Dual Diversified Dynamical Gaussian Process Latent Variable Model for Video Repairing

Hao Xiong, Tongliang Liu, Dacheng Tao, *Fellow, IEEE*, and Heng Tao Shen

***Abstract*— In this paper, we propose a dual diversified dynamical Gaussian process latent variable model (D^{3}GPLVM) to tackle the video repairing issue. For preservation purposes, videos have to be conserved on media. However, storing on media, such as films and hard disks, can suffer from unexpected data loss, for instance, physical damage. So repairing of missing or damaged pixels is essential for better video maintenance. Most methods seek to fill in missing holes by synthesizing similar textures from local patches (the neighboring pixels), consecutive frames, or the whole video. However, these can introduce incorrect contexts, especially when the missing hole or number of damaged frames is large. Furthermore, simple texture synthesis can introduce artifacts in undamaged and recovered areas. To address aforementioned problems, we introduce two diversity encouraging priors to both of inducing points and latent variables for considering the variety in existing videos. In D^{3}GPLVM, the inducing points constitute a smaller subset of observed data, while latent variables are a low-dimensional representation of observed data. Since they have a strong correlation with the observed data, it is essential that both of them can capture distinct aspects of and fully represent the observed data. The dual diversity encouraging priors ensure that the trained inducing points and latent variables are more diverse and resistant for context-aware and artifacts-free-based video repairing. The defined objective function in our proposed model is initially not analytically tractable and must be solved by variational inference. Finally, experimental testing results illustrate the robustness and effectiveness of our method for damaged video repairing.**

***Index Terms*— DGPLVM, inducing points, latect variable, diversity prior.**

I. Introduction

VIDEOS, from black and white films to modern blockbusters, are an indispensible part of contemporary life. For ease of future watching, it is commonplace that these videos are normally conserved on the celluloid or hard disks. It seems that these medias used to store videos inevitably suffer from natural deterioration or, sometimes, deliberate damages.

Manuscript received October 24, 2015; revised April 9, 2016 and May 15, 2016; accepted May 16, 2016. Date of publication May 26, 2016; date of current version June 16, 2016. This work was supported by the Australian Research Council under Project DP-140102164, Project LE-140100061, Project FT-130101457, and Project DP-130103252. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ling Shao.

H. Xiong, T. Liu, and D. Tao are with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: hao.xiong@student.uts.edu.au; tliang.liu@gmail.com; dacheng.tao@uts.edu.au).

H. T. Shen is with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia (e-mail: shenht@itee.uq.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2573581

Fig. 1. Four damaged video examples. First Two Columns: potential damages we aim to repair Last Two Columns: ground truth.

Such damages tend to appear as scattered missing blobs or large holes (see Fig. 1) in video frames, which in turn has an adverse impact on its appreciation or researches in image annotation and retrieval [1], [2], classification [3]–[6], action and face recognition [7]–[11], etc. It is, therefore, useful to recover damaged parts of videos. However, this is not trivial because sometimes it may be hard for individuals to predict the damaged parts, especially when the missing areas are large.

In essence, there are a number of existing methods that aim to repair damaged videos. A naive approach to video repairing is to regard the damaged frame as a single image and then applying the image inpainting methods, such as in [12] and [13], to recover the missing part. Such image inpainting techniques are able to retain linear structures, like object contours, without any assistance from other images. It turns out that simple image inpainting has the capability to accomplish small hole completion, but it thoroughly fails to repair damaged image or video with large missing areas. The reason for that is image inpainting tend to fill in the holes by synthesizing texture according to neighborhood without considering temporal information from other video frames. Since neighboring information in a single image is limited, highly relying on neighboring texture is likely to engender wrong context in the repaired video.

It is acknowledged that the difficulties of video repairing lies in the uncertainty and variety existing in present video scenes and also the fact that objects are often highly dynamic in the video sequence. Therefore, simply utilizing and synthesizing local information, such as neighbouring texture, cannot satisfy the variety and dynamic characteristics of real world videos. The work in [14] proposed to fill in missing

holes by sampling and synthesizing a number of patches from neighboring frames. However, this is still not always reliable, since the video sequences are likely to be dynamic and complex. In such circumstances, an object presents in one frame may have so complicated structures that texture synthesis techniques can create saw-shaped artifacts at the boundaries of the original region and the recovered part. Furthermore, the similar patches selected by aforementioned approach for synthesis are probably too small to satisfy larger holes. This is to say that it is not able to repair damaged videos with large missing areas.

Recently, Damianou *et al.* [15] developed a dynamical Gaussian process latent variable model (DGPLVM) to smoothly repair videos without saw-shaped artifacts in the presence of missing pixels. To achieve this, a Gaussian process (GP) prior based on auxiliary inducing points was introduced so that the variational Bayes approach was tractable. The latent variables were then variationally integrated out and a closed-form lower bound on a log likelihood function computed. The original purpose of inducing points in [16]–[21] was to speed up computation by regarding it as a smaller set representing the entire observed frames. In [15], the inducing point was also critical to obtaining a closed-form lower bound of the defined objective function to render the model robust to overfitting.

In [15], the latent variables are low dimensional representation of the observed data and there is a non linear mapping relationship between them. Meanwhile, the inducing points are regarded as a small subset that is supposed to cover an intact information of the observed data. Both of inducing points and latent variables play a pivotal role in the prediction stage of DGPLVM. However, it transpires that [15] tends to generate relatively similar inducing points that can introduce ghost effects in the recovered frames. This is because the inducing points and latent variables trained in DGPLVM have a tendency to focus on frequent shots or those with salient features within the scenes. We propose a dual dynamic Gaussian process latent variable model (D^{3}GPLVM) to address this problem, in which two diversity encouraging priors are applied to the inducing points and latent variables, respectively, so that they are more diverse and capture more distinct scenes from the observed complete frames. Since inducing points and latent variables are pivotal to predicting missing pixels, more resistant inducing points and latent variables enhance video repairing performance. However, directly integrating out latent variables is infeasible in our model since it is nonlinear. Thus, the variational inference approach is applied to approximate the marginal likelihood by maximizing a Jensen's lower bound and then integrate the latent variables. The parameters of D^{3}GPLVM are optimised by maximizing the Jensen's lower bound using scaled conjugate gradient (SCG). Afterwards, a given damaged video can be repaired by the trained D^{3}GPLVM regardless of its damaged area and form.

In the remainder of this paper, we first review the DGPLVM and then introduce our D^{3}GPLVM. To demonstrate the robustness of D^{3}GPLVM, we perform experiments on a **hundred** of various video clips from black and white films to modern movies.

II. Related Work

Completion approaches are broadly used for two fields: image inpainting and video repairing. In this section, we review works related to both of these fields.

A. Image Inpainting

Most inpainting approaches [22]–[29] exploit consistency in neighboring pixels or textures to recover the missing parts. However, these methods are inclined to fail when the missing region is inhomogeneous with its surroundings. Other methods [30]–[32] have attempted to deal with various real-world objects inpainting tasks by defining a similarity term between the input image and the base eigenvector derived by applying PCA to a subspace of the training samples. Based on [31] and [32], [30] was able to inpaint any object without first specifying the object class beforehand and could inpaint in real time. However, these methods are still imperfect since the damaged objects class for inpainting is much more diverse than expected. Instead of inpainting the images with missing pixels, some mehods aim to deblur the images. A case in point is [33].

Other works focus on face inpainting. In [34], the information on the neighboring unobscured regions are employed to select from the used database a number of similar faces which linearly represent input face with positive weights. Then, the occluded regions may be recovered by minimizing the error between the unobscured regions with its corresponding counterparts on similar faces. In [35] and [36], more complicated occlusion scenarios, like sun glasses, paintings on face, face behind fence, are taken into account. To achieve this, [35] detect the artifacts regions by modelling the original image using a normal distribution, while [36] works under the hypothesis that occlusions are large deviations from low dimensional representation [37] of a face. Afterwards, [35] eliminates the artifacts by incorporating the Poisson editing technique [38] into fast marching inpainting method whilst the Fields-of-Experts (FoE) model [39] is adopted by [36] to infer the missing face pixels.

In general, image inpainting exploits neighboring texture information to fill in the missing pixels. These inpainting techniques can work effectively under the circumstances that the missing pixels in an image is small. This is to say that image inpainting is not applicable to video repairing since video context can be complex and the damaged region can be too large. As a result, the recovered videos may have wrong contexts using image inpainting directly (see Fig. 5).

B. Video Repairing

The video repairing method proposed in [40] inferred occluded background and large motion by sampling and aligning movels (structured moving objects) from the captured video. Missing static background was repaired by constructing a layered mosaic in addition to image repairing [41]. To repair moving pixels, an optimal alignment based on a homographic transform was computed by assuming that the moving pixels were projections of cyclic motions [42], where cyclic motions were detected by time-frequency analysis [43]. As an extension of [40], [44] used tensor voting [45] to address the

pertinent spatio-temporal issues in background and motion repairing. Variable illumination and moving cameras were also studied. However, although [44] utilized texture synthesis to recover the occluded area, it is possible that the occluded region texture may not appear in the neighbouring pixels or frames, resulting in inaccurately repaired pixels with artifacts.

In [14], Wexler *et al.* attempted to fill in missing portions by sampling spatio-temporal patches from the input video. Their method defined and used similarity measurements in space and time domains. This method was successful due to the judicious extension of [46] in which non-parametric sampling was used to handle spatial and temporal information simultaneously. They demonstrated that the patches selected for completion may contain errors if the background is complex (e.g., non-textured) and the result will not preserve speed irregularities and may destroy complex structures.

Unlike image inpainting, these method unexceptionally exploited the global temporal information to achieve video repairing. However, this is still problematic since a video may include complex objects and scenes. In that case, simply searching and synthesizing similar patches from neighboring frames can introduce severe artifacts in the observed and recovered areas. Hence, for the scenes containing complicated objects, such methods may not generate a smooth repaired video by preserving the structure of objects.

III. Dynamical GPLVM

In this section, the basic concepts of dynamical Gaussian process latent variable model (DGPLVM) [15] are first introduced for better comprehension of our model.

In DGPLVM, $Y \in R^{n\times p}$ (with columns $\{y_{:,j}\}_{j=1}^{p}$) denotes the observed data where n is the number of data points and p is the dimensionality of each data point in Y. Here, these data are associated with latent variables $X \in R^{n\times q}$ for the sake of dimensionality reduction. Then the likelihood function is defined as:

$$p(Y \mid X) = \prod_{j=1}^{p} p(y_{:,j} \mid X), \tag{1}$$

where $y_{:,j}$ represents the j^{th} column of Y and

$$p(y_{:,j} \mid X) = \mathcal{N}(y_{:,j}|0, K_{ff} + \sigma^{-1} I_n). \tag{2}$$

Here, K_{ff} is a $n \times n$ kernel matrix and the kernel function here is an exponentiated quadratic (RBF) as follows:

$$k(x_{i,:}, x_{k,:}) = \sigma_f^2 \exp\Big(-\frac{1}{2}\sum_{j=1}^{q} \alpha_j (x_{i,j} - x_{k,j})^2\Big), \tag{3}$$

where each $x_{i,:}$ is the i^{th} row of X and $K_{ff} = k(x_{i,:}, x_{k,:})$. The purpose is to compute the marginal likelihood function:

$$p(Y) = \int p(Y \mid X) p(X) \mathrm{d}X. \tag{4}$$

Since Y is observed data which is supposed to be noisy, another latent variable F is introduced to be considered as the unnoisy version of Y. Here, the variable F has the same size as Y and

$$p(Y|F) = \prod_{j=1}^{p} \mathcal{N}(y_{:,j}|f_{:,j}, \sigma^{-1} I_n). \tag{5}$$

For the sake of dimensionality reduction, the conditional distribution in Eq.(1) becomes

$$p(F|X) = \prod_{j=1}^{p} \mathcal{N}(f_{:,j}|0, K_{ff}). \tag{6}$$

In DGPLVM, time sequencce is also taken into account so each datapoint $y_{i,:}$ is observed at corresponding time t_i. Therefore, the prior distribution of X is:

$$p(X) = \prod_{j=1}^{q} \mathcal{N}(x_{:,j}|0, K_x), \tag{7}$$

where $x_{:,j}$ refers to one column of X and $K_x = k(t_i, t'_i)$ is the covariance matrix obtained by evaluating the covariance function k on the observed times t.

Furthermore, [15] introduced the auxiliary inducing variable $U \in R^{m\times p}$, which is a set of m inducing points $u_{i,:} \in R^p$, evaluated at their associated inducing input locations $Z \in R^{m\times q}$(with columns $\{z_{:,j}\}_{j=1}^{q}$). Likewise,

$$p(U) = \prod_{j=1}^{p} N(0, K_{uu}). \tag{8}$$

Here, $K_{uu} = k(z_{i,:}, z_{k,:})$. The introduced inducing points can not only speed up computation but also render the objective function tractable. Interest readers may refer to [15] for more details.

IV. Dual Diversified Dynamical GPLVM

In DGPLVM, Damianou *et al.* proposed a Bayesian approach to train the GPLVM. By taking inducing points into consideration, the input variables of Gaussian process could be robustly integrated out. Then, the trained model was used to recover videos in the presence of missing pixels. However, the inducing points trained from DGPLVM were highly similar and could not fulfill the repairing of videos with more complex scenes.

Here, we propose the D^{3}GPLVM with repect to the repulsion property of inducing points and latent variables due to the fact that real world videos may have much more diverse scenes than expected. To be exact, given a video with an unknown number of damaged frames, let $Y \in R^{n\times p}$ denote all undamaged frames, where n and p are the number of undamaged frames and pixels in the video respectively. F is the noise-free version of Y, whilst X is the reduced dimension version of Y. Since a video is time sequential, the aforementioned time t_i is referred to as the frame serial number. The corresponding graphical model of D^{3}GPLVM is illustrated in Fig. 2 (a). Meanwhile, the whole repairing process is presented in the Fig. 2 (b)-(f).

Remember that the inducing points $U \in R^{m\times p}$ are a small set representing the entire undamaged frames. Meanwhile, the distribution $p(U)$ has a covariance matrix K_{uu} defined in Eq. (8). Here, a diversity prior of covariance matrix K_{uu} would be modeled by:

$$p(U \in Y) = |K_{uu}|, \tag{9}$$

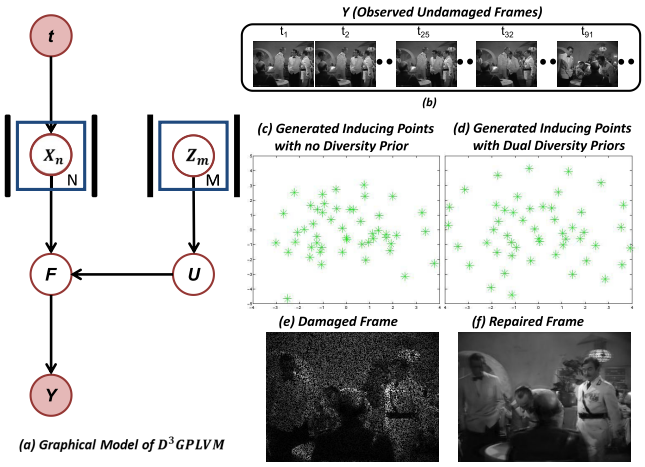

Fig. 2. A schematic of the proposed D^{3}GPLVM. (a) A graphical model of the proposed method. The observed variables are included in the red filled circles while the latent variables are within hollow circles. The blue rectangles with N and M mean that the numbers of X_n and Z_m in our model are N and M, respectively. Meanwhile, the bold black parallel lines outside X_n and Z_m refer to the added diversity encouraging priors. Here, the specific meaning of notations can be found in the Section III and IV. (b) The training frames. (c)-(d) A comparison of the generated inducing points with and without dual diversity encouraging priors. It is clear that the inducing points generated with diversity prior in (d) is more diverse than those in (c). (e) A random damaged frame. (f) The frame repaired using our method.

where $|K_{uu}|$ refers to the determinant of matrix K_{uu}. The inducing points selected with respect to such prior can cover multiple distinct scenes of a video instead of focusing on the most salient ones.

Likewise, another diversity encouraging prior applied to latent variables X is modeled by $|K_{ff}|$. Therefore, the new objective function is:

$$F(\theta) = \log P(Y) + \lambda_1 \log |K_{uu}| + \lambda_2 \log |K_{ff}|, \quad (10)$$

where $\theta = \{\sigma_f^2, \sigma, \alpha_1, \ldots, \alpha_j\}$ are the hyperparameters in our proposed model using the same symbols as in DGPLVM. Furthermore, $\lambda_1, \lambda_2 > 0$ is used to balance the weights between measurements of likelihood and the diversity encouraging prior.

A. Variational Inference

In our proposed model, the log likelihood function we wish to maximise is $F(\theta)$. Therefore,

$$F(\theta) = \log \int |K_{uu}|^{\lambda_1} |K_{ff}|^{\lambda_2} p(Y, F, U, X) \mathrm{d}X \mathrm{d}F \mathrm{d}U. \quad (11)$$

Here, the joint distribution $p(Y, F, U, X)$ can be further factorised as:

$$\begin{aligned} &p(Y, F, U, X) \\ &\quad = p(Y|F)p(F|U, X)p(U)p(X) \\ &\quad = \left(\prod_{j=1}^{p} p(y_{:,j}|f_{:,j}) p(f_{:,j}|u_{:,j}, X) p(u_{:,j}) \right) p(X). \end{aligned} \quad (12)$$

Now, the objective function $F(\theta)$ is:

$$\begin{aligned} F(\theta) = \log \int |K_{uu}|^{\lambda_1} |K_{ff}|^{\lambda_2} \prod_{j=1}^{p} p(y_{:,j}|f_{:,j}) \\ \times \Big(p(f_{:,j}|u_{:,j}, X) p(X) \mathrm{d}X \Big) p(u_{:,j}) \mathrm{d}U \mathrm{d}F. \end{aligned} \quad (13)$$

Note that integration over X is unfeasible since X is an input, in a rather complex non-linear manner, of $p(f_{:,j}|u_{:,j}, X)$, which contains the kernel matrix K_{ff}.

To see that, the specific form of $p(f_{:,j}|u_{:,j}, X)$ is derived based on Eq. (6) and Eq. (8):

$$\begin{aligned} p(F|U, X) &= \prod_{j=1}^{p} p(f_{:,j}|u_{:,j}, X) \\ &= \prod_{j=1}^{p} p(f_{:,j}|a_j, \Sigma_f), \end{aligned} \quad (14)$$

where $a_j = K_{fu} K_{uu}^{-1} u_{:,j}$, $\Sigma_f = K_{ff} - K_{fu} K_{uu}^{-1} K_{uf}$.

Thus, we use variational distribution $q(F, U, X)$ to approximate the true posterior $P(F, U, X|Y)$ with the form:

$$q(F, U, X) = \left(\prod_{j=1}^{p} p(f_{:,j}|u_{:,j}, X) q(u_{:,j}) \right) q(X). \quad (15)$$

Here, $q(X)$ is a variational distribution that follows:

$$q(X) = \prod_{i=1}^{n} N(x_{i,:}|\mu_{i,:}, S_i), \quad (16)$$

where each covariance matrix S_i is diagonal. Another variational distribution $q(U)$ is arbitrary and will be explained later. In terms of Jensen's inequality, the lower bound $F\big(q(X), q(U)\big)$ of the objective function could be derived by:

$$\begin{aligned} &F\big(q(X), q(U)\big) \\ &\quad = \int q(F, U, X) \log \frac{|K_{uu}|^{\lambda_1} |K_{ff}|^{\lambda_2} p(Y)}{q(F, U, X)} \mathrm{d}X \mathrm{d}F \mathrm{d}U. \end{aligned} \quad (17)$$

After inserting Eq. (15) into Eq. (17), we have:

$$\begin{aligned} &F\big(q(X), q(U)\big) \\ &= \int \prod_{j=1}^{p} p(f_{:,j}|u_{:,j}, X) q(u_{:,j}) q(X) \Bigg(\log |K_{uu}|^{\lambda_1} |K_{ff}|^{\lambda_2} \\ &\quad + \log \frac{\prod_{j=1}^{p} p(y_{:,j}|f_{:,j}) p(f_{:,j}|u_{:,j}, X) p(u_{:,j}) p(X)}{\prod_{j=1}^{p} p(f_{:,j}|u_{:,j}, X) q(u_{:,j}) q(X)} \Bigg) \\ &\quad \times \mathrm{d}X \mathrm{d}F \mathrm{d}U. \end{aligned} \quad (18)$$

By cancelling $p(f_{:,j}|u_{:,j}, X)$, we have:

$$\begin{aligned} &F\big(q(X), q(U)\big) \\ &= \int \prod_{j=1}^{p} p(f_{:,j}|u_{:,j}, X) q(u_{:,j}) q(X) \Bigg(\log |K_{uu}|^{\lambda_1} |K_{ff}|^{\lambda_2} \\ &\quad + \log \frac{\prod_{j=1}^{p} p(y_{:,j}|f_{:,j}) p(u_{:,j}) p(X)}{\prod_{j=1}^{p} q(u_{:,j}) q(X)} \Bigg) \mathrm{d}X \mathrm{d}F \mathrm{d}U. \end{aligned} \quad (19)$$

Since $K_{uu} = k(z_{i,:}, z_{k,:})$, the term $\lambda_1 \log|K_{uu}|$ can be placed outside the integral. However, $\lambda_2 \log|K_{ff}|$ contains variable X, so it is still inside the integral.

$$\begin{aligned} F\big(q(X), q(U)\big) = \lambda_1 \log|K_{uu}| + \lambda_2 \int q(X) \log|K_{ff}| \mathrm{d}X \\ + \int \prod_{j=1}^{p} p(f_{:,j}|u_{:,j}, X) q(u_{:,j}) q(X) \\ \times \log \prod_{j=1}^{p} p(y_{:,j}|f_{:,j}) \mathrm{d}X \mathrm{d}F \\ + \int \prod_{j=1}^{p} q(u_{:,j}) \log \frac{\prod_{j=1}^{p} p(u_{:,j})}{\prod_{j=1}^{p} q(u_{:,j})} \mathrm{d}U \\ - \int p(X) \frac{q(X)}{p(X)} \mathrm{d}X. \end{aligned} \tag{20}$$

According to the product formula of logarithm, the term $F\big(q(X), q(U)\big)$ becomes:

$$\begin{aligned} F\big(q(X), q(U)\big) = \lambda_1 \log|K_{uu}| + \lambda_2 \int q(X) \log|K_{ff}| \mathrm{d}X \\ + \int \prod_{j=1}^{p} p(f_{:,j}|u_{:,j}, X) q(u_{:,j}) q(X) \\ \times \sum_{j=1}^{p} \log p(y_{:,j}|f_{:,j}) \mathrm{d}X \mathrm{d}F \\ + \int \prod_{j=1}^{p} q(u_{:,j}) \sum_{j=1}^{p} \log \frac{p(u_{:,j})}{q(u_{:,j})} \mathrm{d}U \\ - \int p(X) \frac{q(X)}{p(X)} \mathrm{d}X. \end{aligned} \tag{21}$$

By applying integral operation, we have:

$$\begin{aligned} F\big(q(X), q(U)\big) = \lambda_1 \log|K_{uu}| + \lambda_2 \int q(X) \log|K_{ff}| \mathrm{d}X \\ + \int p(f_{:,j}|u_{:,j}, X) q(u_{:,j}) q(X) \\ \times \sum_{j=1}^{p} \log p(y_{:,j}|f_{:,j}) \mathrm{d}X \mathrm{d}f_{:,j} \\ + \int q(u_{:,j}) \sum_{j=1}^{p} \log \frac{p(u_{:,j})}{q(u_{:,j})} \mathrm{d}U \\ - \int p(X) \frac{q(X)}{p(X)} \mathrm{d}X. \end{aligned} \tag{22}$$

Let $\langle \cdot \rangle_p$ be a shorthand for expectation with respect to the distribution p, then:

$$\begin{aligned} & F(q(X), q(U)) \\ & = \lambda_1 \log|K_{uu}| + \lambda_2 \int q(X) \log|K_{ff}| \mathrm{d}X \\ & + \sum_{j=1}^{p} \Big(\int q(u_{:,j}) q(X) \langle \log p(y_{:,j}|f_{:,j}) \rangle_{p(f_{:,j}|u_{:,j},X)} \mathrm{d}X \mathrm{d}u_{:,j} \\ & + \langle \log \frac{p(u_{:,j})}{q(u_{:,j})} \rangle_{q(u_{:,j})} \Big) - \int q(X) \log \frac{q(X)}{p(X)} \mathrm{d}X. \end{aligned} \tag{23}$$

To simplify the sum term in the Eq. (23), let the formulation within the sum notation be denoted by:

$$\begin{aligned} & \hat{F}_j\big(q(X), q(U)\big) \\ & = \langle \log \frac{p(u_{:,j})}{q(u_{:,j})} \rangle_{q(u_{:,j})} \\ & + \int q(u_{:,j}) q(X) \langle \log p(y_{:,i}|f_{:,j}) \rangle_{p(f_{:,j}|u_{:,j},X)} \mathrm{d}X \mathrm{d}u_{:,j}. \end{aligned} \tag{24}$$

Thus, the lower bound $F\big(q(X), q(U)\big)$ may be re-expressed in the form:

$$\begin{aligned} F\big(q(X), q(U)\big) = \sum_{j=1}^{p} \hat{F}_j\big(q(X), q(U)\big) - KL\big(q(X)||p(X)\big) \\ + \lambda_1 \log|K_{uu}| + \lambda_2 \int q(X) \log|K_{ff}| \mathrm{d}X. \end{aligned} \tag{25}$$

Now, the lower bound $F\big(q(X), q(U)\big)$ consists of four parts: $\lambda_1 \log|K_{uu}|$, $\lambda_2 \int q(X) \log|K_{ff}| \mathrm{d}X$, $\hat{F}_j\big(q(X), q(U)\big)$ and $KL\big(q(X)||p(X)\big)$.

Here, $\lambda_1 \log|K_{uu}|$ is straightforward and the difficult part is to compute $\lambda_2 \int q(X) \log|K_{ff}| \mathrm{d}X$. In the optimisation stage, it requires to caculate the derivative of $|K_{ff}|$ which results in the inverse of matrix K_{ff}. It is too time consuming to compute K_{ff}^{-1} since it is a $N \times N$ matrix. Instead of calculating $\log|K_{ff}|$, its tight upper bound $tr(K_{ff} - I)$ is introduced here.

$$\begin{aligned} \lambda_2 \int q(X) \log|K_{ff}| \mathrm{d}X &\leq \lambda_2 \int q(X) tr(K_{ff} - I) \mathrm{d}X \\ &= \lambda_2 n(\sigma_f^2 - 1). \end{aligned} \tag{26}$$

Afterwards, we move on to compute $KL\big(q(X)||p(X)\big)$. Since both of $q(X)$ and $p(X)$ are Gaussian distributions, the KL term can be easily calculated:

$$KL\big(q(X)||p(X)\big) = \frac{1}{2} \sum_{i=1}^{n} tr\big(\mu_{i,:} \mu_{i,:}^T + S_i - log S_i\big) - \frac{nq}{2}. \tag{27}$$

So as to derive $\hat{F}_j\big(q(X), q(U)\big)$, it is essential to compute $\langle \log p(y_{:,i}|f_{:,j}) \rangle_{p(f_{:,j}|u_{:,j},X)}$.

$$\begin{aligned} & \langle \log p(y_{:,i}|f_{:,j}) \rangle_{p(f_{:,j}|u_{:,j},X)} \\ & = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log|\sigma^2 I_n| \\ & - \frac{1}{2} tr\Big(\sigma^{-2} I_n \big(y_{:,j} y_{:,j}^T - 2 y_{:,j} \langle f_{:,j} \rangle_{p(f_{:,j}|u_{:,j},X)} \\ & + \langle f_{:,j} f_{:,j}^T \rangle_{p(f_{:,j}|u_{:,j},X)} \big) \Big). \end{aligned} \tag{28}$$

After integrating out $f_{:,j}$, we have:

$$\begin{aligned} & \langle \log p(y_{:,i}|f_{:,j}) \rangle_{p(f_{:,j}|u_{:,j},X)} \\ & = \log N(y_{:,j}|K_{fu} K_{uu}^{-1} u_{:,j}, \sigma^2 I_n) \\ & - \frac{1}{2\sigma^2} tr(K_{ff} - K_{fu} K_{uu}^{-1} K_{uf}). \end{aligned} \tag{29}$$

After inserting Eq. (29) into $\hat{F}_j\big(q(X), q(U)\big)$, $\hat{F}_j\big(q(X), q(U)\big)$ can be easily derived and expressed as:

$$\begin{aligned}\hat{F}_j(q(X), q(U)) &= \frac{1}{2\sigma^2} tr\big(\langle K_{ff}\rangle_{q(X)}\big) - \frac{1}{2\sigma^2} tr\big(K_{uu}^{-1}\langle K_{uf}K_{fu}\rangle_{q(X)}\big) \\ &\quad + \int q(u_{:,j}) \log \frac{e^{\langle \log N(y_{:,j}|a_j, \sigma^2 I_n)\rangle_{q(X)}} p(u_{:,j})}{q(u_{:,j})} \mathrm{d}u_{:,j}. \end{aligned} \tag{30}$$

Note that there is a KL-like quantity in the Eq. (30), such that the optimal $q(u_{:,j})$ is supposed to be proportional to:

$$q(u_{:,j}) \propto e^{\langle \log N(y_{:,j}|a_j, \sigma^2 I_p)\rangle_{q(X)}} p(u_{:,j}). \tag{31}$$

By putting Eq.(8) back into Eq.(31), we have $q(u_{:,j}) = N(u_{:,j}|\mu_u, \Sigma_u)$ where:

$$\begin{aligned}\Sigma_u &= (\sigma^{-2} K_{uu}^{-1}\langle K_{uf}K_{fu}\rangle_{q(X)} K_{uu}^{-1} + K_{uu}^{-1})^{-1} \\ \mu_u &= \sigma^{-2}\Sigma_u K_{uu}^{-1}(\langle K_{fu}\rangle_{q(X)})^T y_{:,j}. \end{aligned} \tag{32}$$

$\hat{F}_j(q(X), q(U))$ can be upper bounded by $\hat{F}_j(q(X))$ after applying the *reversing Jensen's inequality* [47] to the KL-like quantity containing $q(u_{:,j})$:

$$\begin{aligned}\hat{F}_j(q(X)) = \frac{1}{2\sigma^2} tr\big(\langle K_{ff}\rangle_{q(X)}\big) - \frac{1}{2\sigma^2} tr\big(K_{uu}^{-1}\langle K_{uf}K_{fu}\rangle_{q(X)}\big) \\ + \log \int e^{<\log N(y_{:,j}|a_j, \sigma^2 I_p)>_{q(X)}} p(u_{:,j}) \mathrm{d}u_{:,j}. \end{aligned} \tag{33}$$

Now $q(U)$ is optimally eliminated, $\hat{F}_j(q(X))$ can be calculated as follows:

$$\begin{aligned}\hat{F}_j(q(X)) = \left[\log \frac{\sigma^{-n}|K_{uu}|^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}|\sigma^{-2}\psi_2 + K_{uu}|^{\frac{1}{2}}} e^{-\frac{1}{2} y_{:,j}^T W y_{:,j}}\right] \\ - \frac{\psi_0}{2\sigma^2} + \frac{1}{2\sigma^2} tr(K_{uu}^{-1}\psi_2), \end{aligned} \tag{34}$$

where $\psi_0 = tr\big(\langle K_{ff}\rangle_{q(X)}\big)$, $\psi_1 = \langle K_{fu}\rangle_{q(X)}$, $\psi_2 = \langle K_{uf}K_{fu}\rangle_{q(X)}$ and $W = \sigma^{-2} I_n - \sigma^{-4}\psi_1(\sigma^{-2}\psi_2 + K_{uu}^{-1})\psi_1^T$.

The Ψ statistics can be computed separately for each marginal $q(x_{i,:}) = N(x_{i,:}|\mu_{i,:}, S_i)$ taken from the full q(X). Thus, $\psi_0 = \sum_{i=1}^n \psi_0^i$ where:

$$\begin{aligned}\psi_0 &= \sum_{i=1}^n \int k(x_{i,:}, x_{i,:}) N(x_{i,:}|\mu_{i,:}, S_i) \mathrm{d}x_i \\ &= n\sigma_f^2, \end{aligned} \tag{35}$$

Further, Ψ_1 is an $n \times m$ matrix such that

$$\begin{aligned}(\Psi_1)_{i,k} &= \int k(x_{i,:}, (x_u)_{k,:}) N(x_{i,:}|\mu_{i,:}, S_i) \mathrm{d}x_{i,:} \\ &= \sigma_f^2 \prod_{j=1}^q \frac{\exp\Big(-\frac{1}{2}\frac{\alpha_j\big(u_{i,j} - (x_u)_{k,j}\big)}{\alpha_j S_{i,j} + 1}\Big)}{(\alpha_j S_{i,j} + 1)^{\frac{1}{2}}}, \end{aligned} \tag{36}$$

where $(x_u)_{k,:}$ denotes the kth row of X_u.

Finally, Ψ_2 is an $m \times m$ matrix that can be written as $\Psi_2 = \sum_{i=1}^n \Psi_2^i$ where Ψ_2^i is such that:

$$\begin{aligned}(\Psi_2^i)_{k,k'} &= \int k(x_{i,:}, (x_u)_{k,:}) k((x_u)_{k',:}, x_{i,:}) N(x_{i,:}|\mu_{i,:}, S_i) \mathrm{d}x_{i,:} \\ &= \sigma_f^4 \prod_{j=1}^q \frac{\exp\Big(-\frac{\alpha_j\big((x_u)_{k,j} - (x_u)_{k',j}\big)^2 - \frac{\alpha_j(u_{i,j} - \bar{x}_{:,j})^2}{2\alpha_j S_{i,j}}}{\alpha_j S_{i,j} + 1}\Big)}{(2\alpha_j S_{i,j} + 1)^{\frac{1}{2}}}, \end{aligned} \tag{37}$$

where $\bar{x}_{:,j} = \frac{\big((x_u)_{k,j} + (x_u)_{k',j}\big)}{2}$.

With $\hat{F}_j\big(q(X), q(U)\big)$ and $KL\big(q(X)||p(X)\big)$ in hand, we can optimize the parameters θ in our model using a gradient-based algorithm.

B. Repairing Damaged Video

A set of partially observed frames $Y_* = \{Y_*^u, Y_*^o\}$ in a video sequence is given in the prediction stage. Here, Y_*^o denotes the observed part in the video frames and Y_*^u refers to the missing part. Our task is to calculate the following predictive density:

$$P(Y_*|Y) \approx \int P(Y_*|F_*) q(F_*|X_*) q(X_*) \mathrm{d}X_* \mathrm{d}F_*. \tag{38}$$

Like F and X, F_* and X_* are namely the latent variables of new testing data Y_*. To compute above, we need to optimise with respect to the parameters (u_*, S_*) of the Gaussian variational distribution $q(X_*)$. The standard GP prediction is employed here to obtain $q(X_*)$ which can be further factorized as follows:

$$\begin{aligned}q(X_*) &= \int p(X_*|X) q(X) \mathrm{d}X \\ &= \prod_{j=1}^q \int p(x_{*,j}|x_{:,j}) q(x_{:,j}) \mathrm{d}x_{:,j}, \end{aligned} \tag{39}$$

where the variational distribution $q(X)$ is already known and $p(X_*|X)$ can be derived from the *conditional GP prior* [48].

According to the predictive density, we need to compute $q(F_*|X_*)$ now. By following Eq. (15), $q(F_*|X_*)$ is:

$$q(F_*|X_*) = \prod_{j=1}^p \int p(f_{:,j}|u_{:,j}, X) q(u_{:,j}) \mathrm{d}u_{:,j}. \tag{40}$$

Note that $p(f_{:,j}|u_{:,j}, X)$ is already known in training stage and the specific expression of variational distribution $q(u_{:,j})$ is given in Eq. (31).

So, to predict Y_*, we need to first predict its latent function F_* according to:

$$q(F_*) = \int q(F_*|X_*) q(X_*) \mathrm{d}X_*. \tag{41}$$

Clearly, the specific expression of $p(f_{:,j}|u_{:,j}, X)$ is given in Eq. 8 and $q(u_{:,j})$ is already optimised in training phase. So the

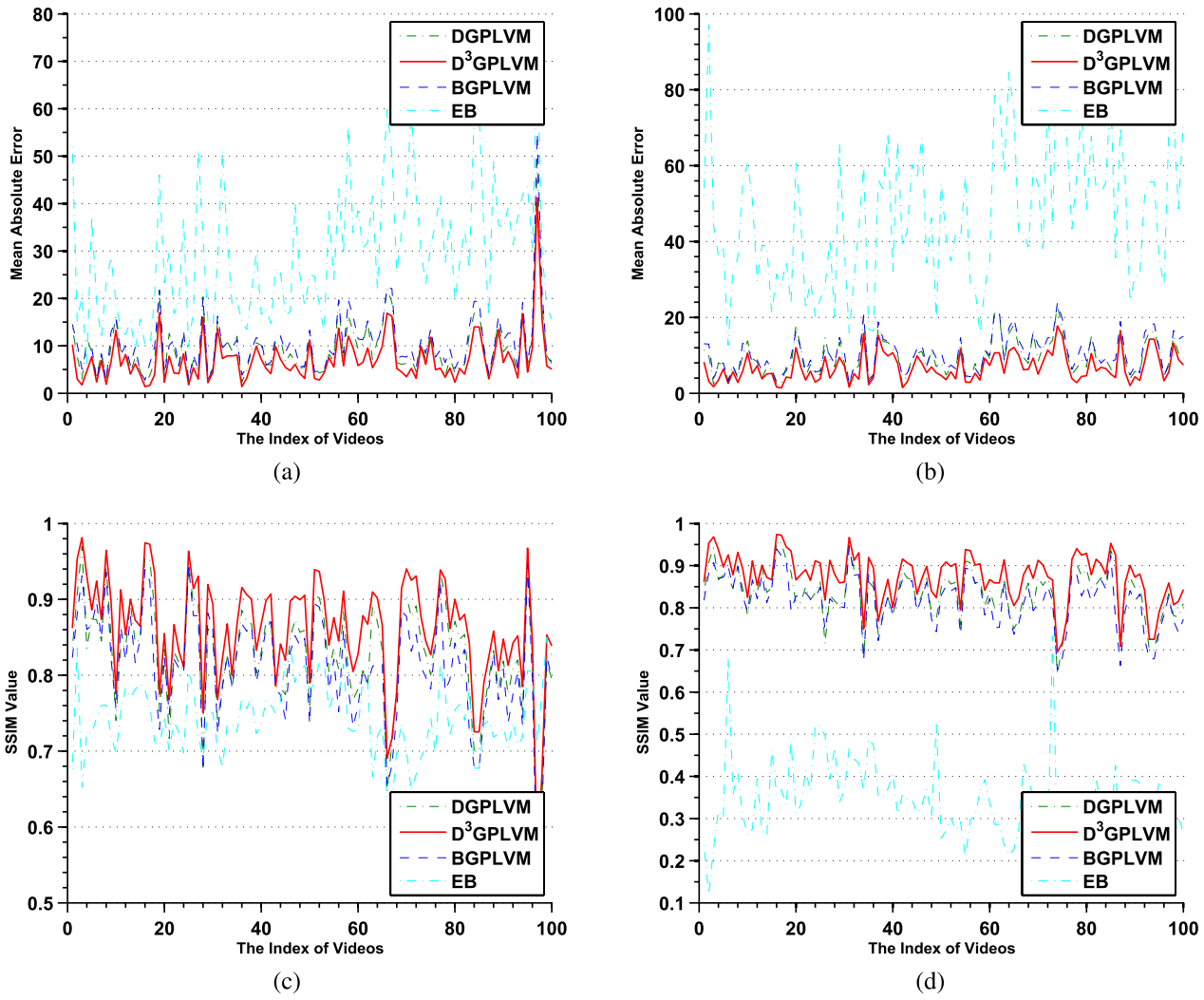

Fig. 3. The performances of four different methods for video repairing. (a)-(c) MAEs and SSIMs of block damage repairing in videos. (b)-(d) MAEs and SSIMs of scattered damage repairing in videos. The red line is our method. The smaller the MAE value is, the more similar the video frames between repaired video and ground truth, and vice versa for SSIM.

expression of $q(F_*|X_*)$ is:

$$q(F_{u*}|X_*) = N\big(F_{u*}|K_{*u}B, K_{**}K_{*u}[K_{uu}^{-1} - (K_{uu} + \sigma^{-2}\Psi_2)^{-1}]K_{*u}^T\big), \quad (42)$$

where B=$\sigma^{-2}(K_{uu}+\sigma^{-2}\Psi_2)^{-1}\Psi_1^T Y$, $K_{**} = k_f(X_*, X_*)$ and $K_{*u} = K(X_*, z_{i,:})$. By substituting $q(F_*|X_*)$ and $q(X_*)$ back into Eq.38 and using the fact that both of them are Gaussian. Now, the mean is $E[F_*] = B^T\Psi_1^*$ and the covariance is:

$$Cov(F_*) = \sigma^2 I + B^T\Big(\Psi_2^* - \Psi_1^*(\Psi_1^*)^T\Big)B + \Psi_0^* I - tr\Big((K_{uu}^{-1} - (K_{uu} + \sigma^{-2}\Psi_2)^{-1})\Psi_2^*\Big)I, \quad (43)$$

where $B = \sigma^{-2}(K_{uu} + \sigma^{-2}\Psi_2)^{-1}\Psi_1^T Y$, $\psi_0^* = tr(K_{uu})$, $\Psi_1^* = \langle K_{u*}\rangle$ and $\Psi_2^* = \langle K_{u*}K_{u*}^T\rangle$. Notice that the Ψ statistics involving the test latent variable x_* appear naturally in these expressions. Using the above expressions, the predicted mean of Y_* is equal to $E(F_*)$ and the predicted covariance is equal to $Cov(F_*) + \sigma^{-1}I$.

V. Results

In this section, we conduct experiments on movie clips from the Hollywood dataset.[1] The Hollywood dataset offers various kinds of movie clips, like actions, loving story, for our experiments. Meanwhile, these clips normally range from a few seconds to several minutes. In this dataset, 100 movie clips are selected for evaluation of our proposed video repairing model.

Since D^{3}GPLVM requires a certain number of frames for training, the clips used for testing were at least seven seconds in length. For each sequence, 40 percent of the frames were randomly selected to generate artificial damage. To be precise, we assumed that half the pixels in one frame were missing, and providing two kinds of damage for testing: block damage and scattered damage. These two kinds of damage are the most common evaluation approaches in the video repairing experiments [14], [15], [44], [49]. More specifically, block damage was generated by cutting off either the left/right or top/bottom half part of one frame. Furthermore, a number

[1] http://www.di.ens.fr/~laptev/actions/hollywood2/

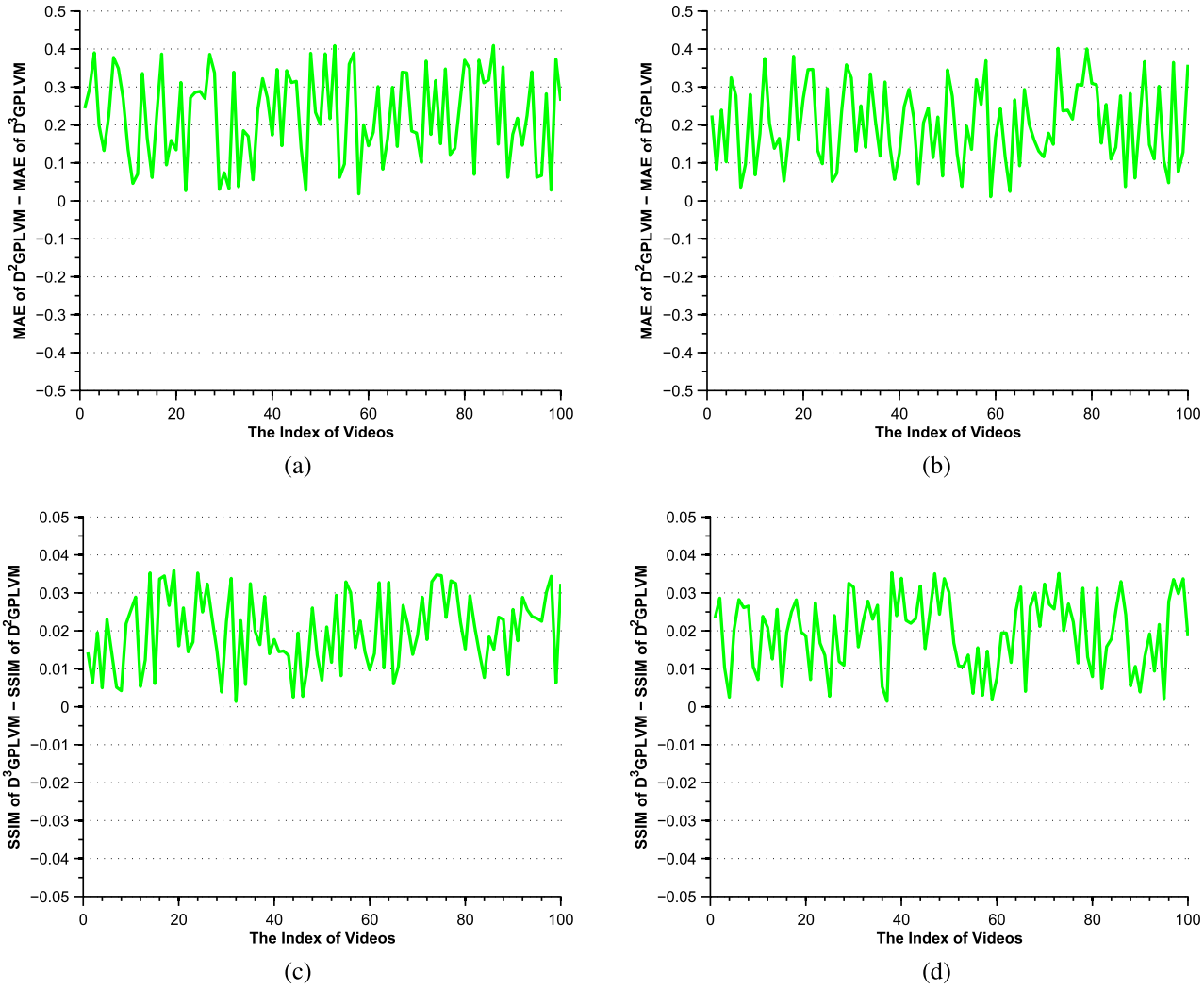

Fig. 4. A comparison of diversified dynamical Gaussian process latent variable model (D^{2}GPLVM using single diversity encouraging prior) and D^{3}GPLVM (using dual diversity encouraging priors). For better illustration, we show the difference on both of the MAE values and SSIM values from D^{2}GPLVM and D^{3}GPLVM respectively.

of missing pixels were distributed across the whole frame to produce scattered damage. Our method was performed on these 100 movie clips in comparison with three classical methods: Bayesian Gaussian process latent variable model (BGPLVM) [50], dynamical Gaussian process latent variable model (DGPLVM) [15] and exemplar based inpainting (EB) [12]. Here, BGPLVM and DGPLVM are the variants of Gaussian process latent variable model (GPLVM). Moreover, BGPLVM aims to solve GPLVM based on the variational Bayesian method, and DGPLVM incorporates BGPLVM with a dynamical prior by considering the sequential dependence between video frames. Additionally, EB method attempts to repair the missing area of video frames by synthesizing similar texture patches from the neighbours of a missing hole.

To demonstrate the robustness of our approach, we tested both black-and-white and color films. Since the test sequence resolution was not fixed, it was not easy to obtain specific times used for training and repairing. However, in general, it took about one minute for training and around two seconds for one frame repairing. The code was run on Matlab 2014a on a computer configured with a 3.2GHz CPU and 8GB memory. Of note, our approach maintained a similar execution time for repairing as DGPLVM and BGPLVM, but was much faster than exemplar-based inpainting, which took over 15 minutes for entire video inpainting. **Note that the parameter λ_1 and λ_2 in our model controls the diversity of the optimized inducing points and latent variables, which are manually set as 0.01 and 0.1 respectively for all 100 videos here. The results can be further improved if the cross-validation method is employed**.

Here, mean absolute error (MAE) and structural similarity (SSIM) [51]–[53] were calculated to assess the fidelity of recovered frames on all clips. Generally, lower MAE value means that the repaired video is closer to the ground truth. In contrast, higher SSIM value indicates better repaired results.

We evaluate both of the MAE and SSIM for each frame in all videos. The MAE and SSIM values for one clip were

Fig. 5. A schematic of repairing videos with block damage. Two recovered frames of the four video sequences are randomly displayed. Our method consistently outperforms the others and provides smooth and clearly recovered results.

TABLE I

AVERAGE MAE AND SSIM OF PLOTS (A), (B), (C), (D) IN FIG. 3 RESPECTIVELY

Method	Avg MAE (lower is better)		Avg SSIM (higher is better)	
	Scattered Missing	Block Missing	Scattered Missing	Block Missing
BGPLVM	10.685	11.845	0.827	0.887
DGPLVM	9.653	10.734	0.844	0.903
Exemplar Based	29.261	52.871	0.763	0.368
D^{3}GPLVM	**7.175**	**7.672**	**0.890**	**0.965**

TABLE II

COMPARISON OF PERFORMANCES OF EXPLOITING SINGLE AND DUAL DIVERSITY ENCOURAGING PRIORS

Num of Priors Exploted in Proposed Model	Avg MAE (lower is better)		Avg SSIM (higher is better)	
	Scattered Missing	Block Missing	Scattered Missing	Block Missing
Single Diversity Prior on Inducing Points	7.335	7.908	0.877	0.945
Dual Diversity Prior	**7.175**	**7.672**	**0.890**	**0.965**

subsequently obtained by averaging the sum of the MAEs and SSIMs for all frames. Exemplar based inpainting consistently showed the worst performance (Fig. 3). DGPLVM incorporates the dynamical sequential prior into BGPLVM, rendering it more competent than BGPLVM for time sequence based video repairing. By taking the diversity encouraging prior into account, our method drastically enhanced damaged video recovery (Fig. 3). Moreover, so as to further clearly

Fig. 6. A schematic of repairing video with scattered damage. Two recovered frames of the two video sequences are randomly displayed.

demonstrate the superiority of our method, we also computed the average value of MAEs and SSIMs in the plots of (a), (b), (c), (d). This can be seen in Table I that our method unexceptionally significantly outperforms other methods in respect of the MAE and SSIM measurements.

The dual diversity priors were defined on the inducing points and latent variables in our model, so we made a comparison of performance of models between using single diversity prior on inducing points and dual diversity priors respectively. More precisely, we performed video repairing on the same dataset with single and dual diversity priors in turn. Likewise, MAE and SSIM measurements are still essential to reflect the performance of repairing with single and dual diversity priors. Apparently, in Fig. 4 by applying dual diversity priors, the repairing effect substantially enhanced in comparsion with single prior. Also, the average value of the MAE and SSIM values in Fig. 4 are illustrated in Table II. Undoutedly, it is more necessary and better to apply dual diversity priors rather than a single one.

So as to show the repairing effect of our proposed model, more examples of video repairing in several video sequences are present in Fig. 5 and Fig. 6. Specifically, Fig. 5 illustrates repairing of videos with block damage and Fig. 6 displays repairing of scatted missing pixels. Also, five distinct methods were ran here and the original frames with half the pixels missing are shown in the first column, while the next five columns are a comparison with the other methods and the last column is the ground truth. As mentioned above, since exemplar based inpainting only exploits local image features for texture synthesis, it is likely to introduce errors during repairing and performs worst. Without special constraints, BGPLVM and DGPLVM tend to generate ghost effects in the recovered frames because the inducing points created by BGPLVM and DGPLVM are usually more similar than diverse. The effects between D^{2}GPLVM (single diversity prior) and D^{3}GPLVM may not be obvious here, but we have demonstrated the superiority of D^{3}GPLVM in the above quantative experiments. Consequently, these methods fail to extract comprehensive distinct scenes to better recover damaged videos, especially those with constantly changing shots, a dynamic background and/or moving objects. Our method offers stable repairing and outperforms the other approaches in different video scenarios.

VI. Conclusion

This paper presents D^{3}GPLVM for video repairing by simultaneously exploring dynamic and diversity properties under the GPLVM framework. Compared to DGPLVM, the diversity encouraging priors are essential in our model to extract more distinct features from the observed training frames. Meanwhile, our method has the inherent advantage of recovering incomplete frames with more complex sceneries since these types of video usually have more diverse characteristics that cannot be captured by traditional DGPLVM. Instead of using local neighboring texture synthesis, our probabilistic model utilizes global temporal information to recover smoother frames with large missing holes. Finally, our model reformulates the objective function of traditional DGPLVM and introduces a lower bound of the objective function by variational inference. Therefore, the objective function of our model is analytically tracktable by maximizing its variational lower bound. We illustrate the effectiveness of our method by comparing it with a number of other methods on a movie dataset of 100 various video clips.

References

[1] R. Hong, M. Wang, Y. Gao, D. Tao, X. Li, and X. Wu, "Image annotation by multiple-instance learning with discriminative feature mapping and selection," *IEEE Trans. Cybern.*, vol. 44, no. 5, pp. 669–680, May 2014.

[2] R. Hong, Y. Yang, M. Wang, and X.-S. Hua, "Learning visual semantic relationships for efficient visual retrieval," *IEEE Trans. Big Data*, vol. 1, no. 4, pp. 152–161, Dec. 2015.

[3] F. Zhu and L. Shao, "Weakly-supervised cross-domain dictionary learning for visual recognition," *Int. J. Comput. Vis.*, vol. 109, nos. 1–2, pp. 42–59, 2014.

[4] T. Liu and D. Tao, "Classification with noisy labels by importance reweighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 447–461, Mar. 2016.
[5] C. Xu, T. Liu, D. Tao, and C. Xu, "Local rademacher complexity for multi-label learning," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1495–1507, Mar. 2016.
[6] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 12, pp. 2531–2544, Dec. 2015.
[7] M. Yu, L. Liu, and L. Shao, "Structure-preserving binary representations for RGB-D action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2015.2491925.
[8] D. Wu and L. Shao, "Deep dynamic neural networks for gesture segmentation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2016.2537340.
[9] L. Shao, L. Liu, and M. Yu, "Kernelized multiview projection for robust action recognition," *Int. J. Comput. Vis.*, to be published, doi: 10.1007/s11263-015-0861-62016.
[10] L. Liu, L. Shao, F. Zheng, and X. Li, "Realistic action recognition via sparsely-constructed Gaussian processes," *Pattern Recognit.*, vol. 47, no. 12, pp. 3819–3827, Dec. 2014.
[11] C. Ding, C. Xu, and D. Tao, "Multi-task pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 980–993, Mar. 2015.
[12] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.
[13] R. Hong, L. Zhang, and D. Tao, "Unified photo enhancement by discovering aesthetic communities from Flickr," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1124–1135, Mar. 2016.
[14] Y. Wexler, E. Shechtman, and M. Irani, "Space-time video completion," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun./Jul. 2004, pp. 120–127.
[15] A. C. Damianou, M. K. Titsias, and N. D. Lawrence. (Sep. 2014). "Variational inference for uncertainty on the inputs of Gaussian process models." [Online]. Available: http://arxiv.org/abs/1409.2287
[16] L. Csató, "Gaussian processes: Iterative sparse approximations," Ph.D. dissertation, Dept. Electron. Eng. Comput. Sci., Aston University, Birmingham, U.K., 2002.
[17] L. Csató and M. Opper, "Sparse on-line Gaussian processes," *Neural Comput.*, vol. 14, no. 3, pp. 641–668, 2002.
[18] M. Seeger, C. Williams, and N. Lawrence, "Fast forward selection to speed up sparse Gaussian process regression," in *Proc. 9th Int. Workshop Artif. Intell. Statist.*, 2003.
[19] E. Snelson and Z. Ghahramani, "Sparse Gaussian processes using pseudo-inputs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1257–1264.
[20] J. Quiñonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *J. Mach. Learn. Res.*, 2005, pp. 1939–1959.
[21] M. K. Titsias, "Variational learning of inducing variables in sparse Gaussian processes," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2009, pp. 567–574.
[22] I. Drori, D. Cohen-Or, and H. Yeshurun, "Fragment-based image completion," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 303–312, 2003.
[23] J. Sun, L. Yuan, J. Jia, and H.-Y. Shum, "Image completion with structure propagation," in *Proc. ACM SIGGRAPH*, 2005, pp. 861–868.
[24] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. ACM SIGGRAPH*, 2000, pp. 417–424.
[25] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 882–889, Aug. 2003.
[26] Y.-W. Wen, R. H. Chan, and A. M. Yip, "A primal–dual method for total-variation-based wavelet domain inpainting," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 106–114, Jan. 2012.
[27] R. H. Chan, Y.-W. Wen, and A. M. Yip, "A fast optimization transfer algorithm for image inpainting in wavelet domains," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1467–1476, Jul. 2009.
[28] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1153–1165, May 2010.
[29] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.
[30] T. Hosoi, K. Kobayashi, K. Ito, and T. Aoki, "Fragment-based image completion," in *Proc. IEEE Int. Conf. Image Process.*, 2011, pp. 3441–3444.
[31] T. Amano and Y. Sato, "Image interpolation using BPLP method on the eigenspace," *Syst. Comput. Jpn.*, vol. 38, no. 1, pp. 87–96, 2007.
[32] T. Amano, "Image interpolation by high dimensional projection based on subspace method," in *Proc. 17th Int. Conf. Pattern Recognit.*, Aug. 2004, pp. 665–668.
[33] H. Liu, X. Sun, L. Fang, and F. Wu, "Deblurring saturated night image with function-form kernel," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4637–4650, Nov. 2015.
[34] Z. Mo, J. P. Lewis, and U. Neumann, "Face inpainting with local linear representations," in *Proc. Brit. Mach. Vis. Conf.*, 2004, pp. 347–356.
[35] A. Sobiecki, A. Telea, G. Giraldi, L. A. P. Neves, and C. E. Thomaz, "Low-cost automatic inpainting for artifact suppression in facial images," in *Proc. Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2013, pp. 1–10.
[36] R. Min and J.-L. Dugelay, "Inpainting of sparse occlusion in face recognition," in *Proc. IEEE Int. Conf. Image Process.*, Sep./Oct. 2013, pp. 1425–1428.
[37] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, 2001, Art. no. 11.
[38] P. Pérez, M. Gangnet, and M. Blake, "Poisson image editing," in *Proc. ACM SIGGRAPH*, 2003, pp. 313–318.
[39] S. Roth and M. J. Black, "Fields of experts," *Int. J. Comput. Vis.*, vol. 82, no. 2, pp. 205–229, Apr. 2009.
[40] J. Jia, W. Tai-Pang, Y.-W. Tai, and C.-K. Tang, "Video repairing: Inference of foreground and background under severe occlusion," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, Jun./Jul. 2004, pp. 364–371.
[41] J. Jia and C.-K. Tang, "Inference of segmented color and texture description by tensor voting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 771–786, Jun. 2004.
[42] S. M. Seitz and C. R. Dyer, "View-invariant analysis of cyclic motion," *Int. J. Comput. Vis.*, vol. 25, no. 3, pp. 231–251, 1997.
[43] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 781–796, Aug. 2000.
[44] J. Jia, Y.-W. Tai, T.-P. Wu, and C.-K. Tang, "Video repairing under variable illumination using cyclic motions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 832–839, May 2006.
[45] G. Medioni, M.-S. Lee, and C.-K. Tang, *A Computational Framework for Feature Extraction and Segmentation*. Amsderstam, The Netherlands: Elsevier, 2000.
[46] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. 7th Int. Conf. Comput. Vis.*, 1999, pp. 1033–1038.
[47] N. J. King and N. D. Lawrence, "Fast variational inference for Gaussian process models through KL-correction," in *Proc. Eur. Conf. Mach. Learn.*, 2006, pp. 270–281.
[48] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
[49] M. Zhou *et al.*, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 130–144, Jan. 2012.
[50] M. K. Titsias and N. D. Lawrence, "Bayesian Gaussian process latent variable model," in *Proc. 13th Int. Workshop Artif. Intell. Statist.*, 2010, pp. 844–851.
[51] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[52] K. Ma, T. Zhao, K. Zeng, and Z. Wang, "Objective quality assessment for color-to-gray image conversion," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4673–4685, Dec. 2015.
[53] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.

Hao Xiong received the master's degree in information technology from the University of Sydney in 2012. He is currently pursuing the Ph.D. degree with the Centre for Quantum Computation and Intelligent Systems, University of Technology Sydney, Ultimo, NSW, Australia. His research interests include latent variable models, image processing, and multiview geometry.

Tongliang Liu received the B.E. degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2012. He is currently pursuing the Ph.D. degree in computer science with the University of Technology Sydney, Ultimo, NSW, Australia. His research interests include statistical learning theory, computer vision, and optimization. He received the best paper award in the IEEE International Conference on Information Science and Technology 2014.

Dacheng Tao (F'15) is a Professor of Computer Science with the Centre for Quantum Computation and Intelligent Systems, and the Faculty of Engineering and Information Technology, University of Technology Sydney. He mainly applies statistics and mathematics to data analytics problems. His research interests spread across computer vision, data science, image processing, machine learning, and video surveillance. His research results have expounded in one monograph and 200+ publications at prestigious journals and prominent conferences, such as the IEEE T-PAMI, T-NNLS, T-IP, JMLR, IJCV, NIPS, ICML, CVPR, ICCV, ECCV, AISTATS, ICDM, and ACM SIGKDD. He received several best paper awards, such as the Best Theory/Algorithm Paper Runner Up Award in the IEEE ICDM07, the Best Student Paper Award in the IEEE ICDM13, and the 2014 ICDM 10-Year Highest-Impact Paper Award. He received the 2015 Australian Scopus-Eureka Prize, the 2015 ACS Gold Disruptor Award, and the 2015 UTS Vice-Chancellors Medal for Exceptional Research. He is a fellow of the OSA, IAPR, and SPIE.

Heng Tao Shen received the B.Sc. (Hons.) and Ph.D. degrees from the Department of Computer Science, National University of Singapore, in 2000 and 2004, respectively. He is a Professor of Computer Science and an ARC Future Fellow with the School of Information Technology and Electrical Engineering, The University of Queensland. He then joined The University of Queensland as a Lecture, Senior Lecture, Reader, and became a Professor in 2011. His research interests mainly include multimedia/mobile/web search, and big data management on spatial, temporal, multimedia, and social media databases.