Statistical Conditional Sampling for Variable-Resolution Video Compression

Alexander Wong[1]*, **Mohammad Javad Shafiee**[2], **Zohreh Azimifar**[2]

[1] Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada, [2] School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

Abstract

In this study, we investigate a variable-resolution approach to video compression based on Conditional Random Field and statistical conditional sampling in order to further improve compression rate while maintaining high-quality video. In the proposed approach, representative key-frames within a video shot are identified and stored at full resolution. The remaining frames within the video shot are stored and compressed at a reduced resolution. At the decompression stage, a region-based dictionary is constructed from the key-frames and used to restore the reduced resolution frames to the original resolution via statistical conditional sampling. The sampling approach is based on the conditional probability of the CRF modeling by use of the constructed dictionary. Experimental results show that the proposed variable-resolution approach via statistical conditional sampling has potential for improving compression rates when compared to compressing the video at full resolution, while achieving higher video quality when compared to compressing the video at reduced resolution.

Citation: Wong A, Shafiee MJ, Azimifar Z (2012) Statistical Conditional Sampling for Variable-Resolution Video Compression. PLoS ONE 7(10): e45002. doi:10.1371/journal.pone.0045002

Editor: Teresa Serrano-Gotarredona, National Microelectronics Center, Spain

Received May 4, 2012; **Accepted** August 11, 2012; **Published** October 8, 2012

Funding: The authors would like to thank the Natural Sciences and Engineering Research Council of Canada for funding this research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: a28wong@uwaterloo.ca

Introduction

Over the last two decades, digital video compression has become one of the fastest growing areas of research and development around the world, where the underlying goal is to take digital video content and encode it in a form that minimizes the requirements for digital storage and/or transmission. There is a continually increasing demand for better digital video compression technologies, particularly since digital video has become an integral part of our daily lives, with mass digital video consumption in a wide range of application scenarios such as digital TV broadcast (via MPEG-2 [1] in most North American systems), real-time Internet video streaming, real-time video telecommunications (e.g., via H.32x [2]), personal video recording, and media disk storage (e.g., DVDs). Given the incredible demand for high quality digital video content consumption, significant progress has been made in the area of digital video compression, cumulating in the current state-of-the-art video compression standards such as H.264/MPEG-4 AVC [3], a block-transform motion-compensated based digital video codec standard that provides significantly improved compression rates when compared to previous standards. Much of the gains in compression performance over the past two decades in digital video compression has been largely due to improvements on rate-distortion optimization techniques [4–7] and motion compensation techniques [8–12] such as improved inter-frame utilization, variable block size motion compensation, multiple motion vectors per macroblock, and sub-pixel motion compensation precision.

Despite the great increases in compression performance gained through rate-distortion optimization and motion compensation, another area of research in digital video compression that has garnered recent interest and is worth investigating is the area of variable resolution compression [13–17], where the underlying video content is stored and compressed at different spatial resolutions. In the work by Wei et al. [13], salient regions are detected within the scene via a visual attention model. The regions with the highest saliency is stored and compressed at its original resolution, the regions with lowest saliency stored at medium resolution, and regions in between stored at medium resolution. In the work by Defroges et al. [15–17], referred to as locally adaptive resolution (LAR) techniques, regions of interest are extracted from the scene via a region segmentation approach. These regions of interest are then reduced in resolution depending on the underlying content, such that smaller regions maintain higher resolution for the underlying video content while larger regions are stored and compressed at a reduced resolution.

One of the main limitations with existing variable resolution compression techniques is that they are largely constrained to exploiting spatial redundancy within a video frame. As such, the significant information redundancy that can be gained by considering the spatial-temporal characteristics of the underlying video content is largely untapped in current methods. Furthermore, existing variable resolution methods require significant modifications and even architectural departures from current state-of-the-art video compression standards. As such, a method that addresses both issues is worth investigating.

The main contribution of this paper is to introduce and investigate the potential for the use of Conditional Random Fields and statistical conditional sampling for variable-resolution video compression, with the aim to improve compression rates while maintain high visual quality. Rather than store individual regions within a frame at different resolutions, as previous approaches have done, we take a radically different approach where different frames within a video are stored and compressed at different resolutions. At encoding, the keyframes are stored at full resolution, while the rest of the frames are stored at reduced resolutions. At decoding, a region-based dictionary from high resolution representative key-frames within a video shot is constructed automatically and statistical conditional sampling based on Conditional Random Field is used to restore the low resolution frames to the original resolution based on information contained within the full resolution dictionary of regions. By incorporating the proposed approach within a H.264 video compression framework, the proposed approach can take advantage of all the advanced rate distortion optimization and motion compensation techniques inherent and available for H.264 while provided an additional value-added component for improving compression performance over the existing framework.

The rest of the paper is organized as follows. First, the underlying methodology behind the proposed use of Conditional Random Fields and statistical conditional sampling for variable-resolution video compression is described in Section. The experimental results and the associated discussion is presented in Section. Finally, conclusion are drawn and future work is discussed in Section.

Methods

The proposed use of statistical conditional sampling for variable-resolution video compression consists of two main stages: i) Variable-resolution compression stage, and ii) Decompression stage. In the variable-resolution compression stage, the identified representative key-frames are compressed at full resolution while the rest of the frames are compressed at a lower resolution. Secondly, in the decompression stage, all of the frames within the video content are decompressed at their respective resolutions, a full resolution region-based dictionary is constructed from the representative key-frames and then the low resolution frames are restored to the original resolution via statistical conditional sampling based on the dictionary and conditional probability of CRF. An flowchart summarizing the proposed approach is shown in Fig. 1.

Representative frame identification stage

At the first step in compression stage, representative key-frames are identified and extracted within a video shot. To achieve this goal, we wish to first determine an appropriate metric for quantifying the similarity between every frame pair within the shot. Given the importance of structures of objects within a scene, we chose to employ the well-known SSIM metric proposed by Wang et al. [18], which has been shown to provide a strong indicator for visual similarity assessment in a local manner. The local nature of the metric is important since:

- video statistical features are usually highly spatially/temporally nonstationary,
- video distortions may also be space/time variant,
- at one time instance and at a typical viewing distance, only a local area in the image can be perceived with high resolution by the human observer, and
- a localized quality measurement can provide a space-time varying quality map of the image, which delivers more information about the quality degradation of the image and may be useful in some applications.

Having considered the above properties, the SSIM metric can be defined as

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+Q_1)(2\sigma_{xy}+Q_2)}{(\mu_x^2+\mu_y^2+Q_1)(\sigma_x^2+\sigma_y^2+Q_2)}, \tag{1}$$

where the constant Q_1 is included to avoid instability when $\mu_x^2+\mu_y^2$ is very close to zero. Specifically, Q_1 is chosen as the squared product of pixel values dynamic range and a small positive constant much less then one. Similarly, the constant Q_2 is assumed as the squared product of pixel values dynamic range and another small constant.

To utilize the SSIM metric for assessing video frame similarity so that we can identify representative keyframes, an SSIM matrix (S) is first constructed, where the elements of the matrix indicate the SSIM value between every two frames within a shot. A distance matrix D is then calculated to obtain the temporal distance map of the given shot,

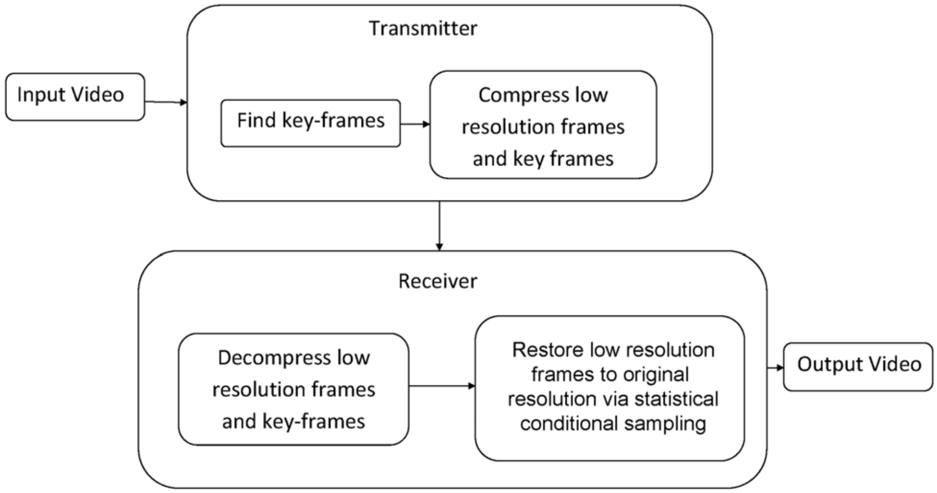

Transmitter
Input Video
Find key-frames
Compress low resolution frames and key frames
Receiver
Decompress low resolution frames and key-frames
Restore low resolution frames to original resolution via statistical conditional sampling
Output Video

Figure 1. Flow diagram of the proposed variable-resolution approach.
doi:10.1371/journal.pone.0045002.g001

$$D=1-S. \tag{2}$$

For each frame i within the video shot, a vector with size n depicts the distance of the frame and other $n-1$ frames, where the j^{th} entry of the i^{th} vector shows the SSIM of frame i and frame j.

In order to avoid identifying uniformly distributed key-frames, a Fuzzy c-means clustering strategy [19] is employed to identify the representative keyframes within the video shot. To determine the number of clusters (i.e., the actual number of representative keyframes to store), a principal component analysis (PCA) approach is utilized, where one can determine the significant eigenvalues within the set of data and use them to determine a reasonable estimate of the number of clusters, i.e., number of representative key-frames within the shot to store. This proposed key-frame identification and selection process is important since the number of keyframes that needs to be stored and compressed can vary greatly from shot to shot depending on the underlying video content.

Based on the above theory, the representation keyframe identification and selection procedure from a video shot can be described in detail as follows (Fig. 2). Suppose that the video shot F contains n frames $\{f_1:f_n\}\in F$. Because we wish to select the most informative and representative frames as the representation keyframes, the similarity of each pair of frames is calculated using the SSIM measure (defined in Eq. 1). The similarity between a reference frame f_i and a secondary frame f_j will be denoted as s_{ij}. Based on the similarity s_{ij}, one can get an assessment of dissimilarity d_{ij} as its inverse: $d_{ij}=1-s_{ij}$. Therefore, the dissimilarity matrix D representing the dissimilarities between all frames in the video shot as:

$$D=\begin{bmatrix} 0 & \cdots & d_{1,n} \\ d_{2,1} & \cdots & d_{2,n} \\ \vdots & 0 & \vdots \\ d_{n,1} & \cdots & 0 \end{bmatrix} \tag{3}$$

where D is sample space which will be utilized to identify the keyframes. Each row i of D is the corresponding feature vector for frame i. As mentioned, first of all, PCA is used to determined the sufficient number of keyframes to be identified for the video shot. Based on the covariance P of the sample space D, the number of keyframes K is specified to be the number of significant eigenvalues.

Once the number of clusters K is determined, the fuzzy c-means (FCM) clustering procedure is used to select the most informative and representative keyframes. The FCM algorithm attempts to partition a finite collection of n elements $D=\{d^1,\ldots,d^n\}$ into a collection of K fuzzy clusters. Given a finite set of data, the algorithm returns a list of K cluster centers $C=\{c^1,\ldots,C^K\}$ and a partition matrix $U=u_{i,j}\in[0,1]$, $i=1,\ldots,n$, $j=1,\ldots,K$ (Eq. (4)). Each element u_{ij} characterizes the degree to which element d^i belongs to cluster c^j:

$$u_{ij}=\frac{1}{\sum_{k=1}^{K}\left(\frac{\|d^i-c^j\|}{\|d^i-c^k\|}\right)^{\frac{2}{m-1}}} \tag{4}$$

$$c^j=\frac{\sum_{i=1}^{n}u_{ij}^{m}d^i}{\sum_{i=1}^{n}u_{ij}^{m}} \tag{5}$$

This procedure is iterated m times until convergence is achieved. After the procedure converges, the K clusters are identified. To find the keyframes, the nearest sample i to each cluster center j is determined based on minimum distance:

$$key\,frame_j=\arg\min_{d^i}\left[distance(d^i,c^j)\right] \tag{6}$$

At this stage, the representative keyframes have been selected and are stored at the original resolution.

Variable-resolution compression stage

In the variable-resolution compression stage, the frames within the video content are stored and compressed via H.264 [3] depending on whether it is one of the identified representative key-frames or not. For the set of frames that are not identified as representative key-frames, they are down-scaled to a lower resolution and compressed as a video sequence at this reduced resolution. For implementation purposes, this set of frames are down-scaled by a factor of 2 in both the vertical and horizontal resolutions. The keyframes are compressed at the original resolution. By compressing them at the original resolution, much of the important details within the frames are well preserved, which is fundamental for the decompression stage when we attempt to restore the lower resolution video frames to their original resolutions. The main advantage of this compression approach is that a state-of-the-art video compression standard such as H.264 can be used directly for variable-rate video compression without the need for significant modifications, making it well suited for integration into consumer level media devices.

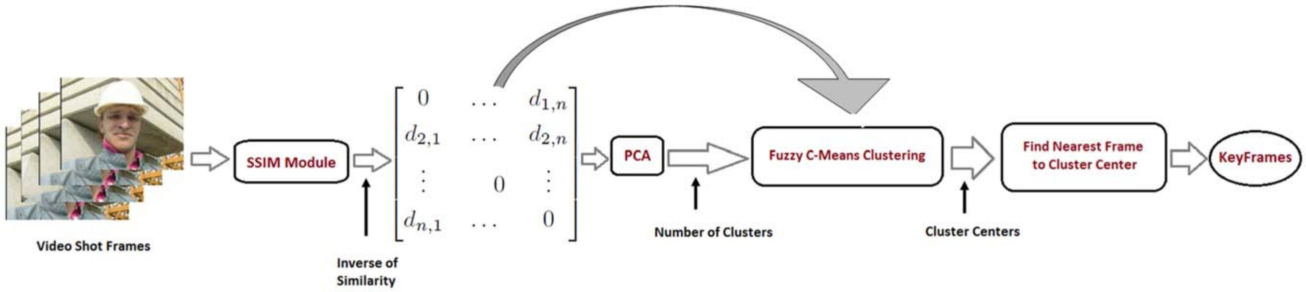

Video Shot Frames
SSIM Module
Inverse of Similarity
$\begin{bmatrix} 0 & \cdots & d_{1,n} \\ d_{2,1} & \cdots & d_{2,n} \\ \vdots & 0 & \vdots \\ d_{n,1} & \cdots & 0 \end{bmatrix}$
PCA
Number of Clusters
Fuzzy C-Means Clustering
Cluster Centers
Find Nearest Frame to Cluster Center
KeyFrames

Figure 2. Flow diagram of the representative keyframe identification and selection procedure.
doi:10.1371/journal.pone.0045002.g002

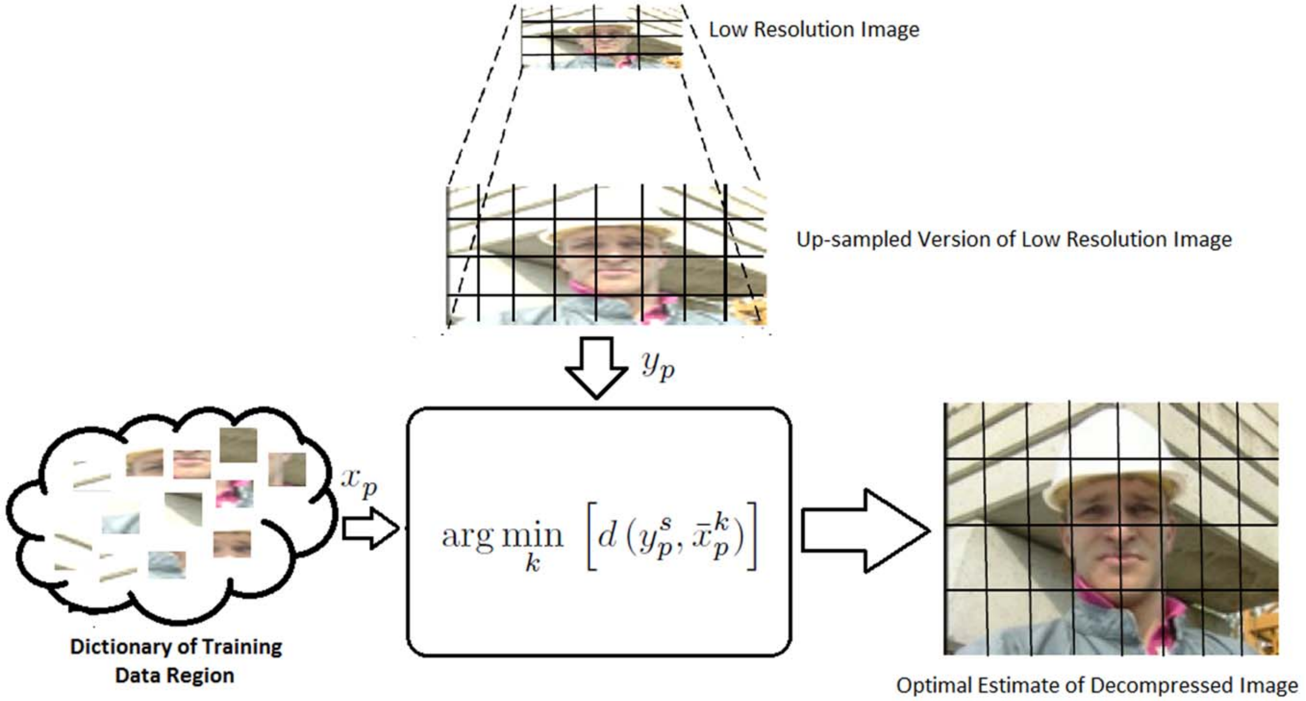

Low Resolution Image
Up-sampled Version of Low Resolution Image
y_p
x_p
$\arg\min_k \left[d\left(y_p^s, \bar{x}_p^k\right)\right]$
Dictionary of Training Data Region
Optimal Estimate of Decompressed Image

Figure 3. Flow diagram of the sampling and inference step.
doi:10.1371/journal.pone.0045002.g003

Decompression stage

In this stage, the goal is to reconstruct the decompressed video content back to the original resolution. First, the representative key-frames are decoded and decompressed at full resolution, while the rest of the frames are decoded and decompressed at the reduced resolution. In this stage the region-based dictionary D is constructed from full resolution key-frames. Once we have the decompressed frames, we restore the low resolution frames to the original resolution via statistical conditional sampling, which is described as follows.

Let y is a realization of low resolution frame $Y=\{Y_s : s\in S_L\}$, and x is a realization of original resolution frame $X=\{X_s : s\in S_H\}$, where S_L is the set of all pixels within the low resolution frame, while S_H is the set of all pixels within the original resolution frame. The conditional probability of x given y can be expressed as:

$$p(x|y;D)\propto \prod_{c\in C} \psi_c(y_c,x_c) \tag{7}$$

where $p(x|y)$ is modeled by Conditional Random Field (CRF) [20] (a parametric model) in which C is the set of clique templates, $\psi_c(.,.)$ is a potential function, and D is the dictionary of high resolution regions which were extracted from key-frames. We can determine the original resolution frame $\hat{x}$ by sampling from $p(x|y)$:

$$\hat{x}\leftarrow p(x|y) \tag{8}$$

while various potential functions $\psi_c(.,.)$ can be applied, the most simplest and effective one is Sum of Squared Difference (SSD). As the objective of this paper is to find the best high resolution frame based on the low resolution compressed video frame and the key-frames, SSD was found to be an appropriate metric.

Training. The only feature function utilizing in this paper is the SSD measure, therefore, the training phase of CRF simply is to determine the key-frames utilized to construct the dictionary D [21]. For efficient implementation purposes, the region-based dictionary D is constructed for each pixel s in the following manner. First, a total of N samples are randomly drawn from a 2-D Gaussian sampling distribution

Table 1. The compression ratio of different sequences for the following scenarios: i) compression at full resolution (FR), ii) compression at low resolution (LR), and iii) compression via variable-resolution (VR) approach.

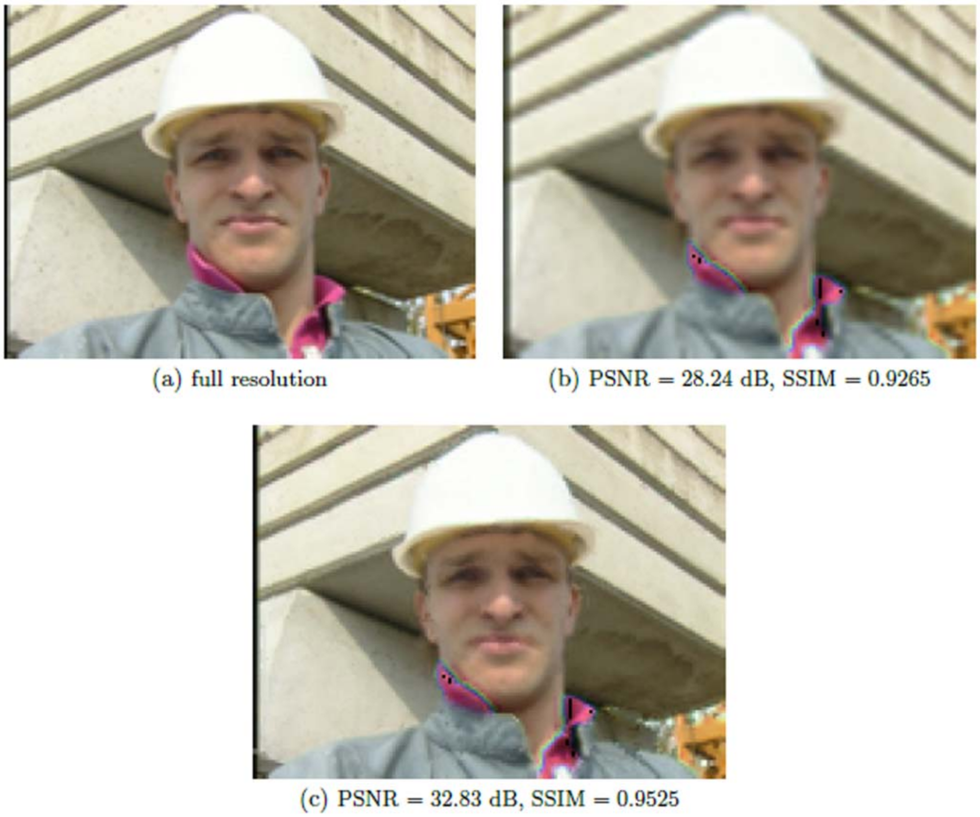
	Compression ratio		
Sequence	**FR**	**LR**	**VR**
Foreman	5.5:1	9.31:1	7.68:1
Table Tennis	3.98:1	14.84:1	12.44:1
Ohaio1	6.48:1	20.90:1	15.90:1
Ohaio2	6.52:1	26.26:1	17.96:1

doi:10.1371/journal.pone.0045002.t001

Table 2. Averaged PSNRs (dB) and SSIMs of the reconstructed frames for low resolution (LR) compression and the proposed variable-resolution (VR) compression approach.

	PSNR (dB)		SSIM	
Sequence	**LR**	**VR**	**LR**	**VR**
Foreman	27.77	31.27	0.9211	0.9448
Table Tennis	26.68	27.40	0.7280	0.7791
Ohaio1	25.29	26.90	0.8596	0.8811
Ohaio2	26.56	28.26	0.8450	0.8853

doi:10.1371/journal.pone.0045002.t002



(a) full resolution

(b) PSNR = 28.24 dB, SSIM = 0.9265

(c) PSNR = 32.83 dB, SSIM = 0.9525

Figure 4. Visual comparison of the proposed method on two example frames from 'Foreman' video sequence [22]. (a) full resolution original frames, (b) Results of low resolution video compression and (c) depicts result of proposed variable-resolution compression approach.
doi:10.1371/journal.pone.0045002.g004

with a mean of s and a standard deviation of n_s. At the pixel locations corresponding to each of the N samples, a $n_r \times n_r$ high resolution region around that pixel is extracted from the representative high resolution keyframes and stored into the dictionary. In this study, $n_s = 9$ and $n_r = 5$ as they were found experimentally to provide strong visual quality.

Sampling and inference. The sampling is done to find the best matching high resolution frames. The original resolution frame $\hat{x}$ is estimated by directly sampling from the dictionary of region-based training data $D = \{\bar{x}_p^k, k \in [0, \cdots, N]\}$ according to $p(x|y)$. This is accomplished by computing the optimal estimate $\bar{x}_p^{s,k^*}$ for reconstructing the original resolution frame by identifying the best regional match for each pixel s in an up-scaled version of the low resolution frame y, denoted as y_p^s, with the dictionary of regions $\bar{x}_p$,

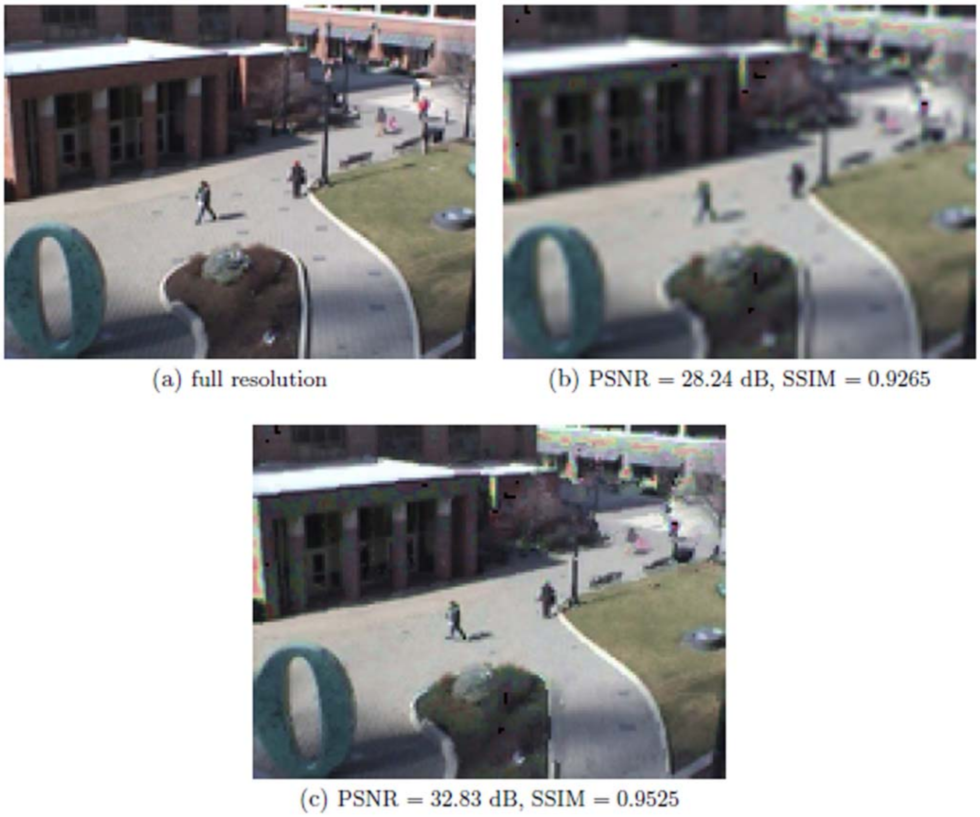
$$k^* = \arg\min_k [d(y_p^s, \bar{x}_p^k)], \quad \forall k \in [0, \cdots, N]. \tag{9}$$

where d is the dissimilarity metric between two regions (for implementation purposes, d is the sum of squared differences between the regions), and the clique definition on this approach is based on the 5×5 neighborhood structure. Once the best matching region from the dictionary D for a pixel s is determined, the value at s in the estimated original resolution frame $\hat{x}$ is set to the value of the center pixel of that best matching region. The overview workflow of the sampling and inference step is shown in Fig. 3.

Results and Discussion

To demonstrate the potential of the proposed use of statistical conditional sampling for variable-rate video compression, a number of different video sequences were tested. Two main performance metrics were evaluated. First, we evaluate the compression rate achieved using the proposed method against the compression rate achieved by: 1) compressing the entire video sequences at full resolution using H.264 [3], and ii) compressing the entire video sequences at a reduced resolution of a factor of two for both horizontal and vertical resolutions using H.264 [3]. H.264 [3] is a state-of-the-art video compression framework that accounts for inter-frame redundancy. This performance metric allows us to evaluate whether the proposed variable-resolution approach's claims for improving compression performance is valid. Second, we evaluate the average peak signal-to-noise ratio (PSNR) and the average structural similarity index (SSIM) values of the video frames produced using the proposed approach, and compare to that achieved by compressing the entire video sequence at a reduced resolution. This performance metric allows us to evaluate whether the proposed approach's claims for improved video quality over compressing at a reduced resolution is valid.

Table 1 demonstrates compression ratio of the tested scenarios for each frame sequence. It can be observed that the compression ratios achieved using the proposed variable-resolution approach is noticeably higher than that achieved using the full resolution approach, which justifies the claim for the proposed approach of improving compression performance. When compared to the compression ratios achieved by the low resolution approach, the proposed approach takes a minor hit in storage overhead for the 'Foreman' and 'Table Tennis' video sequences, while taking a larger hit in storage overhead for the 'Ohaio1' and 'Ohaio2' video sequences. However, despite the storage overhead when compared


(a) full resolution

(b) PSNR = 28.24 dB, SSIM = 0.9265

(c) PSNR = 32.83 dB, SSIM = 0.9525

Figure 5. Visual comparison of the proposed method on two example frames from 'Ohaio1' video sequence. (a) full resolution original frames, (b) Results of low resolution video compression and (c) depicts result of proposed variable-resolution compression approach.
doi:10.1371/journal.pone.0045002.g005

to the low resolution approach, the overall compression performance of the proposed variable-resolution approach is still strong.

Table 2 shows the average PSNR and average SSIM values for the proposed use of statistical conditional sampling for variable-resolution video compression and for the scenario where the entire video sequence is compressed at a lower resolution. To facilitate for comparison purposes, the video frames from the low resolution scenario is up-scaled using bi-cubic interpolation so that it can be compared against the reference full resolution video frame. It can be observed that strong PSNR and SSIM values were obtained using the proposed approach for all video sequences when compared to the low resolution approach. Furthermore, a visual comparison of the proposed approach on two example frames from the 'Foreman' and 'Ohaio1' video sequences are shown in Fig. 4 and Fig. 5, respectively. It can be observed that the frames produced using the proposed approach contains noticeably more detail when compared to the frames produced using low resolution compression, thus validating the claim that improved visual quality can be achieved using the proposed approach. However, as expected, the visual quality of the frames produced using the proposed approach is not as good as the full resolution original frames, thus illustrating the trade off between visual quality and compression performance.

Conclusions

The potential use of statistical conditional sampling for variable-resolution video compression to further improve compression rate while maintaining high quality video was studied. In the proposed approach, the representative key-frames were first identified within a video shot. The representative key-frames were compressed at the original resolution while the remaining frames within the video shot are compressed at a reduced resolution. Upon decompression, the reduced resolution frames are restored to the original resolution via statistical conditional sampling based on the original resolution representative keyframes. Experimental results demonstrate the potential of the proposed approach for improving compression rates when compared to compressing the video at full resolution, while achieving higher video quality when compared to compressing the video at reduced resolution. Future work involves exploring improved key-frame identification methods as well as improved frame restoration approaches.

Author Contributions

Conceived and designed the experiments: AW MS. Performed the experiments: AW MS. Analyzed the data: AW MS ZA. Wrote the paper: AW MS ZA.

References

1. Int Telecommun Union-Telecommun (1994) Generic coding of moving pictures and associated audio information - part 1: Systems.
2. Int Telecommun Union-Telecommun (1999) Narrow-band visual telephone systems and terminal equipment.
3. Bjontegaard G (2000) H.26L test model long term number 4 (TML 4) draft0.
4. Wiegand T, Lightstone M, Mukherjee D, Campbell T, Mitra S (1996) Rate-distortion optimized mode selection for very low bit rate video coding and the emerging h.263 standard. IEEE Trans on Circuits and Systems for Video Technology 6: 182–190.
5. Sullivan G, Wiegand T (1998) Rate-distortion optimization for video compression. IEEE Signal Processing Magazine 15: 74–90.

6. Wiegand T, Schwarz H, Joch A, Kossentini F, Sullivan G (2003) Rate-constrained coder control and comparison of video coding standards. IEEE Trans on Circuits and Systems for Video Technology 13: 688–703.
7. Cook G, Prades-Nebot J, Liu Y, Delp E (2006) Rate-distortion analysis of motion-compensated rate scalable video. IEEE Trans on Image Processing 15: 2170–2190.
8. Wiegand T, Zhang X, Girod B (1999) Long-term memory motion-compensated prediction. IEEE Trans on Circuits and Systems for Video Technology 9: 70–84.
9. Wiegand T, Girod B (2001) Multi-frame Motion-Compensated Prediction for Video Transmission.
10. Li J, Liu H, Wang L, Zhang K (2009) An improved motion estimation for spatially scalable video coding. In: 2nd International Congress on Image and Signal Processing. pp. 1–5.
11. Chou L, Ye C, Liu Y, Jhao B (2007) Fast predictive search algorithm for video motion estimation. In: 14th International Conference on Image Analysis and Processing. pp. 399–406.
12. Xiong R, Xu J, Wu F (2008) In-scale motion compensation for spatially scalable video coding. IEEE Trans on Circuits and Systems for Video Technology 18: 145–158.
13. Wei L, Sang N, Wang Y, Wang D, Wang F (2009) Variable resolution image compression based on a model of visual attention. In: Proceedings of the SPIE. volume 7495, p. 74950P.
14. Sahabi H, Basu A, Fiala M (1995) Vlsi implementation of variable resolution image compression. In: Proceedings of the 8th International Conference on VLSI Design. pp. 214–219.
15. Deforges O, Babel M (2000) Region of interest coding for low bit-rate image transmission. In: Proc. IEEE International Conference on Multimedia and Expo. pp. 107–110.
16. Deforges O, Babel M (2008) LAR method: from algorithm to synthesis for an embedded low complexity image coder. In: Proc. 3rd International Design and Test Workshop. pp. 187–192.
17. Deforges O, Babel M, Bedat L, Ronsin J (2007) Color LAR codec: A color image representation and compression scheme based on local resolution adjustment and self-extracting region representation. IEEE Trans on Circuits and Systems for Video Technology 17: 974–987.
18. Wang Z, Bovik A, Sheikh H, Simoncelli E (2004) Image quality assessment: From error visibility to structural similarity. IEEE Trans Image Processing 13: 600–612.
19. Bezdek J, Ehrlich R, Full W (1984) FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences.
20. Lafferty J, McCallum A, Pereira F (2001) Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proc. 18th International Conf. on Machine Learning. pp. 282–289.
21. Kong D, Han M, Xu W, Tao H, Gong Y (2006) A conditional random field model for video super-resolution. In: 18th International Conference on Pattern Recognition. volume 3, pp. 619–622.
22. Video Trace Library (2012) Trace YUV video sequences.