Short Papers

A Kronecker Product Model for Repeated Pattern Detection on 2D Urban Images

Juan Liu, Emmanouil Z. Psarakis [ID], Yang Feng [ID], and Ioannis Stamos

Abstract—Repeated patterns (such as windows, balconies, and doors) are prominent and significant features in urban scenes. Therefore, detection of these repeated patterns becomes very important for city scene analysis. This paper attacks the problem of repeated pattern detection in a precise, efficient and automatic way, by combining traditional feature extraction with a Kronecker product based low-rank model. We introduced novel algorithms that extract repeated patterns from rectified images with solid theoretical support. Our method is tailored for 2D images of building façades and tested on a large set of façade images.

Index Terms—Repeated pattern detection, low-rank, Kronecker product model, urban façade

◆

1 INTRODUCTION

URBAN scenes contain rich periodic or near-periodic structures, such as windows, doors, and other architectural features. Detection of periodic structures is useful in many applications such as photo-realistic 3D reconstruction, 2D-to-3D alignment, façade parsing, city modeling, classification, navigation, visualization in 3D map environments, shape completion, cinematography and 3D games. However, it is a challenging task due to scene occlusion, varying illumination, pose variation and sensor noise.

A pre-processing rectification step is common to all façade parsing algorithms. This work also relies on such a pre-processing step based on the methods presented in [1] and [2]. Our pipeline focuses on the detection of repeated patterns in rectified facade images (Sections 3 and 4). State-of-the-art methods use classification [3], statistical and grammar-based approaches [4] and [5], as well as feature-based symmetry [6]. We, on the other hand, provide a novel detection method to model repetition as a Kronecker product.

2 RELATED WORK

In recent years, repeated pattern or periodic structure detection has received significant attention in both 2D images ([4], [7]) and 3D point clouds ([8], [9]). Repeated patterns are usually hypothesized from the matching of local image features. They can be modeled as a set of sparse repeated features [10] in which the crystallographic group theory [11] was employed. [12] maximizes local symmetries and separates different repetition groups via evaluating the local repetition quality conditionally for different repetition intervals.

- *J. Liu is with Google Inc, New York, NY 10011. E-mail: juanliu.ustc@gmail.com.*
- *E. Z. Psarakis is with Computer Engineering & Informatics, University of Patras, Patras 26504, Greece. E-mail: psarakis@ceid.upatras.gr.*
- *Y. Feng is with Statistics, Columbia University, New York, NY 10027. E-mail: yang.feng@columbia.edu.*
- *I. Stamos is with Computer Science, Hunter College of CUNY, New York, NY 10021. E-mail: istamos@hunter.cuny.edu.*

Muller et al. [13] proposes an approach to detect symmetric structures in a rectified fronto-façade and to reconstruct a 3D geometric model. The work of [3] describes a method for periodic structure detection upon the pixel-classification results of a rectified façade. Shape grammars have also been used for 2D façade parsing [4]. Other grammar-based approaches include [14].

Moreover, [15] and [5] addressed the façade parsing problem using shape grammar learning. More recently, [16] proposed a parsing method using dynamic programming. Inspired by [16], [6] developed a method to leverage symmetry and repetitions for façade parsing with occlusions.

All the above-mentioned methods require image rectification as a pre-processing step. To solve this problem, low-rank methods were used and attracted a lot of attention in recent years [2]. A similar work was proposed by [17] in which the rank value N is assumed known. Another method for the recovery of both low-rank and the sparse components is presented in [18]. Finally, [19] describes a low-rank based method that detects the repeated patterns in 2D images for the application of shape completion.

3 FAÇADE MODELING VIA KRONECKER PRODUCTS

In this section we describe a Kronecker product modeling approach which is applied on a rectified façade image. It is a novel representation that describes a large subset of façade examples.

3.1 Ideal Façade Modeling

Let us consider the partition of all-ones matrix $\mathbf{1}_{l_v \times l_h}$ of size $l_v \times l_h$ by using the following mutually exclusive $1-0$ matrices $\mathbf{M}_k$, $k=1,$ $2, \cdots,$ K of size $l_v \times l_h$:

$$< vec\{\mathbf{M}_k\},\ vec\{\mathbf{M}_l\} > = \begin{cases} ||vec\{\mathbf{M}_k\}||_0, & k=l, \\ 0, & k \neq l, \end{cases} \quad (1)$$

$$\sum_{k=1}^{K} \mathbf{M}_k = \mathbf{1}_{l_v \times l_h}, \quad (2)$$

where $vec\{\mathbf{X}\}$, $< \mathbf{x},\ \mathbf{y} >$ and $||\mathbf{x}||_0$ denote the *column-wise* vectorization of matrix $\mathbf{X}$, the inner product of vectors $\mathbf{x},\ \mathbf{y}$ and the l_0 *norm* of vector $\mathbf{x}$, respectively. As it is clear from Eqs. (1)–(2), different choices of matrices $\mathbf{M}_k$ result in different partitions of matrix $\mathbf{1}_{l_v \times l_h}$. Let us now associate with each component $\mathbf{M}_k$, $k=1,\ 2,\ \ldots,\ K$ of the partition of matrix $\mathbf{1}_{l_v \times l_h}$ defined in Eq. (2), a 2-D pattern $\mathbf{P}_k$ of size $N_v \times N_h$ that is going to be repeated according to $\mathbf{M}_k$. The patterns should have a piecewise constant surface form, with example structures including windows, doors and balconies.

We now define a subset of building façades that can be expressed as a sum of Kronecker products:

$$\mathcal{F}_{N \times M} = \sum_{k=1}^{K} \lambda_k (\mathbf{M}_k \otimes \mathbf{P}_k), \quad (3)$$

where $\mathbf{X} \otimes \mathbf{Y}$ is the *Kronecker* product of matrices $\mathbf{X},\ \mathbf{Y}$ and λ_k, $k= 1, 2, \cdots, K$ are weights. Here, $N \times M$ is the size of the urban building façade image. It is obvious that $N = l_v N_v$ and $M = l_h N_h$. The urban building façade's model defined in Eq. (3) can be used even in cases where there is no periodic structure.

Generalizing Eq. (3) to include a "wall" gray level λ_0, we get:

$$\mathcal{F}_{N \times M} = \lambda_0 \mathbf{1}_{N \times M} + \sum_{k=1}^{K} \lambda_k (\mathbf{M}_k \otimes \mathbf{P}_k). \quad (4)$$

Using the fact that the components of the partition of array $\mathbf{1}_{l_v \times l_h}$ of Eq. (2) are mutually exclusive, we rewrite Eq. (4) as:

$$\mathcal{F}_{N \times M} = \sum_{k=1}^{K} \lambda_k (\mathbf{M}_k \otimes \hat{\mathbf{P}}_k), \, \hat{\mathbf{P}}_k = \mathbf{P}_k + \frac{\lambda_0}{\lambda_k} \mathbf{1}_{N_v \times N_h}, \quad (5)$$

where $\hat{\mathbf{P}}_k$ are modified patterns as defined above.

3.2 Façade Model Approximation

In this section we would like to compute the components of the Kronecker product that generate an ideal (i.e. noise-free) building façade $\mathbf{F}_{N \times M} \in \mathbb{R}^{N \times M}$ with $N = l_v N_v$ and $M = l_h N_h$. Using the model of Eq. (5) we can define the following cost function:

$$\mathcal{C}_F(\mathbf{M}_k, \, \hat{\mathbf{P}}_k, \, \lambda_k, \, k = 1, \cdots, K) = ||\mathbf{F}_{N \times M} - \mathcal{F}_{N \times M}||_2^2$$
$$= \left\| \mathbf{F}_{N \times M} - \sum_{k=1}^{K} \lambda_k (\mathbf{M}_k \otimes \hat{\mathbf{P}}_k) \right\|_2^2, \quad (6)$$

where $|| \cdot ||_2$ represents the Frobenius norm of a matrix. As it is clear from its definition, $\mathcal{C}_F(.)$ is a Frobenius norm based cost function that quantifies the error between the given matrix $\mathbf{F}_{N \times M}$ and the model $\mathcal{F}_{N \times M}$.

Therefore, the modeling problem of the given façade $\mathbf{F}_{N \times M}$ can be expressed by the following minimization problem

$$\min_{\mathbf{M}_k, \, \hat{\mathbf{P}}_k, \, \lambda_k, \, k = 1, \cdots, K} \mathcal{C}_F(\mathbf{M}_k, \, \hat{\mathbf{P}}_k, \, \lambda_k, \, k = 1, \cdots, K), \quad (7)$$

which is known as the nearest Kronecker product problem [20]. The following partition of $\mathbf{F}_{N \times M}$ is key to the solution of Eq. (7):

$$\mathbf{F}_{N \times M} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \cdots & \mathbf{F}_{1l_h} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}_{l_v 1} & \mathbf{F}_{l_v 2} & \cdots & \mathbf{F}_{l_v l_h} \end{bmatrix}, \quad (8)$$

where $\mathbf{F}_{ij}$ is a block of size $N_v \times N_h$. Define matrix

$$\mathbf{A}_{l_v l_h \times N_v N_h} = \left[vec\{\mathbf{F}_{11}\} \, vec\{\mathbf{F}_{21}\} \ldots vec\{\mathbf{F}_{l_v l_h}\} \right]^T, \quad (9)$$

which constitutes a rearrangement of $\mathbf{F}_{N \times M}$. Using the above defined quantities, the cost function of Eq. (6) can be expressed as:

$$\mathcal{C}_F(\mathbf{m}_k, \, \hat{\mathbf{p}}_k, \, \lambda_k, \, k = 1, \cdots, K) = ||\mathbf{A}_{l_v l_h \times N_v N_h} - \sum_{k=1}^{K} \lambda_k \mathbf{m}_k \hat{\mathbf{p}}_k^T ||_2^2, \quad (10)$$

where $\mathbf{m}_k, \, \hat{\mathbf{p}}_k$ are the column-wise vectorized forms of $\mathbf{M}_k, \, \hat{\mathbf{P}}_k$. By exploiting the above defined equivalent form of the cost function, the Kronecker Product SVD [20] can be used to solve the optimization problem in Eq. (7):

Theorem 1. *Let $\mathbf{A}_{l_v l_h \times N_v N_h} = \mathbf{V} \Sigma \mathbf{U}^T$ be the Singular Value Decomposition (SVD) of the rearranged counterpart of matrix $\mathbf{F}_{N \times M}$. Consider the following diagonal matrix*

$$\Sigma_K = diag\{\sigma_1 \, \sigma_2 \, \cdots \, \sigma_K\} \quad (11)$$

containing the first K singular values of matrix $\mathbf{A}_{l_v l_h \times N_v N_h}$, and let

$$\mathbf{V}_K = [\mathbf{v}_1 \, \mathbf{v}_2 \, \cdots \mathbf{v}_K], \quad \mathbf{U}_K = [\mathbf{u}_1 \, \mathbf{u}_2 \, \cdots \mathbf{u}_K] \quad (12)$$

be the K associated left and right singular vectors respectively. Then, the matrices $\mathbf{M}_k^*$, the patterns $\hat{\mathbf{P}}_k^*$, and the weighting factors λ_k^* that satisfy:

$$vec\{\mathbf{M}_k^*\} = \mathbf{v}_k, vec\{\hat{\mathbf{P}}_k^*\} = \mathbf{u}_k, \lambda_k^* = \sigma_k, \, k = 1, \cdots, K, \quad (13)$$

constitute the solution of the Eq. (7).

Using Theorem 1, we can find an optimal approximation in the desired form, i.e. it is a sum of Kronecker products, that minimizes the cost function defined in Eq. (6). Note, however, that some of the characteristics of the optimal solution are not consistent with the ingredients of the model defined in Eq. (4), which makes the direct use of Theorem 1 problematic. In particular, neither the matrices $\mathbf{M}_k^*$ nor the patterns $\hat{\mathbf{P}}_k^*$ have the desired form in general, i.e. they are not 1-0 matrices and piecewise constant surfaces, respectively.

In order to impose one of the requirements of the proposed model, in the sequel we assume that matrices $\mathbf{M}_k$ have the desired 1-0 form and are known. We thus form the cost function:

$$\hat{\mathcal{C}}_F(\hat{\mathbf{P}}_k, \, \lambda_k, \, k = 1, \cdots, K | \mathbf{M}_k), \quad (14)$$

which is the cost function of Eq. (6) but with the partition matrices known. We would like to minimize it with respect to the patterns $\mathbf{P}_k$ and the weighting factors λ_k. The solution of the new optimization problem is the subject of Lemma 1 (refer to supplemental material for proof).

Lemma 1. *Assuming that the matrices $\mathbf{M}_k, \, k = 1, 2, \cdots, K$ defined in Eqs. (1-2) are known, then the minimization of the cost function defined in Eq. (14) produces patterns $\hat{\mathbf{P}}_k$ and weighting factors λ_k that are related as follows:*

$$\lambda_k^* vec\{\hat{\mathbf{P}}_k^*\} = \frac{\mathbf{U} \Sigma \mathbf{V}^T vec\{\mathbf{M}_k\}}{||vec\{\mathbf{M}_k\}||_2^2}, \, k = 1, 2, \cdots, K. \quad (15)$$

Lemma 1 is a powerful tool that can be used for solving the modeling problem of urban building façades. However, knowledge on the partitioning $1 - 0$ matrices $\mathbf{M}_k, \, k = 1, \cdots, K$, is required. In the next section, inspired by Eq. (15), we present an algorithm to estimate them, which leads to the solution of the minimization problem in Eq. (7).

4 ALGORITHM

Our algorithm starts with the estimation of the spatial periods N_v and N_h (Section 4.1), and continues with the estimation of K along with the actual partition matrices, pattern matrices and weights (Section 4.2) and concludes with pattern refinement (Section 4.3).

4.1 Estimation of Spatial Periods

In the first stage of our algorithm, we estimate the spatial periods of the patterns. Although well known methods ([8], [9]) can be used for that purpose, we propose a K-means based algorithm to estimate the spatial periods in an efficient way.

Intuitively the spatial period indicates the distance two consecutive repeated patterns span, thus finding similar patterns and their position is critical for computing the spatial period. When patterns (like windows and doors) in a façade repeat in a specific period, the wall patches between windows repeat in the same period.

In order to compute the spatial period along horizontal direction, we can measure the repetition of vertical vectors (columns) of the input image. Similarly, we measure the repetition of horizontal vectors (rows) for computing vertical spatial periods. We propose a method that iteratively estimates the number of different repeated vectors and optimal spatial period. We use a K-means clustering to estimate the periods along both vertical and horizontal directions. Assume that we are given the desired number of clusters (denote it by L) to which we will group the rows of $\mathbf{F}$. Define the following set consisting of L clusters:

$$\mathcal{R}_l = \{\mathbf{b}_q^T : ||\mathbf{b}_q^T - \bar{\mathbf{d}}_l^T||_2^2 \leq ||\mathbf{b}_q^T - \bar{\mathbf{d}}_k^T||_2^2, \, \forall \, 1 \leq k \leq L\}, \quad (16)$$
$$l = 1, 2, \cdots, L,$$

where $\mathbf{b}_q^T$ denotes the q-th row of matrix $\mathbf{F}$, and $\bar{\mathbf{d}}_l^T$ denotes the mean of the l-th cluster of the rows computed by K-means.

Let us also define the corresponding *indicator* vectors of length N:

$$\mathbf{I}_{\mathcal{R}_l}[q] = \begin{cases} 1 & \text{if } \mathbf{b}_q^T \in \mathcal{R}_l, \\ 0 & \text{otherwise,} \end{cases} \quad \text{where } q = 1, 2, \cdots, N. \quad (17)$$

and the *element-wise* median vectors of each cluster:

$$\bar{\mathbf{r}}_l^T = median\{\mathcal{R}_l\}, \; l = 1, 2, \cdots, L. \tag{18}$$

In order to be consistent with Lemma 1, Eq. (18) should be using the mean, since it provides the optimal result assuming an ideal noise-free case. However, due to variations caused by occlusions (such as trees, traffic lights, etc.) or shadows and lighting changes, the median is proven more robust in our empirical studies.

We can now define the following matrix:

$$\mathbf{F}_{\mathcal{R}} = \sum_{l=1}^{L} \mathbf{I}_{\mathcal{R}_l} \bar{\mathbf{r}}_l^T, \tag{19}$$

which has the same size as **F**. More importantly, if the given number of clusters L were the correct one, then L should be equal to the rank of **F**. If, on the other hand, L is greater than the rank of **F**, then the rank of $\mathbf{F}_{\mathcal{R}}$ will be smaller than L. Hence, by computing the number of clusters as:

$$L = rank(\mathbf{F}_{\mathcal{R}}), \tag{20}$$

and repeating the above procedure, we expect that after some iterations, $\mathbf{F}_{\mathcal{R}}$ will be the desired approximation of **F**.

Algorithm 1. Kronecker Façade Modeling, Noise-Free Ideal Case. **Input: F**

1: Initialize L: $L \leftarrow rank(\mathbf{F}) - 1$
2: **repeat**
3:    Form clusters $\mathcal{R}_l, \; l = 1, \cdots, L$ via K-means (16)
4:    Form the indicator vectors $\mathbf{I}_{\mathcal{R}_l}$ of (17)
5:    Form the median vectors $\bar{\mathbf{r}}_l^T$ of (18)
6:    Compute the matrix $\mathbf{F}_{\mathcal{R}}$ defined in (19)
7:    Compute its rank L in (20)
8:    Assign $\mathbf{F}_{\mathcal{R}}$ to **F**
9: **until** convergence
10: Output: $\mathbf{F}_{\mathcal{R}}^{\star}, \; L^{\star}, \mathbf{I}_{\mathcal{R}_l}$.
    Note that $\bar{\mathbf{r}}_l^t, l = 1, 2, \cdots, L^{\star}$ are the rows of $\mathbf{F}_{\mathcal{R}}^{\star}$.

Unfortunately, Lemma 1 does not guarantee that the patterns are piece-wise constant. One way to enforce that constraint is to force the clustering in the columns of **F** as well. We thus consider the matrix:

$$\mathbf{G} = \frac{1}{2}(\mathbf{F}_{\mathcal{C}} + \mathbf{F}_{\mathcal{R}}) \tag{21}$$

and the new number of the clusters:

$$L = \min\{rank(\mathbf{F}_{\mathcal{R}}), \; rank(\mathbf{F}_{\mathcal{C}})\}, \tag{22}$$

where $\mathbf{F}_{\mathcal{C}}$ is the column-wise clustering result. It is obtained by following the same K-means clustering, but now in the columns:

$$\mathcal{C}_l = \{\mathbf{b}_p : ||\mathbf{b}_p - \bar{\mathbf{e}}_l||_2^2 \leq \; ||\mathbf{b}_p - \bar{\mathbf{e}}_m||_2^2, \; \forall \, 1 \leq \; m \; \leq \; L\},$$
$$l = 1, 2, \cdots, L, \tag{23}$$

where $\mathbf{b}_p$ denotes the p-th column of matrix **F**, and $\bar{\mathbf{e}}_k$ denotes the mean of the k-th cluster of the columns respectively. The corresponding *indicator* vectors of length M is defined as:

$$\mathbf{I}_{\mathcal{C}_l}[p] = \begin{cases} 1, & \text{if } \mathbf{b}_p \in \mathcal{C}_l, \\ 0, & \text{otherwise,} \end{cases} \quad \text{where } p = 1, \cdots, M. \tag{24}$$

and the *element-wise* median vectors of each cluster:

$$\bar{\mathbf{c}}_l = median\{\mathcal{C}_l\}, \; l = 1, 2, \cdots, \; L. \tag{25}$$

Then,

$$\mathbf{F}_{\mathcal{C}} = \sum_{l=1}^{L} \bar{\mathbf{c}}_l \mathbf{I}_{\mathcal{C}_l}^T. \tag{26}$$
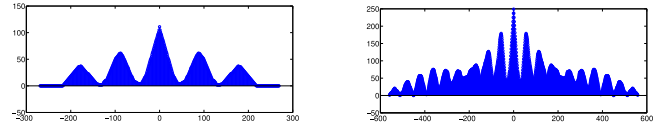
The algorithm is summarized as follows.

Fig. 1. Estimation of the spatial periods of façade shown in Fig. 2. Left: Auto-Correlation sequences used for the estimation of $N_v = 90 \; pixels$. Right: Estimation of $N_h = 56 \; pixels$.

Here, we stop the iteration when $||\mathbf{G}_i - \mathbf{G}_{i-1}||_2$ is sufficiently small, where $\mathbf{G}_i$ denotes the **G** obtained in the ith iteration, or the number of iterations reaches a pre-specified number.

In order to estimate the spatial periods, we run Algorithm 2 for an initial L_0 of the parameter L with input **F**. The initial value L_0 can be any number that is slightly smaller than the rank of **F** for convergence. Then, we can compute the following:

$$||\mathbf{I}_{\mathcal{R}_{k^{\star}}}||_0 = \max_{k=1,2,\cdots,L^{\star}} \{||\mathbf{I}_{\mathcal{R}_k}||_0\}, \tag{27}$$

$$||\mathbf{I}_{\mathcal{C}_{l^{\star}}}||_0 = \max_{l=1,2,\cdots,L^{\star}} \{||\mathbf{I}_{\mathcal{C}_l}||_0\}, \tag{28}$$

and the corresponding auto-correlation sequences:

$$r_{\mathcal{R}_{k^{\star}}} = \mathbf{I}_{\mathcal{R}_k^{\star}} * \mathbf{I}_{\mathcal{R}_k^{\star}}, \tag{29}$$

$$c_{\mathcal{C}_{k^{\star}}} = \mathbf{I}_{\mathcal{C}_l^{\star}} * \mathbf{I}_{\mathcal{C}_l^{\star}}, \tag{30}$$

where " $*$ " denotes the correlation operator. Note that the vectors involved in the computation of the proposed auto-correlation sequences are based on *indicator* vectors, that is $1 - 0$ vectors. We then calculate the distances between adjacent peaks for $r_{\mathcal{R}_{k^{\star}}}$ and $c_{\mathcal{C}_{k^{\star}}}$ sequence respectively. Finally, the spatial periods N_v and N_h are estimated as the mode of the set of distances, which represents the most robust choice.

Algorithm 2. Kronecker Façade Modeling. **Input: F**

1: Initialize L: $L \leftarrow rank(\mathbf{F}) - 1$
2: **repeat**
3:    Form clusters $\mathcal{R}_l, \mathcal{C}_l, \; l = 1, \cdots, L$ via K-means using (16) and (23)
4:    Form the indicator vectors $\mathbf{I}_{\mathcal{R}_l}, \; \mathbf{I}_{\mathcal{C}_l}$ of (17), (24)
5:    Form the median vectors $\bar{\mathbf{r}}_l^T, \bar{\mathbf{c}}_l$ of (18), (25)
6:    Form the matrices $\mathbf{F}_R, \mathbf{F}_C$ and **G** of (19), (26) and (21)
7:    Set L using (22)
8:    Assign **G** to **F**
9: **until** convergence
10: Output: $\mathbf{F}_{\mathcal{R}}^{\star}, \; L^{\star}, \mathbf{I}_{\mathcal{R}_l}, \mathbf{I}_{\mathcal{C}_l}$.
    Note that $\bar{\mathbf{r}}_l^t, l = 1, 2, \cdots, L^{\star}$ are the rows of $\mathbf{F}_{\mathcal{R}}^{\star}$.

The results for the urban building façade in Fig. 2 (top), are shown in Fig. 1. The façade can be partitioned into blocks by using the estimated spatial periods, as illustrated in Fig. 2.

4.2 Estimation of K by Unbiased Estimator of the Degrees of Freedom

Given the computed N_v and N_h, we can rearrange the façade **F** into **A** as described in Eq. (9) (see a representative image in the supplemental material).

If the façade contains repeated patterns, then the partition blocks can be clustered into groups. As we have re-arranged the partitioned blocks into vectors, each group of repeated patterns is in the form of a group of repeated vectors in **A**. Thus K represents the rank of **A** and the problem reduces to the estimation of rank of **A**. Therefore the problem can be formed as the following statistical problem of finding the correct rank of a perturbed low-rank matrix: given a noisy observation $\mathbf{A} = \mathbf{A}^{\circ} + \mathbf{E}$, the goal is to estimate an $m_1 \times m_2$ matrix $\mathbf{A}^{\circ}$, where $m_1 = l_v l_h, \; m_2 = N_v N_h$ and $rank(\mathbf{A}^{\circ}) = K$. We
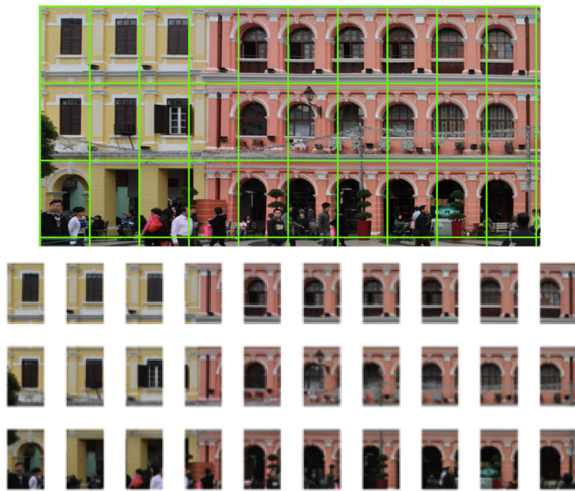
Fig. 2. Top: Partitioning the façade into blocks using the spatial periods estimated in Fig. 1, where the green lines shows the boundary of partition blocks. Bottom: The corresponding 3×10 blocks viewed independently.
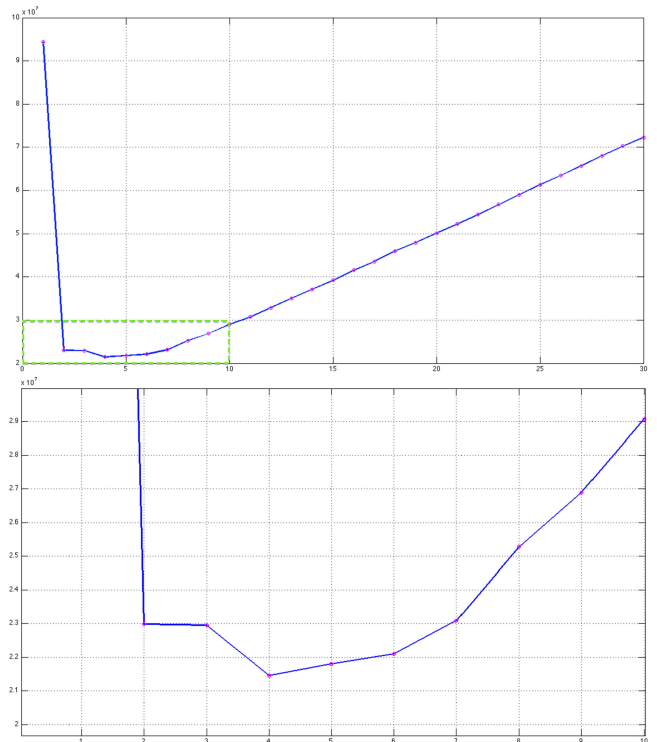

Fig. 3. Rank estimation result for the image shown in Fig. 2. Top: Plot of the function in Eq. (36), where the global minimum comes up at index 4. Bottom: The enlarged figure area within the green box in (a), which shows the global minimum in a clearer way.

assume that the noise matrix $\mathbf{E}$ follows a matrix normal distribution where $E_{ij} \overset{i.i.d.}{\sim} N(0, \tau^2)$ and $m_1 \leq m_2$ so that the full rank is m_1.

A K-means clustering based iterative algorithm was used in [1] to estimate the rank of $\mathbf{A}$. However, the K-means based approach is computationally expensive and unstable in cases where there is occlusion caused by illumination or shadows. In [21], an approach was proposed to address this problem via Degrees of Freedom estimators in an efficient and reliable way. Based on this idea, we will describe a statistical technique that estimates the rank K of $\mathbf{A}$.

In [22] and [23], a rigorous definition of degrees of freedom in the framework of Stein's Unbiased Risk Estimate (SURE) was provided. For the classical linear regression, degrees of freedom is often associated with the number of variables in the model. However the parallel interpretation is unclear in the context of low rank matrix estimation problems where the estimators are highly nonlinear in nature. The number of free parameters in specifying a low rank matrix is often used as the degrees of freedom in this case. It was shown in [21] that the number of free parameters incorrectly measures the complexity of the rank constrained estimator.

Let $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^\top$ be the SVD of $\mathbf{A}$. The estimator of $\mathbf{A}$ with rank K, denoted as $\hat{\mathbf{A}}_K$, is defined as:

$$\hat{\mathbf{A}}_K = \sum_{k=1}^{K} \sigma_k \mathrm{u}_k \mathrm{v}_k^\top, \qquad (31)$$

where u_k and v_k are the kth column of $\mathbf{U}$ and $\mathbf{V}$ respectively.

The formal definition of the optimization function for estimating K is formulated as:

$$\begin{aligned}
\ell(\hat{\mathbf{A}}_K) &= \|\hat{\mathbf{A}}_K - \mathbf{A}^\circ\|_2^2 = \|\hat{\mathbf{A}}_K - (\mathbf{A} - \mathbf{E})\|_2^2 \\
&= \|\hat{\mathbf{A}}_K - \mathbf{A}\|_2^2 + 2 \times \langle \hat{\mathbf{A}}_K - \mathbf{A}, \mathbf{E} \rangle + \|\mathbf{E}\|_2^2 \\
&= \|\hat{\mathbf{A}}_K - \mathbf{A}\|_2^2 + 2 \times \langle \hat{\mathbf{A}}_K, \mathbf{E} \rangle \\
&\quad + (\textit{terms not depending on } \hat{\mathbf{A}}_K), \qquad (32)
\end{aligned}$$

where $K \in \{1, \cdots, m_1\}$ is a tuning parameter and $\langle \cdot, \cdot \rangle$ stands for the inner product. The first term measures the goodness of fit of $\hat{\mathbf{A}}_K$ to the observation $\mathbf{A}$. The second term can be interpreted as the cost of the estimating procedure and can be estimated by using degrees of freedom as shown in [21].

$$df(\hat{\mathbf{A}}_K) = \frac{1}{\tau^2} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} cov(\hat{\mathbf{A}}_{Kij}, \mathbf{E}_{ij}). \qquad (33)$$

We refer the interested readers to [22] and [23] for further discussions about the general theory regarding degrees of freedom. Once the degrees of freedom are defined, the rank estimator can be constructed by using the following C_p type statistic:

$$C_p(\hat{\mathbf{A}}_K) = \|\hat{\mathbf{A}}_K - \mathbf{A}\|_2^2 + 2\tau^2 df(\hat{\mathbf{A}}_K), \qquad (34)$$

where τ^2 is defined as $var(\hat{\mathbf{A}}_{\tilde{K}} - \mathbf{A})$, and $\hat{\mathbf{A}}_{\tilde{K}}$ corresponds to a low-rank estimate which explains λ proportion of the total variance, where $0 \leq \lambda \leq 1$. Here, we fix $\lambda = 0.935$ through all the experiments.

Usually, the degrees of freedom defined in Eq. (33) are not directly computable. The unbiased estimator proposed in [24] lacks analytical expressions and requires numerical methods such as data perturbation and resampling techniques which are computationally prohibitive in large scale problems. For our specific rank regularized estimation problem, we employ the following unbiased estimator of degrees of freedom (see details in Theorem 1 of [21]):

$$\hat{df}(\hat{\mathbf{A}}_K) = (m_1 + m_2 - K)K + 2\sum_{k=1}^{K}\sum_{l=K+1}^{m_1} \frac{\sigma_l^2}{\sigma_k^2 - \sigma_l^2}. \qquad (35)$$

Using Eq. (35), we arrive at the following estimator of the C_p statistic for each candidate rank K:

$$\hat{C}_p(\hat{\mathbf{A}}_K) = \|\hat{\mathbf{A}}_K - \mathbf{A}\|_2^2 + 2\tau^2 \hat{df}(\hat{\mathbf{A}}_K). \qquad (36)$$

The estimated rank $K^\star$ is then defined as follows:

$$K^\star = \arg\min_{1 \leq K \leq m_1} \hat{C}_p(\hat{\mathbf{A}}_K). \qquad (37)$$

The quantity $\hat{C}_p(\hat{\mathbf{A}}_K)$ as a function of K for the façade in Fig. 2 (top) is shown in Fig. 3. As we described in Section 3, each row of this low-rank matrix indicates one partition block of the input façade. By reshaping each row to a matrix in the original partition block size and re-arranging the blocks in their original order, we obtain a "clean" façade image where the noise and occlusions have
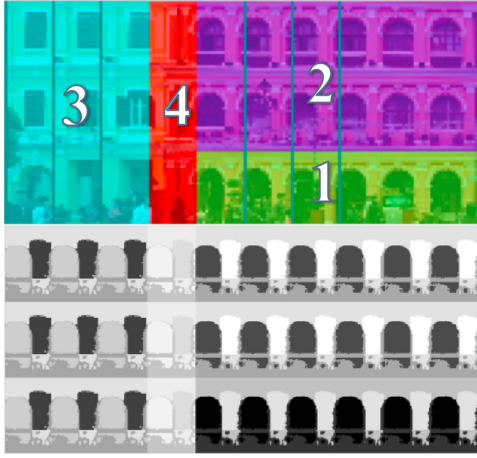
Fig. 4. Top: Each color represents one group. Bottom: The reconstructed façade image.

Fig. 6. Left: The original façade. Right: The recovered "clean" façade structure hidden beneath the noise.

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

Fig. 7. (a) Input image. (b) Partition grid showing the estimated periods. (c) Low-rank component generated by the method in Section 4.2. (d) Estimated 1-0 repeated patterns computed by Algorithm 3 and the refinement step. Each color represents one group. (e) Ground truth.

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

Fig. 8. (a) A failure case in [3]. (b) Our method detects all patterns. (c) In [3], the bottom two patterns are not detected. (d) Ground truth.

been largely removed, as shown in Fig. 4 (bottom). Simultaneously these re-shaped patterns can be easily clustered into K groups as illustrated in Fig. 4.

The algorithm described in this section can be summarized as follows.

Algorithm 3. Kronecker Façade Modeling: estimating rank K.

Input: $\mathbf{A}$ of size $m_1 \times m_2$. $\mathbf{A}$ is the rearranged matrix as defined in Eq. (9).

1:   $C_{min} \leftarrow \infty$, $K^* \leftarrow m_1$
2:   Compute the SVD of $\mathbf{A}$, $\mathbf{A} = \mathbf{U\Sigma V}^\top$
3:   **for** $K = 1$ to m_1 **do**
4:      $\hat{\mathbf{A}}_K \leftarrow \sum_{k=1}^{K} \sigma_k \mathbf{u}_k \mathbf{v}_k^\top$
5:      $f_1(K) \leftarrow \|\hat{\mathbf{A}}_K - \mathbf{A}\|_2^2$
6:      $\hat{df}(\mathbf{A}_K) \leftarrow (m_1 + m_2 - K)K + 2\sum_{k=1}^{K}\sum_{l=K+1}^{m_1}\frac{\sigma_l^2}{\sigma_k^2 - \sigma_l^2}$
7:      $\tau^2 \leftarrow var(\hat{\mathbf{A}}_K - \mathbf{A})$
8:      $C_p(K) \leftarrow f_1(K) + 2\tau^2 df(K)$
9:      **if** $C_p(K) < C_{min}$ **then**
10:        $C_{min} \leftarrow C_p(K)$, $K^* \leftarrow K$
11:      **end if**
12:   **end for**
13:   **Output**: $\hat{\mathbf{A}}_{K^*}$, K^*.

We apply K-means to the rows of the matrix $\hat{\mathbf{A}}_{K^*}$ with K^* clusters. The result can be represented in terms of length-$l_h \times l_v$ indicator vectors (similar to Eq. (17)) which are denoted as $\mathbf{m}_k^*$, $k = 1, \cdots, K^*$, i.e. $\mathbf{m}_k^*(i)$ has value 1 if and only if the row i belongs to cluster k. The mean vector for cluster k from the K-means is denoted by $\mathbf{p}_k^*$ for $k = 1, \cdots, K^*$. To get the detected repeated patterns along with the cluster information, we estimate matrices $\mathbf{M}_k^*$ and $\mathbf{P}_k^*$, $k = 1, 2, \cdots, K^*$ in Eq. (3) by reshaping each indicator vector $\mathbf{m}_k^*$ and $\mathbf{p}_k^*$ into a rectangular array of size $l_v \times l_h$ and $N_v \times N_h$, respectively.
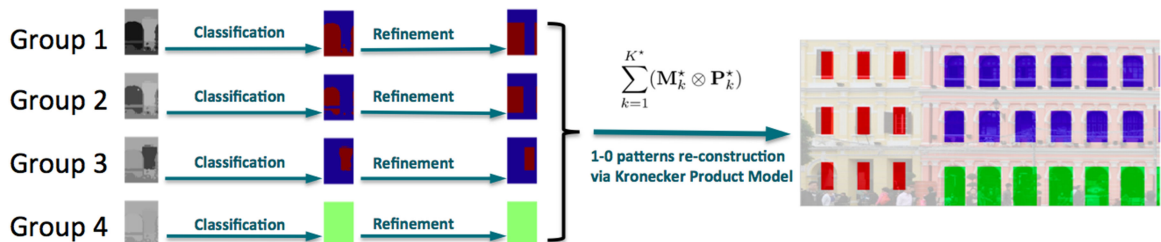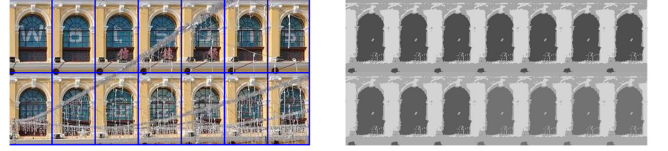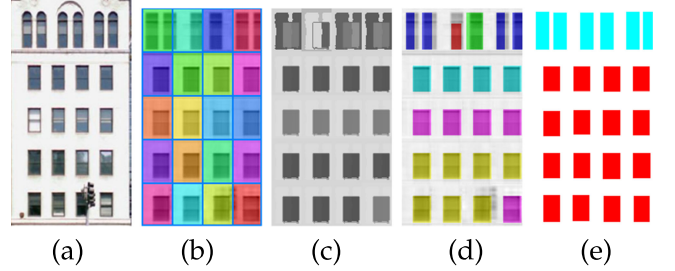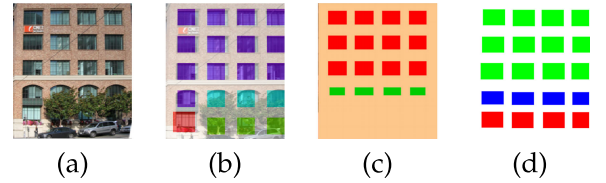
4.3 Pattern Refinement

The low-rank method described above enables us to remove occlusions, small illumination variations and photometric distortions as seen in Fig. 6. As a result, we have very accurate detection of repeated patterns. We next classify each detected pattern $\mathbf{P}_k^*$ into wall vs. non-wall pixels, using the pre-trained random forest classifier in [3]. That classifier was trained on 140 facade images that do not overlap with our testing images. Each classified pattern is further refined based on the rank-one algorithm of [3], which minimizes the l_0-norm of the approximation error. The classification and refinement steps are illustrated in Fig. 5 with an example of detected 1-0 patterns shown in Fig. 7d.

5 EXPERIMENTS AND DISCUSSION

The experiments were executed in Matlab, and run on a computer with an 1.8 GHz Intel Core i7 CPU and a 4GB memory.

We have created a ground truth dataset for a set of 104 façade images from [3], [4]. See Fig. 8(d) for an example. To this end we have manually marked the groups of repeated patterns for each image. The number of groups for each image is the parameter K

Fig. 5. Illustration of classification and refinement steps.

TABLE 1
Performance Comparison between the Proposed Method and [1] Regarding the Sample Mean and Variance of $\hat{K} - K$

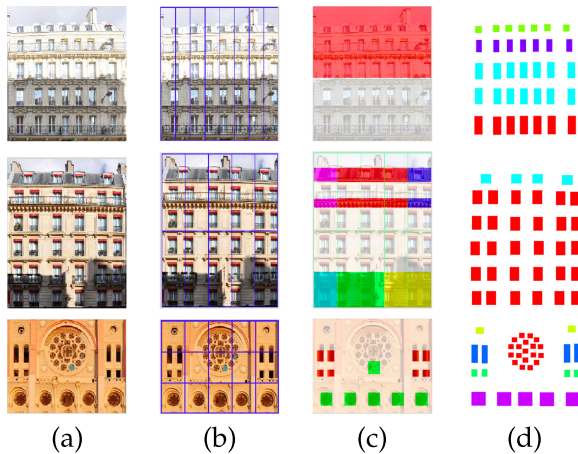| Method	Sample mean	Sample variance
Method of [1]	1.95	6.09
Our method	0.81	2.01

Fig. 9. The first two rows fail at period detection due to strong illumination variation. In the third row, the input façade contains a set of Penrose tiling style windows in the middle top. (a) Original images. (b) Partition grid. (c) Detected patterns. (d) Ground truth.
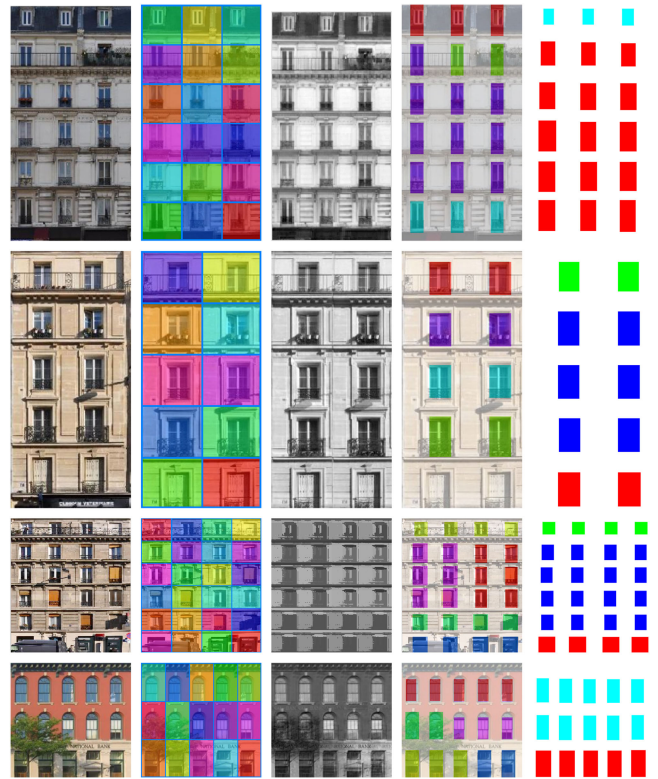

Fig. 10. From left to right: same order as in Fig. 7. The results demonstrate that the proposed method is robust to occlusions and different architectural styles.
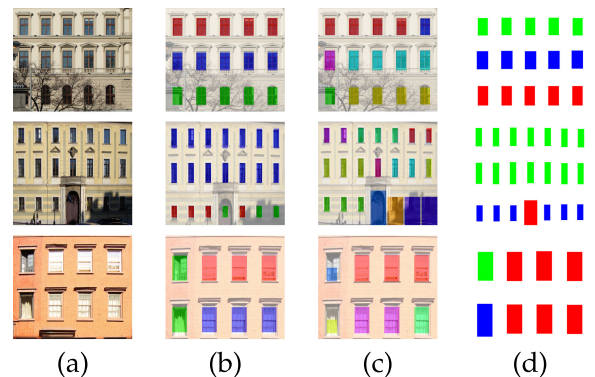

Fig. 11. Examples that demonstrate the improvement of DoF based estimation method over [1]. (a) Input image. (b) Our result: estimated number of groups ($\hat{K}$) for all images is 3. (c) [1]: Estimated number of groups ($\hat{K}$) is 6 for all images. (d) Ground truth. $K = 3$ for all images.

being estimated by Algorithm 3. We tested our method on all 104 images. The detection of repeated patterns in one image is considered a failure when the estimated K is smaller than the true K, because in that case it is impossible to recover all the repeated patterns. On the other hand, when the estimated K is the same as or larger than the true K we could recover the repeated patterns. But the larger the estimated K, the more fragmented the detection will be. That is why in addition of counting the number of failures, we also report the estimation error $\hat{K} - K$, where $\hat{K}$ represents the estimated K. The smaller this difference the better.

Out of the 104 images, our method has 4 failures, while the method in [1] produces 6 failures in a subset consisting of 95 images. The sample means and variances of $\hat{K} - K$ for our method and [1] are presented in Table 1. It is clear that by using our low-rank approach, the resulting estimate is more accurate (in terms of sample mean) and more robust (in terms of sample variance). To show the improvement over [1], we present some representative images in Figs. 10 and 11.

We evaluated the accuracy of the detected repeated pattern pixels versus the wall pixels, using an additional metric. For this we first created a binary image, in which 1 represents a detected pattern pixel, and 0 the wall pixels. We overlaid this binary image with the corresponding ground truth binary image pixel by pixel and had exact matches for 93 percent of the pixels on average over the 104 images.

Fig. 8 presents an image where the proposed method recovered all the patterns while the method in [3] did not. A more extensive collection of results can be found in supplemental material.

We found that in the experiments most of the common building façades can be captured by the Kronecker product structure. One limitation is that our method fails when a façade contains repeated structures that do not follow the Kronecker product model, such as the Penrose tiling style in the third row of Fig. 9. Another limitation is the inability to handle large photometric variations, since they are causing ambiguity in the block partition (refer to the first two rows of Fig. 9 for examples).

In conclusion, this paper describes a novel method for detecting repeated patterns that follow a Kronecker product formulation. The Kronecker product model can be applied to a variety of applications, such as image completion and 3D reconstruction. In addition, the proposed technique can be easily adapted to detect nested Kronecker product structures, in which some of the patterns detected in the first level of decomposition follows another (possibly different) Kronecker product structure, in a hierarchical way.

ACKNOWLEDGMENTS

REFERENCES

[1]   J. Liu, E. Psarakis, and I. Stamos, "Automatic Kronecker product model based detection of repeated patterns in 2D urban images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 401–408.

[2]   Z. Zhang, X. Liang, A. Ganesh, and Y. Ma, "TILT: Transform invariant low-rank textures," in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 314–328.

[3]   C. Yang, T. Han, L. Quan, and C.-L. Tai, "Parsing façade with rank-one approximation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1720–1727.

[4]   O. Teboul, I. Kokkinos, L. Simon, P. Koutsourakis, and N. Paragios, "Shape grammar parsing via reinforcement learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2273–2280.

[5]   R. Gadde, R. Marlet, and N. Paragios, "Learning grammars for architecture-specific facade parsing," *Int. J. Comput. Vis.*, vol. 117, no. 3, pp. 290–316, 2016.

[6]   A. Cohen, M. R. Oswald, Y. Liu, and M. Pollefeys, "Symmetry-aware facade parsing with occlusions," in *Proc. Int. Conf. 3D Vis.*, Oct. 2017, pp. 393–401.

[7]   P. Zhao, T. Fang, J. Xiao, H. Zhang, Q. Zhao, and L. Quan, "Rectilinear parsing of architecture in urban environment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 342–349.

[8]   S. Friedman and I. Stamos, "Online detection of repeated structures in point clouds of urban scenes for compression and registration," *Int. J. Comput. Vis.*, vol. 102, pp. 112–128, 2013.

[9]   C.-H. Shen, S.-S. Huang, H. Fu, and S.-M. Hu, "Adaptive partitioning of urban facades," *ACM Trans. Graph.*, vol. 30, no. 6, 2011, Art. no. 184.

[10]   G. Schindler, P. Krishnamurthy, R. Lublinerman, Y. Liu, and F. Dellaert, "Detecting and matching repeated patterns for automatic geo-tagging in urban environments," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–7.

[11]   Y. Liu, R. T. Collins, and Y. Tsin, "A computational model for periodic pattern perception based on frieze and wallpaper groups," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 3, pp. 354–371, Mar. 2004.

[12]   C. Wu, J.-M. Frahm, and M. Pollefeys, "Detecting large repetitive structures with salient boundaries," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 142–155.

[13]   P. Muller, G. Zeng, P. Wonka, and L. V. Gool, "Image-based procedural modeling of facades," in *Proc. ACM Siggraph*, 2007, Art. no. 85.

[14]   O. Barinova, V. Lempitsky, E. Tretiak, and P. Kohli, "Geometric image parsing in man-made environments," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 57–70.

[15]   M. Kozinski, R. Gadde, S. Zagoruyko, G. Obozinski, and R. Marlet, "A MRF shape prior for facade parsing with occlusions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2820–2828.

[16]   A. Cohen, A. G. Schwing, and M. Pollefeys, "Efficient structured parsing of facades using dynamic programming," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3206–3213.

[17]   S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Problems*, vol. 27, no. 2, 2011.

[18]   E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, Jun. 2011.

[19]   J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[20]   C. F. V. Loan, "The ubiquitous Kronecker product," *J. Comput. Appl. Math.*, vol. 123, pp. 85–100, 2000.

[21]   M. Yuan, "Degrees of freedom in low rank matrix estimation," *Sci. China Math.*, vol. 59, no. 12, pp. 2485–2502, 2016.

[22]   J. Ye, "On measuring and correcting the effects of data mining and model selection," *J. Amer. Statistical Assoc.*, vol. 93, no. 441, pp. 120–131, 1998.

[23]   B. Efron, "The estimation of prediction error: Covariance penalties and cross-validation," *J. Amer. Statistical Assoc.*, vol. 99, no. 467, pp. 619–642, 2004.

[24]   C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, 1981.