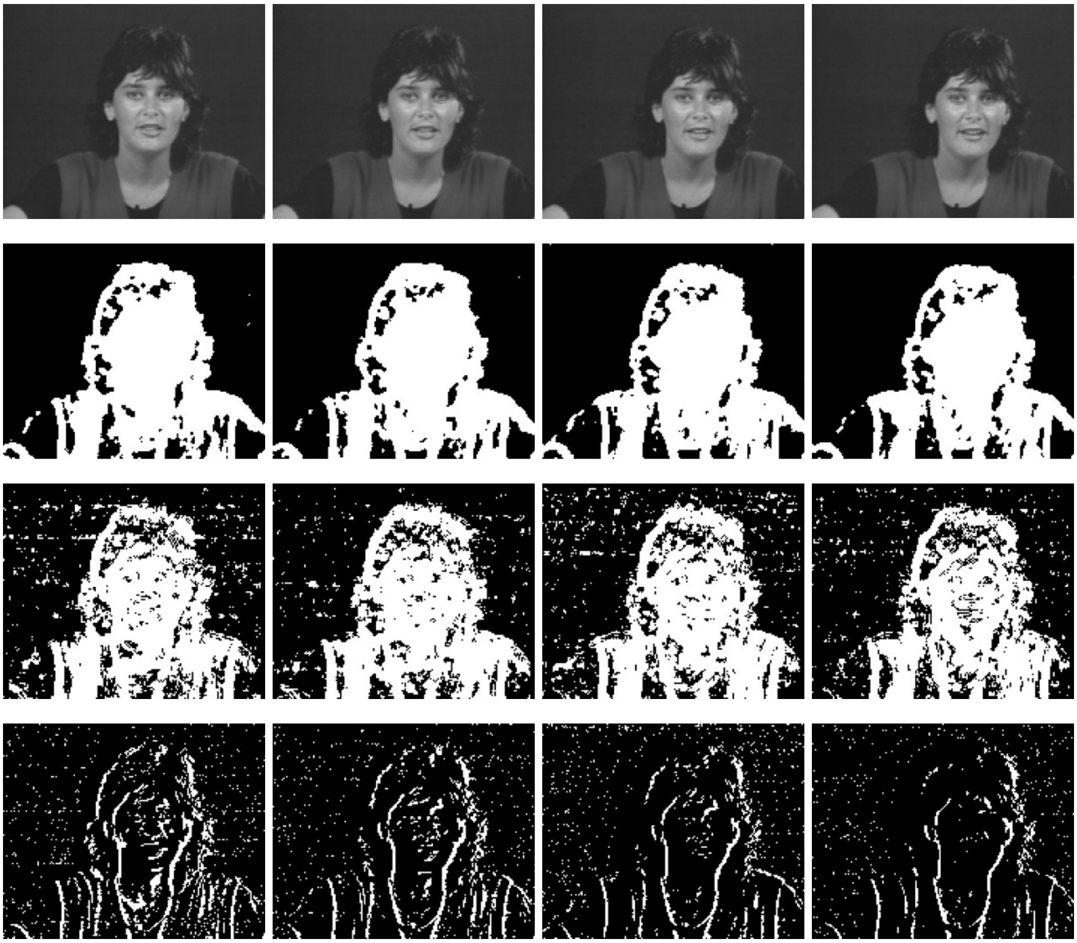


Figure 9: Experimental results of test sequence *Miss America*. From top to bottom: frames 75, 78, 81, and 84; CDMs detected by our method, by the QPF method, and by the M3 method.

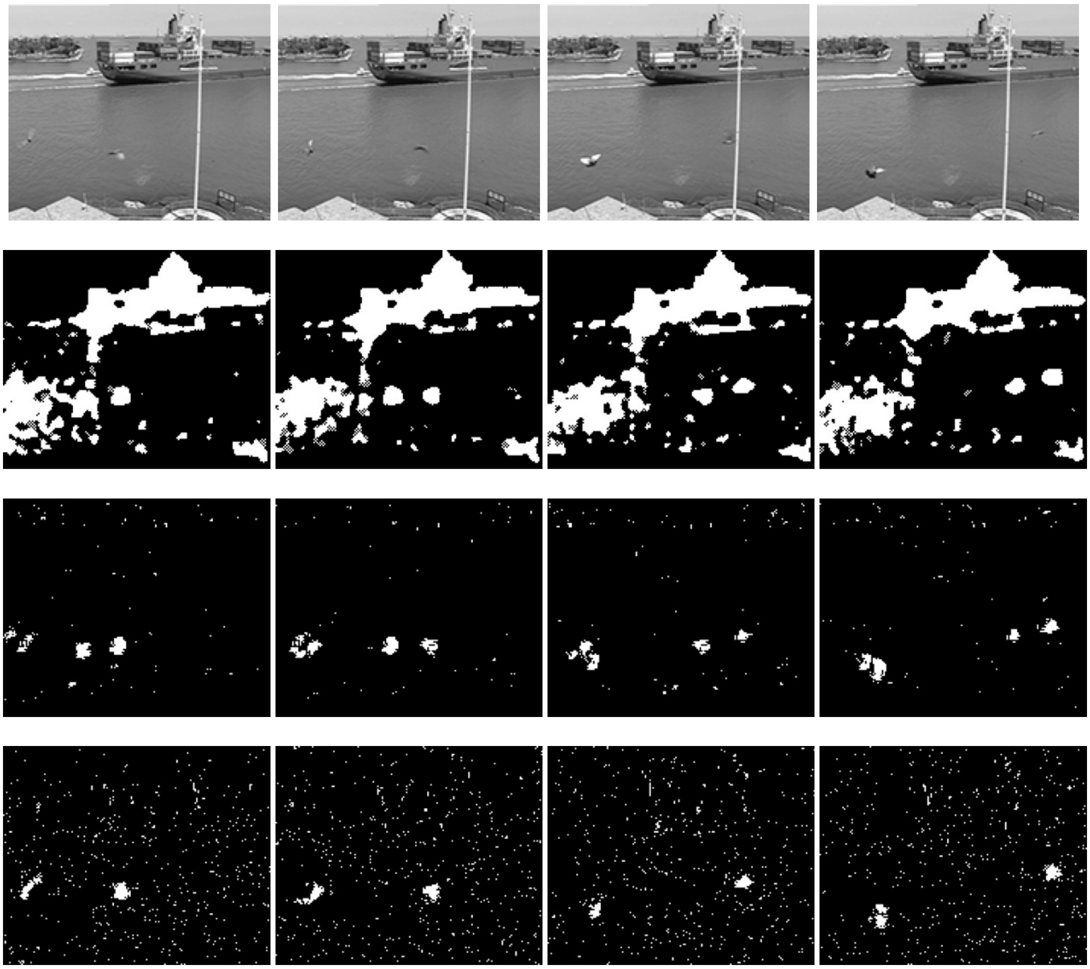


FIGURE 10: Experimental results on test sequence *Container*. From top to bottom: frames 252, 255, 258, and 261; CDMs detected by the MRF method, by the QPF method, and by the M3 method.

In the following, the comparisons with QPF and M3 methods are reported. Several representative change detection results on *Miss America*, *Container*, *Table Tennis*, and *News* are shown in Figures 9–12. In the selected frames of *Miss America* (Figure 9), the subject's head and body were moving to her left. It can be observed that the CDMs detected by the MRF approach reflected this motion, where changes in the face region were very well detected. The results from QPF captured most of the changes; however, the disturbance from noise appeared in the background area. The CDMs detected by M3 had certain errors and also suffered from noise.

The results on the container sequence, a typical outdoor video, are presented in Figure 10. In the sample frames, the container was moving slowly to the right and two birds flew by quickly from the left to the right. It can be seen that all the three methods captured the changes caused by the flying birds. However, the motions of the container and rippling water were only well identified by the MRF method, which shows that the proposed method is more efficient in detecting small changes than the other two methods.

Figure 11 demonstrates the results on the table tennis sequence, which contains very fast motion. Again, the MRF method detected moving regions more completely than the other two methods. The scenes selected in news sequence contain both small motion (e.g., face of the male journalist) and large motion (e.g., the spinning stage and dancers). It can be seen in Figure 12 that, although all three methods were robust against background noise, the MRF approach was superior to the other two methods in detecting more completely changing regions, including both the journalists and the moving stage and dancers.

All the algorithms were implemented in C++ and compiled with Microsoft Visual C++6.0. Experiments were performed on an AMD Athlon 1900 (1.66 GHz) PC with 512 M DDR2100 RAM. Among the three methods implemented, the M3 has the least computational complexity. The MRF requires iterations to compute the mean field, thus is slower than M3. The QPF required the most computation in all the experiments. In Table 2, the average computing time of each method on each testing sequence is listed. The computing time of the QPF and M3 is determined by the number of

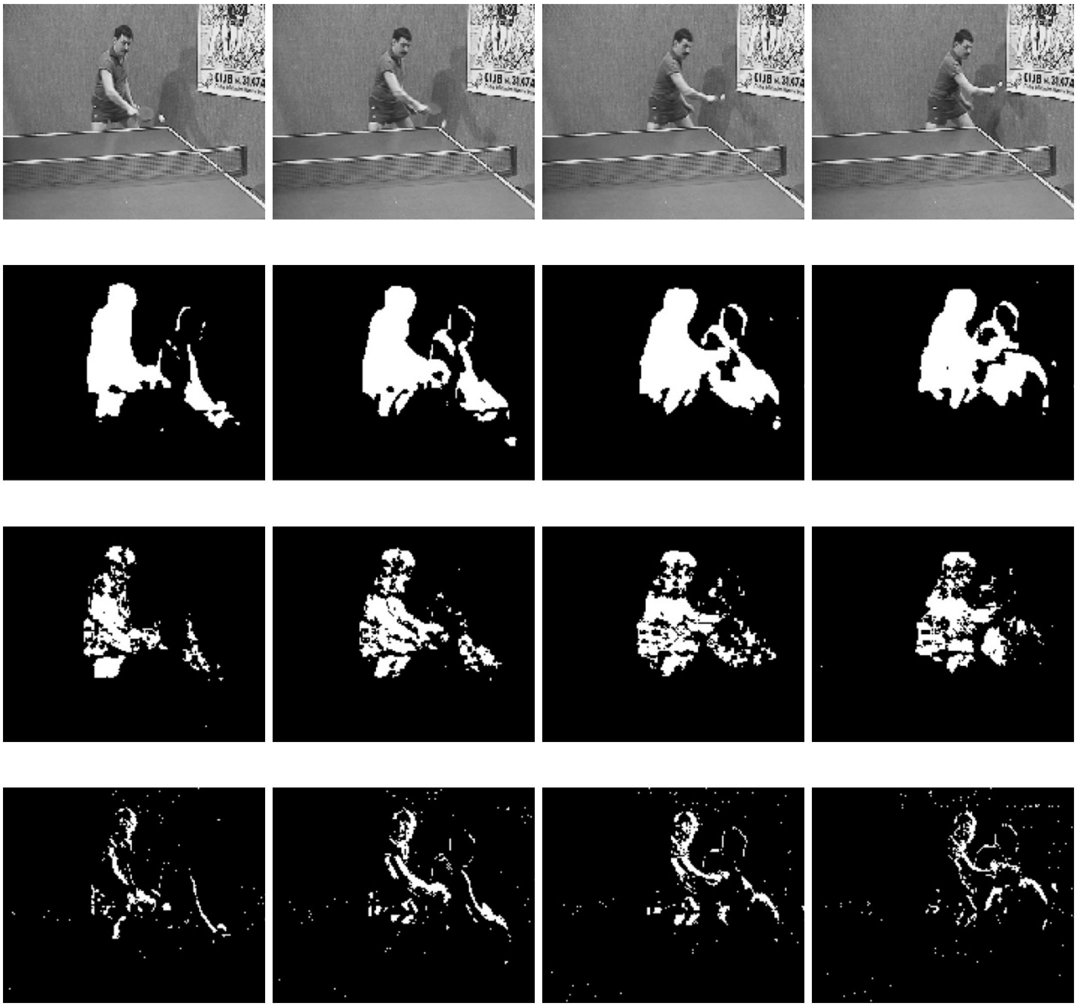


Figure 11: Experimental results of the test sequence *Table Tennis*. From top to bottom: frames 132, 135, 138, and 141; CDMs detected by the MRF method, by the QPF method, and by the M3 method.

pixels contained in a video frame, and therefore is largely fixed for all the testing sequences. The computing time of the MRF depends not only on the spatial resolution, but also the number of iterations taken to compute the mean field values. In practice, if the time of computation is critical, a maximum number of iterations can be specified. For example, our system required an average of 9.4 milliseconds per iteration, so a maximum number of iterations of 10 was utilized in detecting changes in image sequence at 10 frames per second.

5. CONCLUSION

In this paper, we have presented a new approach to the change detection problem in image sequences. This approach employs two well-established theories: MRF and MFT. Based upon the MRF theory, change detection is modeled as an optimization problem, namely, the CDM is calculated in the sense of MAP. Our approach differs from the previous statistical methods which rely on thresholding. In order to carry out an efficient computation, we utilized MFT, which simplifies the procedure of searching for the optimal detection of CDM. Experimental results are reported based on this optimization approach. Both the synthetic and real-world data indicate that this approach accurately detects changes between frame pairs. One remaining problem, however, is to determine the values of control parameters in the associated functions. Currently, the parameters are chosen in an experimental manner. In the future, a meaningful cost function of these parameters may be designed to provide the values in a certain optimal sense.

ACKNOWLEDGMENT

This work was supported in part by the National Institutes of Health, Grants nos. NS/MH38494 and EB002309, and by the Computational Diagnostics, Inc.

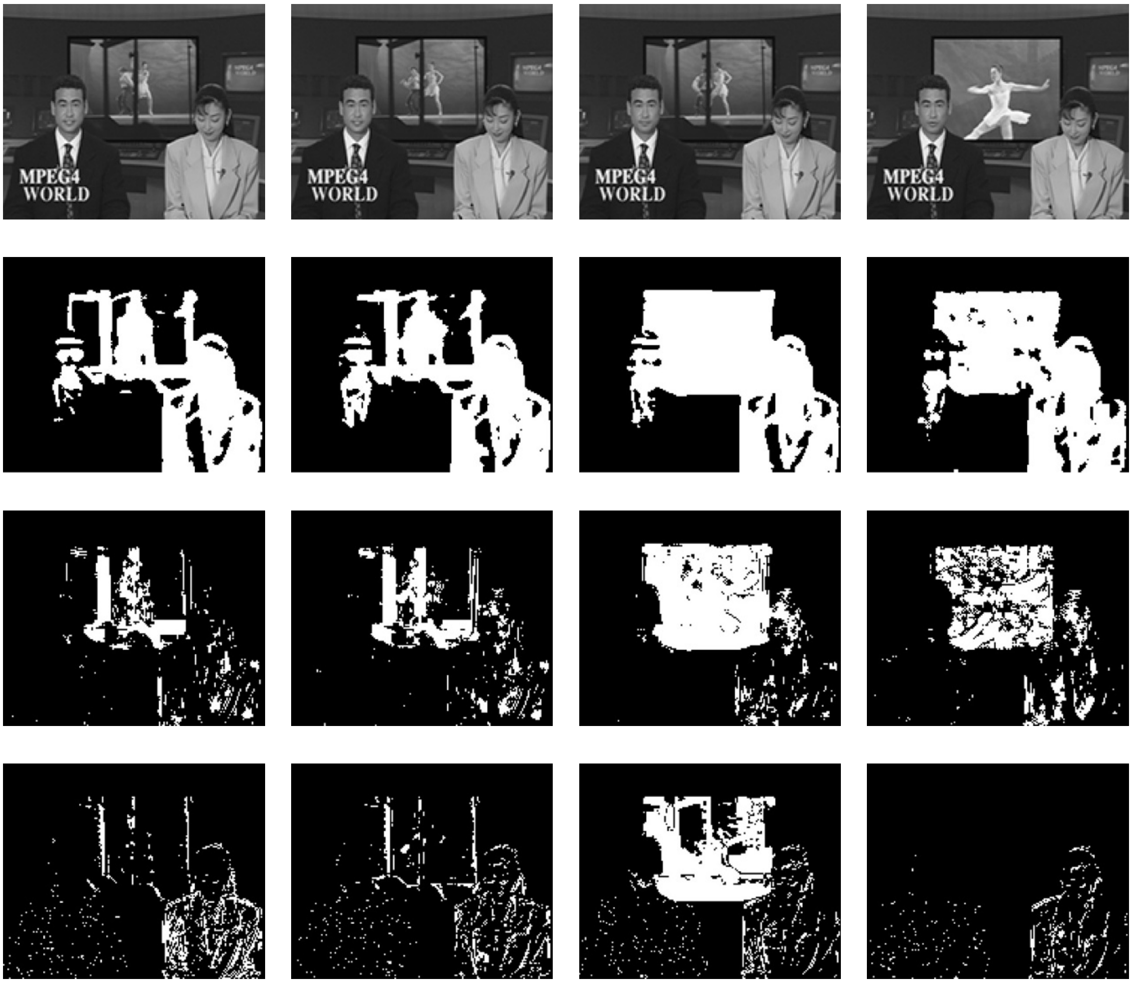


Figure 12: Experimental results on *News*. From top to bottom: frames 84, 87, 90, and 93; CDMs detected by the MRF method, by the QPF method, and by the M3 method.

Table 2: Computational cost of MRF, QPF, and M3 methods.

Test sequence	MRF (average loops/time)	QPF (time)	M3 (time)
Miss America	4.31 loops/40.51 ms	147.2 ms	5.56 ms
Container	6.49 loops/61.0 ms	147.2 ms	5.56 ms
Table Tennis	5.37 loops/50.48 ms	147.2 ms	5.56 ms
News	3.93 loops/36.94 ms	147.2 ms	5.56 ms

REFERENCES

[1] *MPEG-4 Video Verification Model Version 15.0*, ISO/IEC JTC1/SC29/WG11 N3093,1999.

[2] T. Sikora, "The MPEG-7 visual standard for content description an overview," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 696–702, 2001.

[3] T. Meier and K. N. Ngan, "Segmentation and tracking of moving objects for content-based video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1190–1203, 1999.

[4] M. Kim, J. G. Choi, D. Kim, et al., "A VOP generation tool: automatic segmentation of moving objects in image sequences based on spatio-temporal information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1216–1226, 1999.

[5] K. Skifstad and R. Jain, "Illumination independent change detection for real world image sequences," *Computer Vision, Graphics and Image Processing*, vol. 46, no. 3, pp. 387–399, 1989.

[6] Y. Z. Hsu, H. H. Nagel, and G. Rekers, "New likelihood test methods for change detection in image sequences," *Computer Vision, Graphics and Image Processing*, vol. 26, no. 1, pp. 73–106, 1984.

[7] T. Aach, A. Kaup, and R. Mester, "Statistical model-based change detection in moving video," *Signal Processing*, vol. 31, no. 2, pp. 165–180, 1993.

[8] E. Durucan and T. Ebrahimi, "Change detection and background extraction by linear algebra," *Proc. IEEE*, vol. 89, no. 10, pp. 1368–1381, 2001.

[9] T. Aach and A. Kaup, "Bayesian algorithms for adaptive change detection in image sequences using Markov random fields," *Signal Processing: Image Communication*, vol. 7, no. 2, pp. 147–160, 1995.

[10] R. Chellappa and A. Jain, *Markov Random Fields Theory and Applications*, Academic Press, Boston, Mass, USA, 1993.

[11] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 6, pp. 721–741, 1984.

[12] S. Kirkpatrick, C. D. Gellatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[13] J. E. Besag, "On the statistical analysis of dirty pictures (with discussion)," *Journal of the Royal Statistical Society. Series B*, vol. 48, no. 3, pp. 259–302, 1986.

[14] J. Zhang, "The mean field theory in EM procedures for blind Markov random field image restoration," *IEEE Trans. Image Processing*, vol. 2, no. 1, pp. 27–40, 1993.

[15] J. Wei and Z.-N. Li, "An efficient two-pass MAP-MRF algorithm for motion estimation based on mean field theory," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 6, pp. 960–972, 1999.

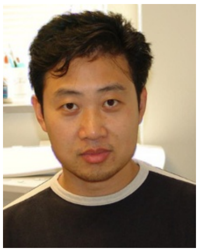
[16] M. Hassner and J. Sklansky, "The use of Markov random fields as models of texture," *Computer Graphics and Image Processing*, vol. 12, no. 4, pp. 357–370, 1980.

[17] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, New York, NY, USA, 1987.

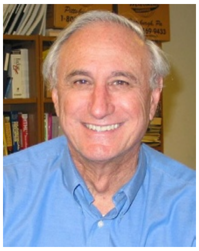
[18] L. Bruzzone and D. F. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 452–466, 2002.

[19] J. Zhang and G. G. Hanauer, "The application of mean field theory to image motion estimation," *IEEE Trans. Image Processing*, vol. 4, no. 1, pp. 19–33, 1995.

[20] M. De Geyter and W. Philips, "A noise robust method for change detection," in *Proc. IEEE International Conference on Image Processing (ICIP '03)*, vol. 2, pp. 391–394, Barcelona, Spain, September 2003.

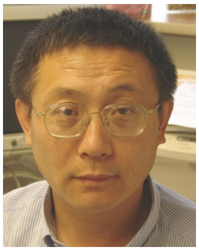


Qiang Liu received his B.S. and M.S. degrees in telecommunications from Xidian University, Xian, China, in 1996 and 1999, respectively. His research experience includes hardware and software designs in the State Key Lab on ISDN, Xidian University, Xian, China, and ZTE telecommunication corporation, Shanghai, China. He is currently a Ph.D. student at the University of Pittsburgh, Pittsburgh, USA. His recent research has been on image representation, object-based video coding, biomedical signal processing, medical imaging, image understanding, and computer vision.



Robert J. Sclabassi is currently a Professor of neurological surgery, electrical engineering, neuroscience, mechanical engineering, psychiatry, and biomedical engineering at the University of Pittsburgh. He completed his undergraduate education at Loyola University of Los Angeles and earned his Masters and Engineers degrees in electrical engineering at the University of Southern California. He was employed at TRW in a new product development for approximately 8 years and is a professional electrical engineer. He received a Ph.D. degree in electrical engineering from the University of Southern California in 1971 and then was a postdoctoral scholar studying basic and clinical neurophysiology at the Brain Research Institution, University of California, Los Angeles, before joining the faculty at that institution in the Departments of Neurology and Biomathemtics. In 1981, he earned an M.D. degree from the University of Pittsburgh. He has published over 330 articles and book chapters and 150 abstracts.

Ching-Chung Li received the B.S.E.E. degree from the National Taiwan University, Taipei, in 1954, and the M.S.E.E. and Ph.D. degrees from the Northwestern University, Evanston, Ill, in 1956 and 1961, respectively. He is presently a Professor of electrical engineering and computer science at the University of Pittsburgh, Pittsburgh, Pa. He was a Visiting Professor of electrical engineering at the University of California, Berkeley, in the spring of 1964, and a Visiting Principal Scientist at the Biodynamics Laboratory, Alza Corporation, Palo Alto, Calif, in the summer of 1970. On his sabbatical leaves, he was with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, in the fall of 1988, and with the Robotics Institute, Carnegie-Mellon University, in the spring of 1998. His research interests are in pattern recognition, image processing, biocybernetics, and applications of wavelet transforms. He is a Fellow of the IEEE.



Mingui Sun received a B.S. degree from the Shenyang Chemical Engineering Institute, China, in 1982, and the M.S. and Ph.D. degrees in electrical engineering from the University of Pittsburgh in 1986 and 1989, respectively. He was a graduate student researcher from 1985 to 1989 working on signal and image processing projects. Currently, he is an Associate Professor and an Associate Director of the Center for Clinical Neurophysiology, Department of Neurological Surgery, University of Pittsburgh, and a Director of Research at Computational Diagnostics, Inc. His current research and development interests include advanced biomedical electronic devices, biomedical signal and image processing, sensors and transducers, biomedical instruments, artificial neural networks, wavelet transforms, time-frequency analysis, and the inverse problem of neurophysiological signals. He has over 160 publications in these areas.