Significance Evaluation of Video Data Over Media Cloud Based on Compressed Sensing

Jie Guo, Bin Song, *Member, IEEE*, and Xiaojiang Du, *Senior Member, IEEE*

***Abstract*—Given the varying communication environment between the media cloud and users, there is a need to ensure the most significant part of a video will be successfully transmitted. Although there exist some techniques to evaluate the significance of video data in traditional video coding methods, such as H.264, the evaluation algorithms are often simple and inaccurate. This paper presents a novel significance evaluation method for video data based on compressed sensing. Specifically, we propose a method to obtain a trained dictionary directly by using the measurements of the video data, and then keep the sparse components and generate a saliency map. Since the sparse components can reflect the essential parts of videos, we discuss how to analyze the area and distribution of salient regions. At last, we present a computing method that gives the degree of significance of a frame. Experimental results show that the proposed saliency map reflects the focus points of humans. The method can be used in the distribution of video data over "wireless" transmissions and provide good video quality to mobile users.**

***Index Terms*—Compressed sensing, media cloud, significance evaluation, video.**

I. Introduction

MOBILE video services have become an inseparable part of our daily lives because of the improvement of wireless transmission technologies, the popularity of smart mobile devices and the transition of the video services into business models [1]–[3]. For example, the number of videos played per day in YouTube (the world's largest video website) is as many as 100 millions. It should be noted that 60% of the video services are required to be online, and there are 65 000 new video clips every day. In addition, according to the research of several companies (e.g., Cisco Systems), mobile video services will become a dominating data service in the near future [2]. Therefore, one of the most important tasks for mobile multimedia is to provide users with good enough experience in mobile video services. Various techniques emerge, one of which is the media cloud.

Manuscript received September 01, 2015; accepted April 27, 2016. Date of publication May 05, 2016; date of current version June 15, 2016. This work was supported by the National Natural Science Foundation of China under Grant 61271173 and Grant 61372068, by the Research Fund for the Doctoral Program of Higher Education of China 20130203110005, by the Fundamental Research Funds for the Central Universities K5051301033, by the 111 Project B08038, by the ISN State Key Laboratory at Xidian University, and by the U.S. National Science Foundation under Grant CNS-1065444. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Honggang Wang.

J. Guo and B. Song are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: irene2010guojie@163.com; bsong@mail.xidian.edu.cn).

X. Du is with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122 USA (e-mail: dxj@ieee.org).

Digital Object Identifier 10.1109/TMM.2016.2564100

Media cloud is a cloud-based service suitable for media environment such as storage, sharing and management of high-quality video and serves as a middleware to provide coordination between service providers and consumers [4]. Media cloud reduces the implementation and maintenance costs of computing resources for services because it provides video service providers with specially designed infrastructure and management of services. Recently, media cloud has become a hot research topic. In [5], the authors propose a novel video transcoding method for media cloud. [6] presents a secure and efficient virtualization framework for media service. [7] proposes an optimal transcoding and caching method for adaptive streaming in media cloud.

However, few existing works pay attention to the different degree of significance of video streaming and the inner property of the original video data. Compared with other data services, a major characteristic of video services is the need to provide a relatively continuous high speed for transmitting a large amount of data in a certain period [8]–[10]. Therefore, it is usually a difficult task to guarantee the real time performance of mobile video services [11], [12]. For example, the wireless network connection often goes bad when a user moves from outdoor to indoor. When a user is on a high-speed train or car, the wireless network changes rapidly and the connection is not guaranteed. The user will suffer bad experience if the current video service remains unchanged.

On the other hand, compared to traditional video services, the devices used by mobile video users are diverse, including the size and resolution of the device. What's more, the channel status and the moving speed of users are also different from each other. As a consequence, it is a challenging task as how to provide various users with satisfying video experiences even in poor communication environment. In this work, our goal is to evaluate the significance of different video frames. In specific, we analyze the saliency map of frames based on compressed sensing (CS) and we propose a method to measure the significance of a frame. By using the proposed method, we can transmit videos according to wireless channel status, e.g., only transmitting the significant frames when there is limited wireless connection. Our proposed method can provide users with better experience when the wireless connection is not good.

Although there are some techniques to evaluate the significance of video data in traditional video coding methods, such as H. 264, the evaluation algorithms are often simple and inaccurate. Meanwhile, nowadays with the development of CS, there is an emerging encoder based on CS [13], which transforms the original pixel domain images into the measurements and combines the sampling and compression into

TABLE I
NOTATION TABLE

y	measurements	Φ	the measurement matrix
M	the row number of Φ	N	the column number of Φ
A	the projection matrix	x	the original signal
D	the trained dictionary	f_b	the recovered saliency block
x_b	sparse coefficient vector	Ψ	sparse basis
S	training sample	$D^{(0)}$	the initial dictionary
S_c	the c column of S	T_0	preset sparse criterion
$E_k^{(t)}$	representation error	q	coefficient matrix
α	positive weight coefficients	β	positive weight coefficients
γ	positive weight coefficients	δ_P	isometry constant
N_s	number of salient blocks	Q_s	ratio of $var(I_r)$ and $var(I_c)$
Q_c	change between neighbour frames	I_s	saliency map

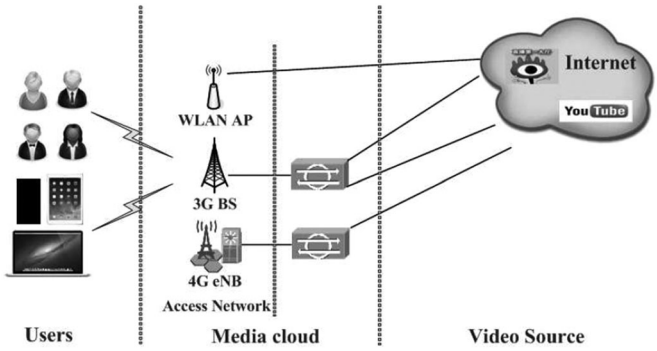


Fig. 1. Media cloud system.

one single step. Since the measurements have the information-preserving and dimensional-deduction property, this kind of encoder can have a low complexity and less output compared to the traditional encoder. Therefore it can alleviate the network congestion. However, after compressive sampling, what we only have is the measurements, it is a challenge to analyze the significance of the original videos. To solve this issue, in this paper, we propose a novel method that evaluates the significance of video frames based on CS. Specifically, we train a dictionary by using measurements and then remove the redundancy part in a frame, then we propose a calculation method to evaluation the significance degree of the frames, thus the uppermost part of the videos can be guaranteed during transmission between the media cloud and users.

This paper is organized as follows. Section II gives a brief overview of media cloud and the sparsity analysis based on CS theory. Section III presents the proposed significance evaluation method. Section IV discusses the experimental results, followed by conclusions in Section V.

II. BACKGROUND

First we give the notation table of this paper as shown in Table I.

A. Media Cloud

Media cloud is an emerging paradigm which can efficiently distribute video streaming services and has attracted significant attentions. Fig. 1 shows a systematic view of delivering video streaming over media cloud. It includes three parts: video source from the content providers, media cloud service providers, and end-users with different devices. The video source comes from service providers and are transmitted through the media cloud to users. Users receive media services by using various devices. Compared to traditional video services, the devices used by mobile video users are diverse, such as different OS (operating system), size and resolution of the devices. What's more, the wireless channel status and the moving speed of users are different from each other.

Most existing literatures study the framework or the scheduling strategies of video streaming, largely ignoring the video content analysis. In this paper, we propose a novel idea: videos are divided according to their degrees of significance, and then the significant frames are guaranteed during transmission with high priority. This will provide much better video experiences for the users.

B. Sparsity Analysis Based on CS

Saliency mechanism has been considered crucial in the human visual system and is also helpful to object detection and recognition. Meanwhile, in terms of the sparse theory, the salient parts are the sparse components of a frame. This means that we can get the saliency map by obtaining the sparse components of an image. Now we describe the CS theory. The CS theory claims that the signal f could be sampled using the following linear random projections:

$$y = \Phi f \tag{1}$$

where $y \in R^M$ is the sampled measurement, $\Phi \in R^{M \times N}$ is the measurement matrix, $M < N$, and the ratio between the height and width of the measurement matrix is defined as the measurement rate (MR), i.e.

$$MR = \frac{M}{N}. \tag{2}$$

Thus the recovery of the sparse coefficients (with respect to orthogonal basis $\Psi \in R^{N \times N}$) can be done by finding the set of coefficients that agrees with the measurements, and especially, with the minimum l_0 norm, i.e.

$$\min ||x||_0 \quad \text{s.t.} \quad y = \Phi\Psi x. \tag{3}$$

Intractable as the problem is NP-hard for typical values of N, it is still solvable if the product of Φ and Ψ, denoting $A = \Phi\Psi$, obeys the Restricted Isometry Property of order [14], that is, $(1 - \delta_P)||s||_2^2 \leq ||As||_2^2 \leq (1 + \delta_P)||s||_2^2$, holds for all P-sparse vectors s for a small "isometry" constant $0 < \delta_P < 1$. Then the signal can be recovered by solving the following unconstrained optimization problem (4):

$$\min ||x||_1 \quad \text{s.t.} \quad y = \Phi\Psi x. \tag{4}$$

This convex optimization problem, namely basis pursuit (BP), can be recasted as a linear program problem, which can be efficiently solved with the available optimization algorithms. However, the complexity of (4) still makes it intractable for many practical applications. Hence, in the literature, many iterative greedy techniques have been proposed to solve the above problem, e.g., the matching pursuit (MP) method. It has been proven

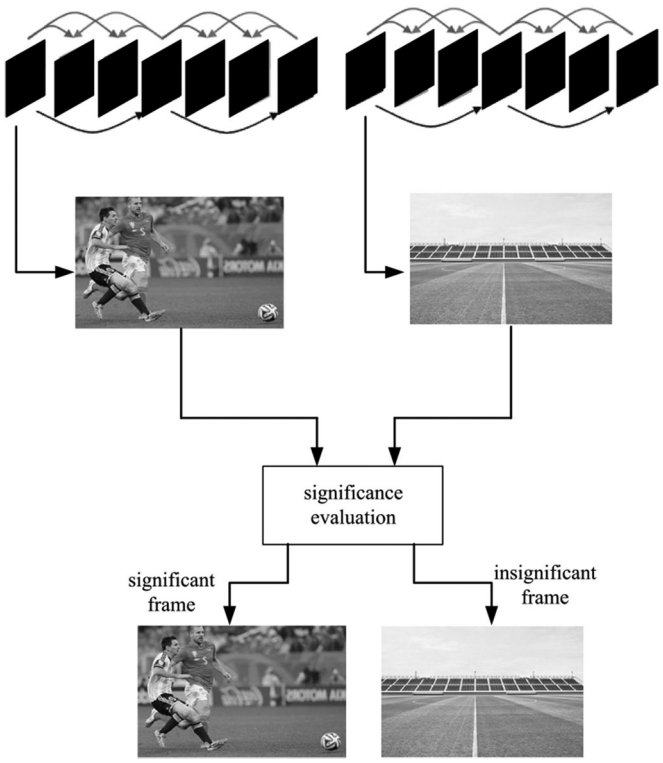


Fig. 2. Significance division of the encoded frames.

that MP could successfully reconstruct the compressively sampled signal with high probability. Other greedy algorithms such as orthogonal matching pursuit, stage-wise orthogonal matching pursuit, and subspace pursuit have also been shown to attain similar guarantees to those of their optimization-based counterparts. Besides, it should be noted that many signals of interest in practice are often "approximately" sparse rather than "exactly" sparse, i.e., the transforming coefficients are generally close to zero and only few of them have significant values. In this case, the solution to (4) could still reconstruct the most sparse coefficients as revealed in the CS theory [14], [15], and hence, provide a good approximation of the original signal.

By using CS to analyze the sparse property, we can extract the sparse components of an image and remove the redundant parts in order to clearly observe the essential parts of the frames.

III. The Proposed Method

In this section, we propose a significance evaluation method for video data based on CS. Since only measurements are available after compressive sampling, we should analyze the significance of videos directly in the measurement domain. As shown in Fig. 2, we extract the coded frames based on CS and then divide them into significant frames and insignificant frames. The first step is to obtain a dictionary D by using measurements, then we remove the redundancy part based on D. The last step is to evaluate the significance of the videos. The details are given below.

A. Saliency Map Based on CS

From the view of cognitive science, the image information-$Info(Image)$ can be decomposed into two parts

$$Info(Image) = Info(Redundancy) + Info(Saliency) \tag{5}$$

where $Info(Redundancy)$ denotes the information with high regularities, and $Info(Saliency)$ represents the novel part [16], [26]. Based on this observation, we can decompose an image into salient regions and redundant regions. It is also shown that the redundancy part should have approximately to be low-rank and the salient part is equivalent to the sparse elements to some extent. Meanwhile, if a frame include more saliency regions, it will attract us more. Therefore, by analyzing salient regions inside a frame, we can identify the important frames, which can be guaranteed during transmission with high priority. This method will provide users with better video experiences.

In (3), the sparse basis Ψ can also be obtained by training, which is often defined as the dictionary learning process. Compared to the common orthogonal basis, the trained dictionary gives a more sparse representation of the original signal. In this paper, we use the K-singular value decomposition (K-SVD) algorithm [18] to train the dictionary D. However, the K-SVD algorithm is quite different from the traditional one. Since we have the measurements instead of original signals after compressive sampling, we adopt measurements to train the dictionary D. Note that the benefits of using distinct measurement matrix for each image block can be understood intuitively [17]. Therefore, while sampling, we adopt different Gaussian measurement matrix for each image block.

Recall the process of K-SVD algorithm. It initializes with a (often randomized) dictionary $D^{(0)}$ by iterating between two stages: a sparse coding stage and a dictionary update stage [18]. Specifically, for a fixed estimate of the dictionary $D^{(t-1)}$ at the start of iteration $t \geq 1$, the sparse coding stage in K-SVD involves solving the sparse coefficient $x^{(t)}$ as follows:

$$\forall s, {x_s}^{(t)} = \arg\min_{x \in R^K} \left\| S_c - D^{(t-1)} x \right\|_2^2 \text{ such that } \|x\|_0 \leq T_0 \tag{6}$$

where S_c denotes the cth column of the training sample S, T_0 is the pre-set sparse criterion [18]. Note that while (6) in its stated form has combinatorial complexity, it can be easily solved by either convexifying (6) [19] or using greedy pursuit algorithms [20].

After the sparse coding stage, K-SVD fixes $X^{(t)}$ and moves to the dictionary update stage. The main novelty in K-SVD lies in the manner in which it carries out dictionary update, which involves iterating through the K atoms of $D^{(t-1)}$ and individually updating the kth atom, $k = 1, 2, ..., K$, as follows:

$$\begin{aligned} d_k^{(t)} &= \arg\min_{d \in R^n} \left\| \left(S - \sum_{j=1}^{k-1} d_j^{(t)} x_{j,T}^{(t)} - \sum_{j=1}^{k-1} d_j^{(t-1)} x_{j,T}^{(t)} \right) - d x_{k,T}^{(t)} \right\|_F^2 \\ &= \arg\min_{d \in R^n} \left\| E_k^{(t)} - d x_{k,T}^{(t)} \right\|_F^2 . \end{aligned} \tag{7}$$

Here, $E_k^{(t)}$ is the representation error for F using first $K-1$ atoms of $D^{(t)}$ and last $k+1, \ldots, K$ atoms of $D^{(t-1)}$. In order to simplify computations, K-SVD in [18] further defines an or-

dered set $\omega_k{}^{(t)} = \{s : 1 \leq s \leq S, x_{k,T}^{(t)}(s) \neq 0\}$ where $x_{k,T}^{(t)}(s)$ denotes the sth element of $x_{k,T}^{(t)}$, and an $S \times |\omega_k^{(t)}|$ binary matrix $\Omega_k^{(t)}$ that has ones in $(\omega^{(t)}(s), s)$ locations and zeros everywhere else. Then, defining $E_{k,R}^{(t)} = E_k^{(t)}\Omega_k^{(t)}$ and $x_{k,R}^{(t)} = x_{k,T}^{(t)}\Omega_k^{(t)}$, it is easy to see from (7) that

$$d_k^{(t)} = \arg\min_{d \in R^n} \left\| E_{k,R}^{(t)} - d x_{k,R}^{(t)} \right\|_F^2 \tag{8}$$

where $\|\cdot\|_F$ is the Frobenius norm. Solving (8) is equivalent to finding the best rank-one approximation of $E_{k,R}^{(t)}$. The algorithm then moves to the sparse coding stage and continues alternating between the two stages till a stopping criterion (e.g., a prescribed representation error) is reached.

For the proposed dictionary learning method by using measurements, different measurement matrix Φ_i is adopted for the block i. Then the above optimization problem can be changed into the following formula. That is, the goal is to find the trained dictionary D in

$$\begin{aligned} \min \|Y - \Phi_i D q\|_F^2 &= \min \left\| (Y - \sum_{j \neq k} \Phi_i d^{(j)} \alpha^j) - \Phi_i d^{(k)} \alpha^k \right\|_F^2 \\ &= \min \left\| R_k - \Phi_i d^{(k)} \alpha^k \right\|_F^2 \end{aligned} \tag{9}$$

where Y stands for the measurement samples obtained by the compressive sampling upon the original signals, $d^{(j)}$ is the jth dictionary atom, α^j is the corresponding coefficients for it for each signal, and $R_k = y - \sum_{j \neq k} \Phi_i d^{(j)} \alpha^j$ is the representation error for training signals when the kth dictionary atom is removed. In the dictionary update process, it is assumed at each step K, $d^{(j)}$, α^j are fixed. We then minimize the criterion over $d^{(k)}$ and α^k which is equivalent to finding the best rank-1 approximation of R_k. The minimizer might typically be obtained by applying SVD to the matrix R_k, but the strict sparsity constraint also must be considered. Therefore, in the K-SVD algorithm, we shrink the matrix R_k by eliminating columns corresponding to those training signals. This strategy preserves the support of the coefficient matrix q. Indeed, joint optimization of the dictionary atoms and the corresponding coefficients in the dictionary update step leads to a much more efficient minimization compared to other dictionary learning methods. Another method to solve (9) can be found in [21], which decompose the dictionary into the fixed part and the atom representation part. It is called compressive K-SVD.

Note that in this paper, the mentioned images are all split into blocks in order to reduce the computation complexity since the amount of pixels in one frame can be very large [22]. Meanwhile, S_c in (6) is obtained by extracting and then vectorizing a block in the neighbour frame (arranging the pixels in a block row by row in order to turn the two-dimensional (2D) matrix block to be a 1D vector). After getting the trained dictionary D, we use BP algorithm to obtain the sparse coefficient x_b of each image block. Then, we keep the sparse components and set the redundant regions to be black area. In specific, we remove the relatively larger absolute value of sparse coefficents, and then

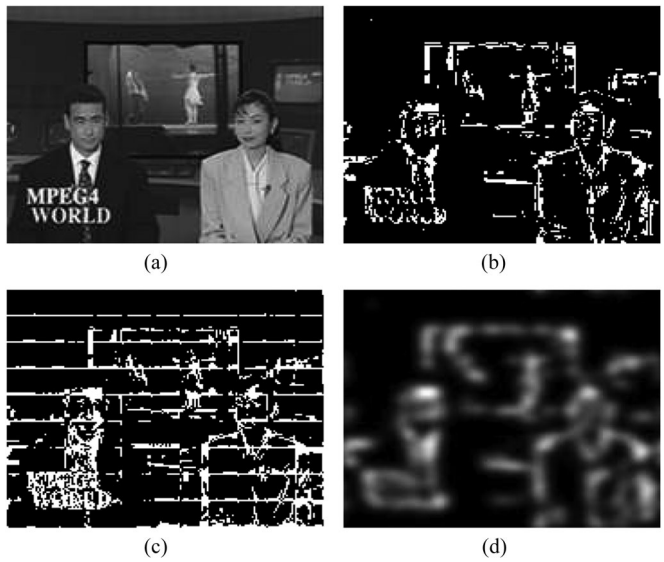


Fig. 3. Saliency map based on CS. (a) Original image. (b) Using the dictionary learning by K-SVD method. (c) Using DCT as the sparse basis. (d) Filtering result of the image in (b).

get the saliency map of a block by using

$$f_b = D * x_b \tag{10}$$

where D is a trained dictionary, x_b is the corresponding coeffficents, f_b denotes the recovered saliency block. Then all f_b are regrouped to be the saliency map. This saliency map generation is quite different from the method proposed in [16]. We define the redundancy as the most frequently arise components in D which is inspired by the frequency division effect of discrete cosine transform (DCT).

Here, the video sequence News QCIF (quarter common intermediate format, 176×144) is taken as an example.

As shown in Fig. 3, the saliency map [in Fig. 3(b)] obtained by the dictionary learning method based on CS outperforms the method [in Fig. 3(c)] by using DCT basis since it overcomes the blocking artifact.

B. Determining Significant Frames

After obtaining the saliency map, we can analyze the distribution of salient objects. It can be noted that frames that have a larger area of salient objects tend to be more important than the smaller ones [23]. Meanwhile, if the salient regions of a frame have a more disperse distribution, it means that it has more edges and outlines so it has much more information. In addition, for video sequences, the changes between frames should also be considered [24].

Therefore, we define the degree of significance of a frame from three aspects: 1) the area of salient regions; 2) the distribution of the salient objects; and 3) the difference of 1) as well as 2) between the current frame and its former one.

In specific, after getting the saliency map, we set a threshold to classify saliency blocks which have more salient information (here we use the threshold 20 by multiple experiments), then

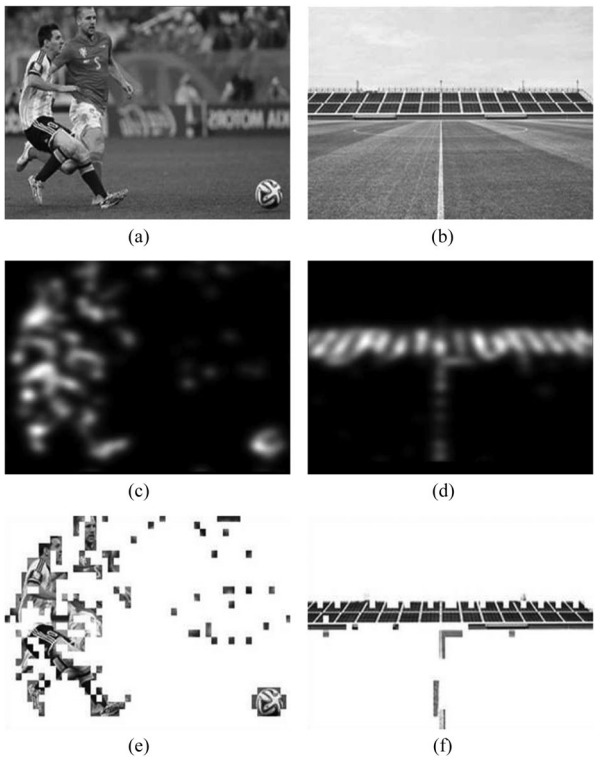


Fig. 4. Saliency map based on CS of Frame 1 and Frame 2: (a), (b) original image; (c), (d) using the dictionary learning by K-SVD method; and (e), (f) salient blocks.

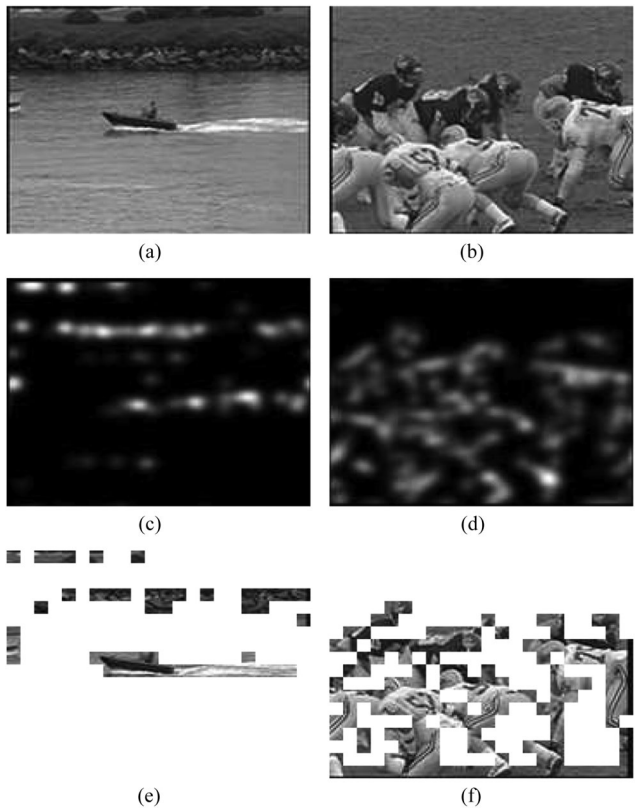


Fig. 5. Saliency map based on CS of Coastguard QCIF and Football QCIF: (a), (b) original image; (c), (d) using the proposed dictionary learning method; and (e), (f) salient blocks.

record the number of salient blocks and divided by the number of the whole block in a frame to obtain a normalized value denoted as N_s.

We denote the saliency image as I_s, which has p rows and q columns. By summing each column of I_s, we can get a row vector, denote as I_r. Similarly, we can get a column vector, denote as I_c. Then, we calculate their variances $var(I_r)$ and $var(I_c)$ respectively. The ratio of $var(I_r)$ and $var(I_c)$ is denoted as Q_s, namely $Q_s = \min(var(I_r), var(I_c))/\max(var(I_r), var(I_c))$. The larger value of Q_s means the larger variety degree of the salient regions.

In addition, the difference between the current frame and its former one is measured by the comparing N_s and Q_s, which is represented as Q_c. Obviously, Q_c measures the changes of N_s and Q_s between frames. Larger Q_c corresponds to a larger movement and changes compared with the former frame.

Finally, we calculate the significance degree of the frame i as

$$Q_i = \alpha N_s + \beta Q_s + \gamma Q_c \tag{11}$$

where α, β and γ are the positive weight coefficients. It is assumed that larger Q_i corresponds to higher significance degree of a frame. Obviously, larger value of Q_i means more salient blocks, varieties and changes in a frame.

IV. Experimental Results

In this part, we use different kinds of frames to evaluate our proposed method. Fig. 4 has two contrastive frames: one frame contains the sportsmen and a soccer which often catches attentions, denoted as Frame 1 and the other is a plain soccer field which has less information, denoted as Frame 2. Each image is tailored to the size of 400×240 and then split into 8×8 image blocks. Figs. 5 and 6 are the obtained images by using our proposed method of Coastguard and Football video sequences in the formats of QCIF (176×144) and CIF (common intermediate format, 352×288) respectively. Figs. 7 and 8 are the obtained images by using our proposed method of Hall and Foreman video sequences in the formats of QCIF and CIF respectively. Note that the video sequence Coastguard is relatively static compared with the sequence Football while Hall and Foreman are similar sequences since their backgrounds do not change.

As mentioned in Section III-A, by using the neighbour frames as the training samples, we obtain the dictionaries of Frame 1 and Frame 2, denoted as D_1 and D_2 respectively. Since we split each frame into blocks, we can calculate the sparse coefficient of each block by using the BP algorithm and keep the sparse part of the frame. The experimental results are shown in Fig. 4.

As shown in Fig. 4, an image with a larger area of salient objects tends to be more significant than that with a smaller area. Also, if the salient regions of an image have a more disperse distribution, it means that it has much more information.

By using (11), the significance degree of the Frame 1 is: $Q_1 = N_{s1} + Q_{s1} + Q_{c1} = 30/1500 + 0.5 + 0.22 = 0.74$ (All the data are normalized to 1); and the significance degree of the Frame 2 is, $Q_2 = N_{s2} + Q_{s2} + Q_{c2} = 24/1500 + 0.005 +$

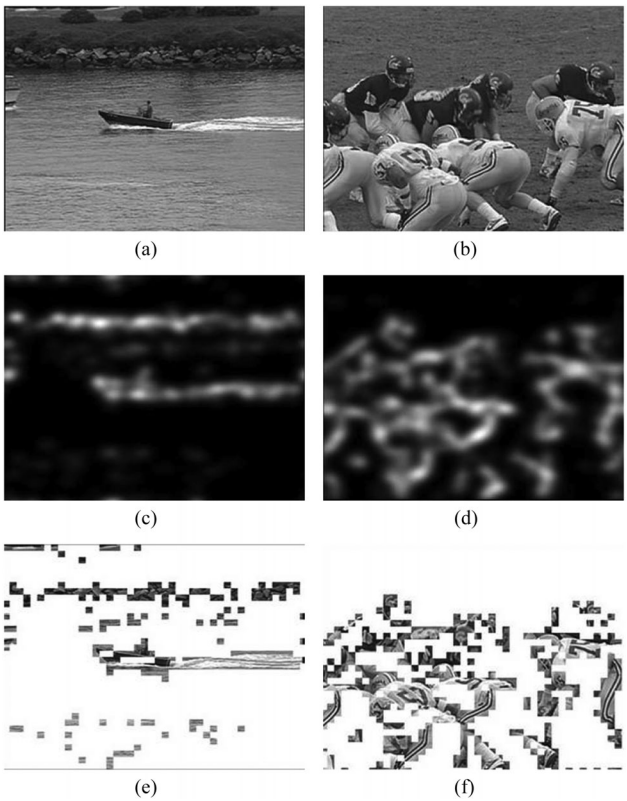


Fig. 6. Saliency map based on CS of Coastguard CIF and Football CIF: (a), (b) original image; (c), (d) using the proposed dictionary learning method; and (e), (f) salient blocks.

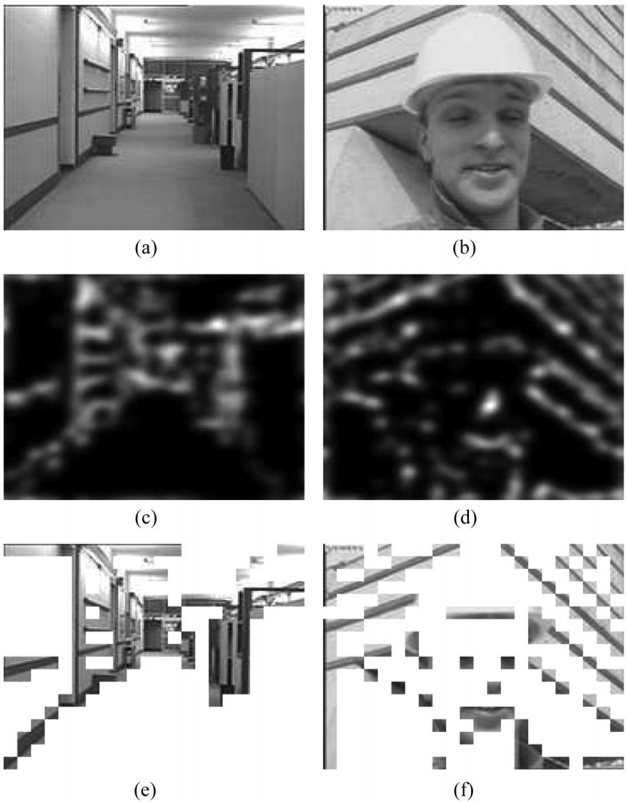


Fig. 7. Saliency map based on CS of Hall QCIF and Foreman QCIF: (a), (b) original image; (c), (d) using the proposed dictionary learning method; and (e), (f) salient blocks.

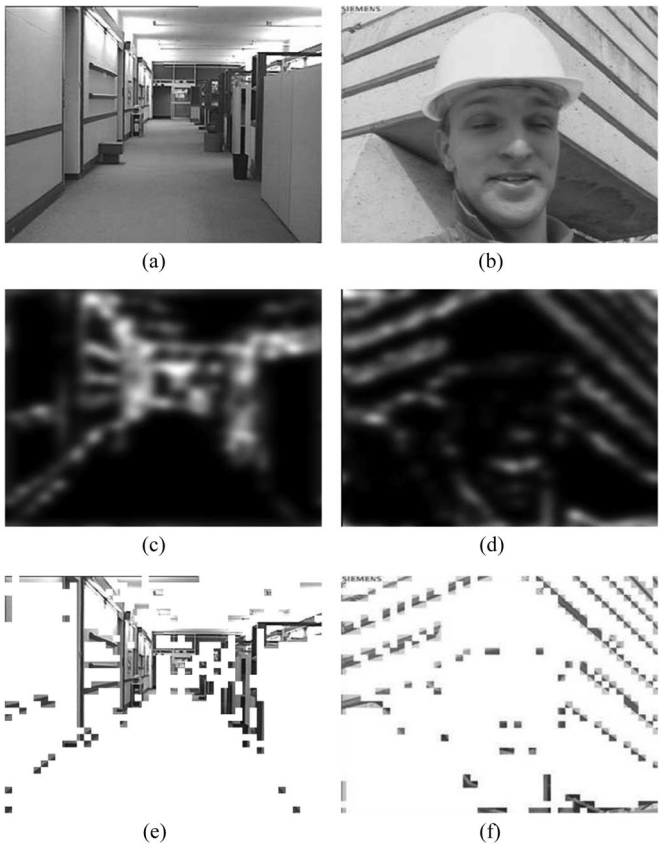


Fig. 8. Saliency map based on CS of Hall CIF and Foreman CIF: (a), (b) original image; (c), (d) using the proposed dictionary learning method; and (e), (f) salient blocks.

TABLE II
SIGNIFICANCE DEGREE OF DIFFERENT VIDEO SEQUENCES

QCIF	Coastguard	Football	Hall	Foreman
N_s	0.12	0.4	0.32	0.3
Q_s	0.09	0.164	0.42	0.38
Q_c	0.11	0.7	0.34	0.32
Q	0.31	1.264	1.04	1.0
CIF	Coastguard	Football	Hall	Foreman
N_s	0.13	0.27	0.21	0.14
Q_s	0.069	0.3	0.53	0.42
Q_c	0.12	0.69	0.32	0.39
Q	0.319	1.26	1.06	0.95

$0.1 = 0.121$. It should be noted that, for simplicity, α, β and γ are all set to be 1.

The significance degrees tell us that Frame 1 is more important than Frame 2. This result is consistent with the ground true (i.e., Frame 1 is more important and users would be more interested in Frame 1). Similarly, the significance degree of QCIF sequences of Coastguard, Football, Hall, Foreman are 0.31, 1.264, 1.04, 1.0 respectively; the significance degree of CIF sequences of Coastguard, Football, Hall, Foreman are 0.319, 1.26, 1.06, 0.95 respectively. The calculation details are illustrated in Table II.

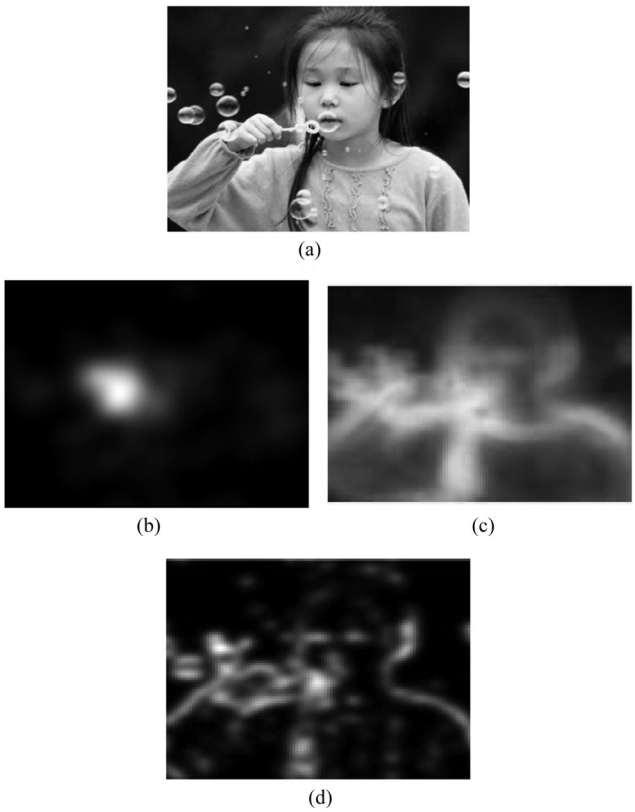


Fig. 9. Saliency map comparison. (a) CAT2000 database 001 JPG. (b) Human fixation. (c) Saliency map in [25]. (d) Proposed saliency map.

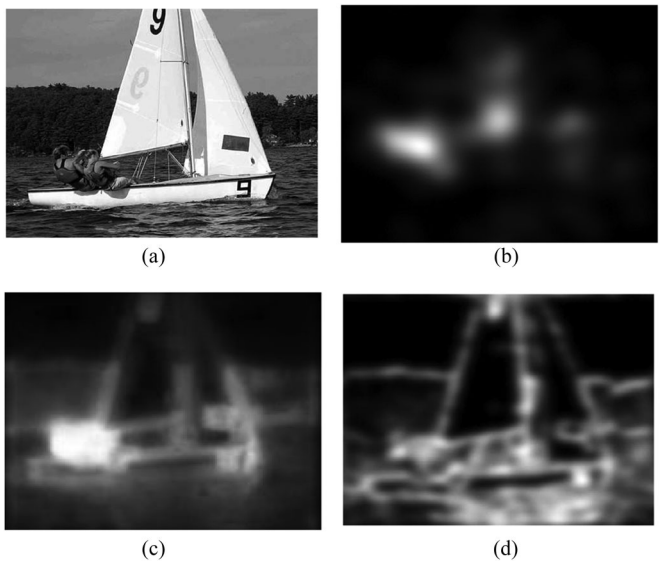


Fig. 10. Saliency map comparison. (a) CAT2000 database 085 JPG. (b) Human fixation. (c) Saliency map in [25]. (d) Proposed saliency map.

In addition, to further evaluate our proposal, the CAT2000 database which contains 4000 images from 20 different categories. For simplicity, the sizes of the images are tailored into 480×400. We give two example images from the CAT2000 database and compare the results of the saliency map proposed in this paper and that in [25], as shown in Figs. 9 and 10.

TABLE III
AUC VALUE OF CAT2000 DATABASE

CAT2000	Proposed	Pixel Domain Method
001 JPG	0.85	0.86
035 JPG	0.89	0.90
045 JPG	0.88	0.90
055 JPG	0.85	0.87
065 JPG	0.86	0.90
085 JPG	0.87	0.91

TABLE IV
COMPARISON OF TIME COMPLEXITY

CAT2000	Proposed	Pixel Domain method
001 JPG	6.2 s	8 s
035 JPG	5.1 s	6 s
045 JPG	6.3 s	7.1 s
055 JPG	5.3 s	5.5 s
065 JPG	6.2 s	6.4 s
085 JPG	5.9 s	7 s

It can be observed that the proposed saliency map compares favourable with that in [25]. To see the comparison clearly, the AUC (Area Under the Curve, shown in Table III) are adopted. We can see that the AUC value of our proposed method is close to the saliency map by using pixels. Meanwhile, since the proposed method is performed in measurement domain, it saves processing time, as shown in Table IV).

It can be concluded that the proposed evaluation method agrees with the video contents and human's perceptual intuitions for the reason that images which would catch more attentions have a higher value of Q. Besides, by using the proposed saliency map based on CS, we can calculate the significance of videos with a relatively fast speed and favourable significance evaluation effect.

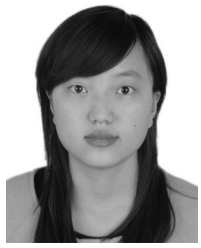
V. CONCLUSION

Considering the changeable property of the wireless network between the media cloud and users, in this paper, we proposed a novel significance-evaluation method for video frames based on CS. Specifically, by using the trained dictionary in measurement domain, we generate the saliency map by keeping the sparse part of one frame, then we analyze the area and distribution of salient regions. We also presented a computing method that gives the degree of significance of a frame. This proposal can be used to guarantee the transmission of videos which have higher significance. The experimental results showed that our method works very well. In the future, we will simplify the evaluation process of video data. In addition, combining with users' experience, the values of α, β and γ in (11) and their respective influence on the significance degree of frames should be investigated.

REFERENCES

[1] "Supporting wireless video growth and trends," 4G Americas, Bellevue, WA, USA, Apr. 2013.

[2] "Cisco Visual Networking Index: Global mobile data traffic forecast update, 2013–2018," Cisco Syst., Inc., San Jose, CA, USA, Feb. 2014.
[3] "14 LTE Broadcast business cases," Expway, Paris, France, Jun. 2014.
[4] D. Diaz-Sanez *et al.*, "Media cloud: An open cloud computing middleware for content management," *IEEE Trans. Consum. Electron.*, vol. 57, no. 2, pp. 970–978, May 2011.
[5] G. Gao *et al.*, "Cost optimal video transcoding in media cloud: Insights from user viewing pattern," in *Proc. IEEE Int. Conf. Multimedia. Expo.*, Jul. 2014, pp. 1–6.
[6] J. Lee, J. Son, R. Hussain, and H. Oh, "Media cloud: A secure and efficient virtualization framework for media service," in *Proc. IEEE Int. Conf. Consum. Electron.*, pp. 79–80, Jan. 2014.
[7] Y. Jin, Y. Wen, and C. Westphal, "Optimal trancoding and caching for adaptive streaming in media cloud: An analytical approach," *IEEE Trans. Circuts Syst. Video Technol.*, vol. 25, no. 12, pp. 1914–1925, Dec. 2015.
[8] S. Soltani, K. Misra and H. Radha, "Delay constraint error control protocol for real-time video communication," *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 742–751, Jun. 2009.
[9] H. Lee and S. Sull, "A VBR video encoding for locally consistent picture quality with small buffering delay under limited bandwidth," *IEEE Trans. Broadcast.*, vol. 58, no. 1, pp. 47–56, Mar. 2012.
[10] D. Ong and T. Moors, "Deferred discard for improving the quality of video sent across congested networks," in *Proc. IEEE 38th Conf. Local Comput. Netw.*, Oct. 2013, pp. 743–746.
[11] S. Chang, R. Chang, J. Ho, and Y. Oyang, "A priority selected cache algorithm for video relay in streaming applications," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 79–91, Mar. 2007.
[12] K. C. Lin, W. Shen, C. Hsu, and C. Chou, "Quality-differentiated video multicast in multirate wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 21–34, Jan. 2013.
[13] T. T. Do *et al.*, "Distributed compressed video sensing," in *Proc. IEEE 16th Int. Conf. Image Process.*, Nov. 2009, pp. 1393–1396.
[14] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[15] E. Cands, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[16] J. Yan, M. Zhu, H. Liu, and Y. Liu, "Visual saliency detection via sparsity pursuit," *IEEE Signal Process. Lett.*, vol. 17, no. 8, pp. 739–742, Aug. 2010.
[17] M. Aghagolzadeh and H. Radha, "Adaptive dictionaries for compressive imaging," in *Proc. IEEE Global Conf. Signal Inform. Process.*, Dec. 2013, pp. 1033–1036.
[18] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[19] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
[20] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[21] F. P. Anaraki and S. M. Hughes, "Compressive K-SVD," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013, pp. 5469–5473.
[22] Z. Liu, H. V. Zhao, and A. Y. Elezzabi, "Block-Based adaptive compressed sensing for video," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1649–1652.
[23] L. Zhou *et al.*, "Salient region detection via integrating diffusion-based compactness and local contrast," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3308–3320, Nov. 2015.
[24] W. Cheng, Y. Tsai, and C. Lin, "Prioritized retransmission for error protection of video streaming over WLANs," in *Proc. Int. Symp. Circuits Syst.*, May 2004, vol. 2, pp. 65–68.
[25] A. Borji and L. Itti, "CAT2000: A large scale fixation dataset for boosting saliency research," *CoRR*, 2015. [Online]. Available: http://arxiv.org/abs/1505.03581
[26] G. Xue, Li Song, and J. Sun, "Foreground estimation based on linear regression model with fused sparsity on outliers," *IEEE Trans. Circuts Syst. Video Technol.*, vol. 23, no. 8, pp. 1346–1357, Aug. 2013.



Jie Guo received the B.S. degree in communication engineering from Zhengzhou University, Zhengzhou, China, in 2011, and is currently working toward the Ph.D. degree at Xidian University, Xi'an, China.

Her research interests include video compression, transmission, and compressed video sensing.

Bin Song (M'10) received the BS, M.S., and Ph.D. degrees in communication and information systems from Xidian University, Xi'an, China, in 1996, 1999, and 2002, respectively.

In 2002, he joined the School of Telecommunications Engineering, Xidian University, where he is currently a Professor of Communications and Information Systems. He is also the Associate Director of the State Key Laboratory of Integrated Services Networks, Xidian University, Xi.an, China. He has authored or coauthored more than 50 journal or conference papers and 30 patents. His research interests and areas of publication include video compression and transmission technologies, video transcoding, distributed video coding, and video signal processing based on compressed sensing, big data, and multimedia communications.

Xiaojiang Du (M'04-SM'09) received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland at College Park, College Park, MD, USA, in 2002 and 2003, respectively.

He is a Tenured Associate Professor with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA, USA. He has authored or coauthored more than 170 journal and conference papers, as well as a book published by Springer. His research interests include wireless communications, wireless networks, security, and systems.

Prof. Du served as the Lead Chair of the Communication and Information Security Symposium of the IEEE International Communication Conference 2015, and a Co-Chair of the Mobile and Wireless Networks Track of the IEEE Wireless Communications and Networking Conference 2015. He was a Technical Program Committee Member of several premier ACM/IEEE conferences such as INFOCOM (2007–2016), IM, NOMS, ICC, GLOBECOM, WCNC, BroadNet, and IPCCC.