A DCT-DOMAIN APPROACH TO IMAGE CHANGE DETECTION AND ITS APPLICATION TO PATIENT VIDEO MONITORING

Qiang Liu1,2, Robert J. Sclabassi1,2,3, and Mingui Sun1,2,3
Laboratory for Computational Neuroscience
Departments of 1Neurological Surgery, 2Electrical Engineering, and 3 Bioengineering
University of Pittsburgh, Pittsburgh, PA 15261, USA
E-mail: qlin@neuronet.pitt.edu

Abstract—Change detection is a useful tool to identify content changes in images. Its significance is found in the applications of video surveillance, editing, coding and analysis. In this work, we present another application of change detection to patient video monitoring, based on a novel approach in the discreet cosine transform (DCT) domain. This approach is built upon a statistical test on the difference between DCT-blocks, where the noise disturbances and illumination variations are excluded from the meaningful changes. Our experimental results show that the proposed method is robust in detecting content changes and may be used to facilitate high-level segmentation tasks.

I. INTRODUCTION

Change detection performs preliminary tasks in video segmentation, where the pixels in one or more video frames are classified into two sets, “changed” and “unchanged”. The former indicates that “there are significant differences between the test image and the reference image(s) at the pixel locations”, and the latter denotes the opposite. The definition of “significant” is largely associated with human visual perception and may vary from application to application. In common cases, image differences caused by relative motion between objects and camera, appearance/disappearance of objects, and shape, color, and texture changes of objects are considered to be “significant”, while those caused by ambient and sensor noise and illumination variation are “insignificant”.

The applications of change detection are broad [1], including video surveillance, remote sensing, object-based video coding, medical diagnosis, driving and traffic assistance, etc.. In this work, we discuss a medical application of patient monitoring via video recording, which is commonly used in hospitals for clinical diagnosis. Automatic detection of the dynamic changes in the monitoring video may facilitate this application in the following aspects: 1) detecting unusual changes and behavior in a scene so that an alarm can be generated in case of an abnormal event; 2) tracking the subject and controlling the camera automatically to present an optimal view of the subject; 3) organizing the video data according to motion activity to facilitate archiving and retrieval of video contents; and 4) compressing the video contents to reduce the bandwidth in data communication by encoding only the changing portion of a video frame.

Considerable research effort has been devoted to developing change detection algorithms [2]. These methods operate in the spatial domain testing whether a pixel or a region between two images are significantly different. However, many image processing tasks are carried out in the transform domain, such as discrete cosine transform (DCT) used by JPEG and MPEG. It would be useful that change detection could also be carried out in the transform domain than in the spatial domain. In this paper, we present a DCT domain approach to detect changes between image pairs.

II. METHODOLOGY

In most compression methods, images are decomposed into blocks containing $N \times N$ pixels which are then transformed into DCT coefficients. Considering a pair of corresponding blocks in two images, we model the intensity values by

$$x^{(k)} = I^{(k)}S^{(k)} + \theta^{(k)}, \quad k = 1, 2$$

where $k$ is the image index, $x^{(k)}$ is an $N \times N$ matrix denoting the intensities of a block, $I^{(k)}$ is a scalar representing illuminance, $S^{(k)}$ is an $N \times N$ matrix representing reflectance of the patch surface, and $\theta^{(k)}$ is an $N \times N$ matrix denoting noise. It should be noted that (1) incorporates the Shading Model of Phong [3] and assumes that the illuminance is uniformly distributed on the patch surface.

Let $X^{(k)}$ denote the DCT block, i.e. the DCT coefficients of a block, formulated by

$$X^{(k)} = M x^{(k)} M^T, \quad k = 1, 2$$

where $M$ is the $N \times N$ transformation matrix. Equivalently, we have

$$X^{(k)} = I^{(k)} M S^{(k)} M^T + M \theta^{(k)} M^T, \quad k = 1, 2.$$ (3)

Essentially, given $X^{(k)}$ the task of a change detection algorithm is to identify whether there is meaningful change between the two blocks. Usually, the hypothesis of “no change” can be interpreted as $S^{(1)} = S^{(2)}$, which means that the patch surface does not change between the two images. Therefore, change detection can be carried out by testing the
null hypothesis which states $S^{(1)} = S^{(2)}$. Notice that the variations of $I^{(k)}$ and $θ^{(k)}$ are considered to be irrelevant changes.

To test the null hypothesis, one needs to have knowledge of $I^{(k)}$ and $θ^{(k)}$. For the former, let us define

$$I^{(2)} = γI^{(1)}$$

(4)

where $γ$ denotes the ratio of the illumination between two blocks. Then, instead of comparing $X^{(1)}$ and $X^{(2)}$, we can examine $γX^{(1)}$ and $X^{(2)}$, where illumination variation is compensated. Now the question is how to obtain $γ$. An easy way is to use the input data to estimate it. To do that, let us consider the DC components of the two blocks, denoted by $X^{(k)}_{11}$, which can be formulated by

$$X^{(k)}_{11} = I^{(k)}M_1S^{(k)}M_1^T + M_1θ^{(k)}M_1^T, \quad k = 1, 2$$

(5)

where $M_1$ is the first row of $M$, and $M_1^T$ is its transverse. Let $η_{11}$ denote the second term on the right side of (5). Then, since the entries $M_{1j} = \frac{1}{\sqrt{N}}$, $j = 1, 2, ..., N$, one has

$$η_{11}^{(k)} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} θ_{ij}^{(k)}, \quad k = 1, 2$$

(6)

where $θ_{ij}^{(k)}$ denotes the $(i, j)$th entry of noise matrix $θ^{(k)}$. Assuming that $θ_{ij}^{(k)}$ is independently and identically distributed (i.i.d.) and has a Gaussian distribution with a zero mean and a variance $σ^2$, we have $η_{11}^{(k)}$ obeying the same distribution as $θ_{ij}$. Considering that, for natural images, $S^{(k)}$ has highly correlated entries, one may assume that $I^{(k)}M_1S^{(k)}M_1^T$ usually has a much larger value than that of $η_{11}$. Therefore, we have

$$X^{(k)}_{11} \approx I^{(k)}M_1S^{(k)}M_1^T, \quad k = 1, 2$$

(7)

As a result,

$$γ = \frac{I^{(2)}}{I^{(1)}} \approx \frac{X^{(2)}_{11}M_1S^{(1)}M_1^T}{X^{(1)}_{11}M_1S^{(2)}M_1^T}$$

(8)

Under the null hypothesis, i.e., $S^{(1)} = S^{(2)}$, we have

$$γ \approx \frac{X^{(2)}_{11}}{X^{(1)}_{11}}$$

(9)

Now let us define the difference between the pair of DCT blocks as

$$ξ = X^{(2)} - γX^{(1)}$$

$$= M(I^{(2)}S^{(2)} - γI^{(1)}S^{(1)} + θ^{(2)} - γθ^{(1)})M^T$$

(10)

Under the null hypothesis, one has

$$ξ = M(θ^{(2)} - γθ^{(1)})M^T = Mθ^{(2)}M^T$$

where $θ = θ^{(2)} - γθ^{(1)}$. With the assumption on $θ_{ij}^{(k)}$ stated previously, the entries of $θ$ are i.i.d. Gaussian random variables, with a mean of zero and a variance of $(1 + γ^2)σ^2$.

In order to perform the hypothesis test, we need to know the probability density function of the random variables of $ξ$ on condition of the null hypothesis. Let $ξ_{ij}$ denote the $(i, j)$th entry of $ξ$, we have

$$ξ_{ij} = M_iθ_{jk}M_j^T$$

(12)

where $M_i$ denotes the ith row of $M$ and $M_{ij}$ the $(i, j)$th entry of $M$. Note that $ξ_{ij}$ is a linear combination of i.i.d. Gaussian random variables, we have $ξ_{ij}$ obeying Gaussian distribution. Also, since $M$ is a unitary matrix, $ξ_{ij}$, $i, j \in \{1, 2, ..., N\}$ have the same mean and variance as $θ_{ik}, k, l \in \{1, 2, ..., N\}$, and they are independently distributed. Since the former is obvious, we only show the proof of the latter as follows

$$E(ξ_{ij}ξ_{gh}) = E(\sum_k M_{jk}M_{gh}θ_{lk}) \cdot \sum_s M_{gs}M_{ht}E(θ_{lk}θ_{st})$$

$$= \sum_k M_{jk}M_{lk}M_{gs}M_{ht}E(θ_{lk}θ_{st})$$

$$= \sum_k M_{hk}M_{gh}M_{il}M_{st}E(θ_{lk}θ_{ks})$$

(13)

where, if $j \neq h$ or $i \neq g$, then $\sum_k M_{jk}M_{hk} = 0$ or $\sum_i M_{il}M_{il} = 0$, therefore $E(ξ_{ij}ξ_{gh}) = 0$.

In brief, under the null hypothesis, $ξ_{ij}, i, j = 1, 2, ..., N$ are i.i.d. Gaussian random variables having the following conditional probability density function

$$p(ξ_{ij} | H_0) = \frac{1}{\sqrt{2\pi(1 + γ^2)σ^2}} e^{-\frac{ξ_{ij}^2}{2(1 + γ^2)σ^2}}$$

(14)

where $H_0$ denotes the null hypothesis.

With the previous results, the hypothesis test can be carried out as follows. First, define

$$y = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{ξ_{ij}^2}{(1 + γ^2)σ^2}$$

(15)

as the measure of of the difference between two given blocks. Since $\frac{ξ_{ij}^2}{(1 + γ^2)σ^2}$ are i.i.d. standard normal random variables, $y$ obeys a $χ^2$ distribution with $N^2$ degrees of freedom. Next, determine a threshold $τ$ so that when $y > τ$ the null hypothesis is rejected, otherwise established. Threshold $τ$ is usually obtained by specifying the significance level denoted by $α$, such that

$$α = P(y > τ | H_0)$$

(16)

where $P$ denotes probability.
III. EXPERIMENTAL RESULTS

In this section, experimental results based on the proposed method are reported. We tested the method on an MPEG reference sequence Hallway, a test sequence car toy recorded by a regular digital camera, and patient monitoring video recorded at the epilepsy monitoring unit at the University of Pittsburgh Medical Center. The Hallway sequence is in QCIF format (176 x 144 pixels in spatial dimension) and the other two sequences are in SIF (352 x 240 pixels). The size of a DCT block was 8 x 8. The significance level \( \alpha \) was set to \( 10^{-6} \) for all the testing sequences.

The Hallway sequence contains high level noise, where \( \sigma \) was approximately 2.95. Fig. 1 shows the results on Hallway sequence, where (a),(b), (c) and (d) are frame 1, frame 250, the CDM, and the masked image respectively. It is seen that the human subject and the suitcase (placed on the deck) were very well identified.

![Image of Hallway sequence](image)

In Fig. 2, the results on illumination varying image sequence are presented. The image sequence was taken indoor by SONY F828 digital camera with fixed shutter speed and aperture size. In order to test the robustness of the method against illumination changes, we varied light source intensity. The results on Car Toy sequence are illustrated in Fig. 2, where (a),(b),(c) and (d) show the reference frame, the changed frame, detected CDM and masked image respectively. The lighting conditions of (a) and (b) were very different as shown by (c), which provides the histogram of the intensity difference between the two frames in the “unchanged” area (the black regions in the CDM). It can be seen that the proposed method excluded this dramatic illumination variation reasonably well, while still detected the truly changed region (the toy car).

For the patient monitoring sequence, we show two examples. 1) detecting changes between consecutive video frames (shown in Fig. 3), and 2) detecting changes between a background image and a test video frame (shown in Fig. 4). In Fig. 3, the sample video frames show the patient in bed moving his hand, which was successfully identified by the change detection method. In this case, the illumination change was insignificant because of the short temporal duration between the test video frames. In the other case shown in Fig. 4, the content changes were also well detected between a reference frame containing only background content and a test frame with the subject in bed. The illumination changes in the background regions were excluded regardless of the significant intensity changes (up to 35 in intensity difference). Missing detections occurred within the foreground area, which was due to the intensity homogeneity in those regions. These missing detections may be eliminated by simple postprocessing (e.g. hole filling).

The computational complexity of the proposed method is very low. Given DCT blocks, each comparison requires 2 divisions, \( N^2 + 2 \) multiplications, \( N^2 \) additions, and 1 logical operation. Equivalently, that is \( \frac{2}{N^2} \) divisions, \( 1 + \frac{2}{N^2} \) multiplications, 1 additions, and \( \frac{1}{N^2} \) logical operations per
REFERENCES


Fig. 3. Results on patient monitoring sequence, where the content changes between consecutive video frames were detected.

Fig. 4. Results on patient monitoring sequence, where the content changes between a background image and a test video frame were detected reasonably well.

IV. DISCUSSION

In this work, we presented a change detection approach operating in the DCT domain. This approach was based upon hypothesis test, where device noise and illumination variation were modeled as disturbances. Experimental results show that this method is robust against noise as well as illumination change. However, the compression noise is not considered in the present model, which will be investigated in the future.

ACKNOWLEDGMENT

This research was supported by NIH grants No. R01EB002309 and R01NS38494.