INPAINTING OF SPARSE OCCLUSION IN FACE RECOGNITION

Rui Min and Jean-Luc Dugelay

EURECOM, Department of Multimedia Communications
2229 Route Crêtes, 06560, Sophia Antipolis, France
{min, jld}@eurecom.fr

ABSTRACT

Facial occlusion is a critical issue in many face recognition applications. Existing approaches of face recognition under occlusion conditions mainly focus on the conventional facial accessories (such as sunglasses and scarf) and thus presume that the occluded region is dense and contiguous. Yet due to the wide variety of natural sources which can occlude a human face in uncontrolled environments, methods based on the dense assumption are not robust to thin and randomly distributed occlusions. This paper presents the solution to a newly identified facial occlusion problem – sparse occlusion in the context of face biometrics in video surveillance. We show that the occluded pixels can be detected in the low-rank structure of a canonical face set under the Robust-PCA framework; and the occluded part can be inpainted solely based on the non-occluded part and a Fields-of-Experts prior via spatial inference. Experiments demonstrate that the proposed approach significantly improve various face recognition algorithms in presence of complex sparse occlusions.

***Index Terms* —** Sparse Occlusion, Inpainting, Face Recognition, Robust-PCA, Fields-of-Experts

1. INTRODUCTION

Face recognition, the least intrusive biometric technique from the acquisition point of view, has been applied to a wide range of commercial and law enforcement applications. With the emphasis on real world scenarios (e.g. face recognition in video surveillance), a number of challenges including pose /illumination changes, image degradation as well as partial occlusion is required to be deliberately handled. Facial occlusion, as one of those major challenges, has been extensively studied in the literature [1-5]. However, previous works mainly focus on facial occlusions which are dense and contiguous (e.g. sunglasses, scarf, beards, hat and hand on face), whereas neglecting the other types of facial occlusions.

De facto, there exists a large variety of facial occlusions in uncontrolled environments. In addition to the well-studied facial occlusions in the literature, in this paper, we point out that occlusions caused by facial painting, face dirt, and face behind fence (where the occluded part is often not dense) can also greatly hinder many popular face recognition systems. Inherent from the sparsity/density dichotomy in graph theory [6], we categorize facial occlusions into 2 classes: the *dense occlusion* and the *sparse occlusion* (see Fig. 1). Unlike the traditional studies, the aim of this paper is to address the newly identified sparse occlusion problem in face recognition.

(a) Dense Occlusions **(b) Sparse Occlusions**

Fig. 1: Examples of various kinds of facial occlusions: (a) densely occluded faces, (b) sparsely occluded faces.

Recently, researchers have revealed that imposing prior knowledge of occlusion can significantly improve results of face recognition under occlusion conditions [3-5]. Hence, explicit occlusion analysis is an essential step in occluded face recognition. However, since the previous focus is primarily on dense occlusions, the dense assumption is made intentionally or undeliberately in the detection of occluded regions. In [3] and [4], faces are divided into pre-defined local patches for occlusion detection, where the occluded part is supposed to be larger/equal to the patch size and condensed; the occlusion-free patches are then used in local feature based face recognition. Zhou et al. [5] used a Markov Random Fields (MRF) model to incorporate spatial continuity constraints in the modeling of contiguous occlusions (in order to exclude the information from the occluded part) which improves sparse representation based face recognition [2]. Apparently, such methods are inappropriate when dealing with sparsely occluded faces.

Towards the problem of sparse occlusion, we detect occluded pixels with emphasis on the sparsity by explicit occlusion modeling. In addition, inspired by the work of image inpainting [7], instead of simply excluding information from the occluded part (as suggested by [3-5]), we propose to further recover the occluded part from the non-occluded part via spatial inference. The detection and inpainting of sparse occlusion are then achieved based on the methods of Robust Principal Component Analysis (Robust-PCA) [8] and Fields-of-Experts (FoE) [9] respectively.

Contributions: In this paper, we identified a new type of facial occlusion (sparse occlusion) which is an important issue

This work is partially funded by the French national project FR OSEO BIORAFALE.

however overlooked in state-of-the-art of occluded face recognition. The main contribution of this paper is the idea to detect and then inpaint such sparse occlusions via spatial inference so as to improve face recognition, which is not studied in any prior work according to our best knowledge. Based on this idea, we built an automatic system to detect and inpaint sparsely occluded faces and demonstrated significant improvements of various face recognition algorithms (we tested PCA [10], SIFT [11] and LBP [12] based face recognition respectively) under different sparse occlusions.

The rest of this paper is structured as follows. The proposed algorithm is described in details in section 2. Section 3 presents the experimental results and analysis. Finally, we draw the conclusion in Section 4.

2. APPROACH

In this paper, we propose the solution to handle sparse occlusions in face recognition. We assume that occlusions are large deviations from a low-dimensional face space. A probe face (well-aligned) is thus represented as the lowest rank reconstruction from a canonical face set (where the faces are well-aligned and non-occluded) added with a sparse error vector under the context of Robust-PCA [8]. Based upon the computed sparse error vector, we can discriminate the large error entries from the small facial appearance agitations so as to detect the occluded pixels on a face image. For the inpainting of occluded pixels, we adopt a generic FoE prior [9] to infer the missing part, which demonstrates significant improvements in both visual quality and recognition results for faces with sparse occlusions. Finally, the recovered face is utilized as input to the face recognition system. A high-level work flow diagram of the proposed method is shown in Fig. 2.

2.1. Definition of Dense/Sparse Occlusion

Let us consider an image as an undirected adjacency graph $G = (V, E)$ in which the pixels represent the vertices and the pixel-neighborhoods represent the edges. Given the occluded part of a face image as an induced subgraph $G' = (V', E')$ of the entire graph G, the occlusion is called dense when the number of edges $|E'|$ in G' is close to the maximal number of edges in an adjacency graph with $|V'|$ vertices and vice versa. By this definition, facial occlusions like sunglasses, scarf, and hat (Fig. 1a) are regarded as dense occlusions; whereas examples like facial painting, face dirt and face behind fence (Fig. 1b) belong to the sparse occlusion category.

2.2. Sparse Occlusion Detection

We are given a set of $\mathcal{K}$ well-aligned and non-occluded faces $\mathcal{C} = \{c_1, c_2, \ldots, c_{\mathcal{K}}\}$ represented by pixel-vectors $\{c_i\} \in \mathbb{R}^m$, where m is the feature dimension. Given a probe face $y \in \mathbb{R}^m$, the occlusion modeling method suggested by [2][5] is to find a sparse error vector $e \in \mathbb{R}^m$ by solving:

$$\arg\min_{(x,e)} \|x\|_1 + \|e\|_1 \quad \text{subj. to} \quad \mathcal{C}x + e = y \quad (1)$$

where x is a sparse[1] coefficients vector. The prerequisite to correctly find e relies on a sufficient number of well-aligned training samples $\{a_1, \ldots, a_k\} \in \mathcal{C}$ from the same subject of y, so that y can be linearly approximated in a low-dimensional space by:

$$y \cong \sum_{i=1}^{k} a_i \quad (2)$$

However, in many practical face recognition scenarios, the training samples of each subject in $\mathcal{C}$ are often insufficient (the "curse of the dimensionality" [13] problem, in the extreme case only one template face per subject is available) to correctly resolve equation (1). Therefore we loosen the prerequisite by only assuming faces in $\mathcal{C}$ are well-aligned and non-occluded. In this sense, we model occlusions as large deviations from a low-dimensional face space derived by the canonical set $\mathcal{C}$ (such a set can be different from the gallery set, preferably a smaller set to accelerate computation).

To do so, we first integrate the probe face y with set $\mathcal{C}$ to build an observation matrix $\mathcal{C}^+ = \{y, c_1, \ldots, c_{\mathcal{K}}\}$, where $\mathcal{C}^+ \in \mathbb{R}^{m\times(\mathcal{K}+1)}$ was generated by a low-rank matrix $\mathcal{A} \in \mathbb{R}^{m\times(\mathcal{K}+1)}$ with large corruption (occlusion) on $\mathcal{C}_1^+$ and minor errors (small facial appearance agitations) on $\mathcal{C}_{2\sim\mathcal{K}+1}^+$. The corruptions are represented by an additive matrix $\mathcal{E} \in \mathbb{R}^{m\times(\mathcal{K}+1)}$, where $\mathcal{C}^+ = \mathcal{A} + \mathcal{E}$. Our goal is thus to recover the correct corruptions $\mathcal{E}'$, more specifically $\mathcal{E}_1'$. Because facial occlusion affects only a portion of the entire face, thanks to the "blessing of the dimensionality" [13], the sparse error $\mathcal{E}$ can be efficiently and exactly separated from the low-rank structure of $\mathcal{C}^+$. This problem formulation can be effectively resolved by the Robust-PCA framework [8] via the following optimization relaxation:

$$\arg\min_{(\mathcal{A},\mathcal{E})} \|\mathcal{A}\|_* + \gamma\|\mathcal{E}\|_1 \quad \text{subj. to} \quad \mathcal{A} + \mathcal{E} = \mathcal{C}^+ \quad (3)$$

where $\|\cdot\|_*$ is the nuclear norm which pursues the lowest rank $\mathcal{A}$ that aims to regenerate the observations; and $\|\cdot\|_1$ is the ℓ1-norm which pursues the sparsity of errors. We adopt the inexact Augmented Lagrange Multiplier [14] (ALM) method to solve equation (3) due to its reported accuracy and efficiency.

Once the sparse error vector $\mathcal{E}_1'$ is computed, we exploit it to discriminate the large error entries (regarded as the occluded pixels) from the small facial appearance agitations by giving a pre-defined threshold:

$$M(i) = \begin{cases} 1, & |\mathcal{E}_1'(i)| > \tau, \forall \mathrm{i} \in [1, m] \\ 0, & otherwise \end{cases} \quad (4)$$

τ is selected empirically to minimize the detection error. M is the indicator of occlusion (i.e. $M(i) = 1$ if pixel i is occluded), where M will be served as the mask which supervises the

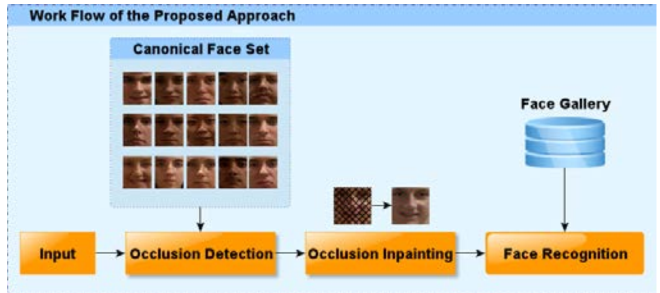

Fig. 2: A high-level workflow of the proposed method

[1] Please notice that here the term 'sparse' is different from the term we used in the dense/sparse occlusion dichotomy. In a vector/matrix, the term 'sparse' indicates the small number of non-zero entries, but not refer to the sparse definition in graph theory.

sparse occlusion inpainting approach introduced in the next section.

2.3. Sparse Occlusion Inpainting

In this paper, we apply an image inpainting method based on the Fields-of-Experts (FoE) model [9] to infer the sparsely occluded pixels. The FoE model learns a generic image prior $\mathcal{P}_{FoE}$ (a high-order Markov Random Field model) from a large number of nature image patches. $\mathcal{P}_{FoE}$ models the local image structures over extended image neighborhoods (i.e. the local spatial properties), and therefore can be used to predict the missing part from existing observations via probabilistic inference.

In a common image inpainting setting, ground truth mask of the region to be inpainted is known. In contrast, we are facing a more challenging scenario where the part to be inpainted (occluded pixels) is unknown. Hence, we supply the mask of pixels which should be inpainted by our automatic occlusion detection ($M(i), i \in [1, m]$, given in Section 2.2). Let M be the mask used in the inpainting process, given the learned image prior $\mathcal{P}_{FoE}$, the image inpainting method based on the FoE prior described in [9] can be casted as the following gradient ascent-based process:

$$x^{(t+1)} = x^{(t)} + \eta * M * \left[\nabla_{x^{(t)}} \log \mathcal{P}_{FoE}(x^{(t)})\right] \qquad (5)$$

where t is the iteration index and η is the update rate; the mask M sets the gradient to zero for all pixels outside the masked region.

Fig. 3 shows the inpainting results of various sparse occlusions of the same face. In the figure, it is clear that all four faces have large distortions where their PSNR are all below 20 dB. If the ground truth masks are given, using FoE prior based image inpainting, the PSNR of inpainted faces improve significantly (up to 39.12 dB). When using the masks returned by our automatic occlusion detection, the recovered images also achieve good visual quality improvements (where their PSNR are all above 26 dB); although some occluded pixels are not detected and thus not inpainted in the eyes and eyebrows region due to the similar appearances.

When the input face is inpainted by the proposed approach, it is then fed into the system for face recognition. Our experiments demonstrate that the proposed method cannot only improves the visual quality of sparsely occluded faces but also significantly improves the results of face recognition systems.

3. EXPERIMENTS

To assess the performance of our proposed approach, we performed a series of experiments on AR face database [15], with different types of artificially generated sparse occlusions. The dataset and detailed configurations of our experiments are introduced in Section 3.1. In Section 3.2, we will illustrate that the recognition results of face recognition algorithms accompanied with our proposed occlusion detection and inpainting approach which can significantly surpass the results from the standard algorithms (Eigenface [10], SIFT [11] and LBP [12]) in presence of sparse occlusions.

3.1. Dataset and Configurations

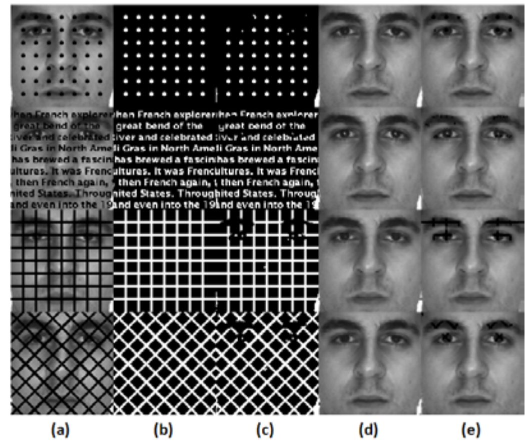

Fig. 3: Illustration of our sparse occlusion inpainting: (a) faces with different sparse occlusions (stain, text, orthogonal grid, and diagonal grid), PSNR={19.12 dB, 13.92 dB, 13.25 dB, 12.81 dB }; (b) ground truth masks of the sparse occlusions; (c) results of our sparse occlusion detection (τ = 0.004); (d) faces after inpainting using the masks in (b), PSNR={39.12 dB, 34.05 dB, 33.26 dB, 32.51 dB}; (e) faces after inpainting using the masks in (c), PSNR={30.43 dB, 28.30 dB, 26.05 dB, 26.50 dB}.

The AR face database contains more than 4000 face images of 126 subjects (70 men and 56 women) with different facial expressions, illumination conditions, and occlusions. For each subject, 26 pictures were taken in two separate sessions (two weeks interval between the two sessions). The original image resolution is 768x576 pixels. Using the eye coordinates, we cropped, normalized and down-sampled the face region into 128x128 pixels. In our experiments, 300 non-occluded faces (with facial expression and illumination variations) are randomly selected to form the canonical face set $\mathcal{C}$. For face recognition, 100 subjects (half of male and half of female) are selected. For each subject, we chose 14 images with different illumination conditions and facial expressions: 7 images from session 1 as the template faces and 7 images from session 2 as the probe faces. The probe faces are imposed by 4 kinds of artificially generated sparse occlusions (stain, text, orthogonal grid and diagonal grid as shown in Fig. 3) to simulate the real-world scenarios.

We tested 3 different face recognition algorithms on the proposed dataset, namely PCA, SIFT and LBP based face recognition, with and without the proposed occlusion detection and inpainting, respectively. For PCA based method, template faces (occlusion-free) are used to train the Eigenspace for both template and probe faces' representation. For SIFT and LBP based method, features are extracted from both template and probe faces for the Nearest-Neighbor (NN) based classification. The SIFT feature extraction is adopted from [11], and the LBP operator $\mathrm{LBP}_{8,2}^{\mathrm{u2}}$ is used in our experiment. Other settings of the experiments are listed here: $\gamma = 1/\sqrt[2]{m}$, $\tau = 0.004$; the inpainting process are consists of 2 steps: a rough step with $t = 500$ and $\eta = 10$, and a refined step to "clean up" the image with $t = 250$ and $\eta = 0.01$ as suggested by [9].

3.2. Results

Fig. 4 shows the recognition rates of PCA, SIFT and LBP based algorithms on the clean face set and the faces with

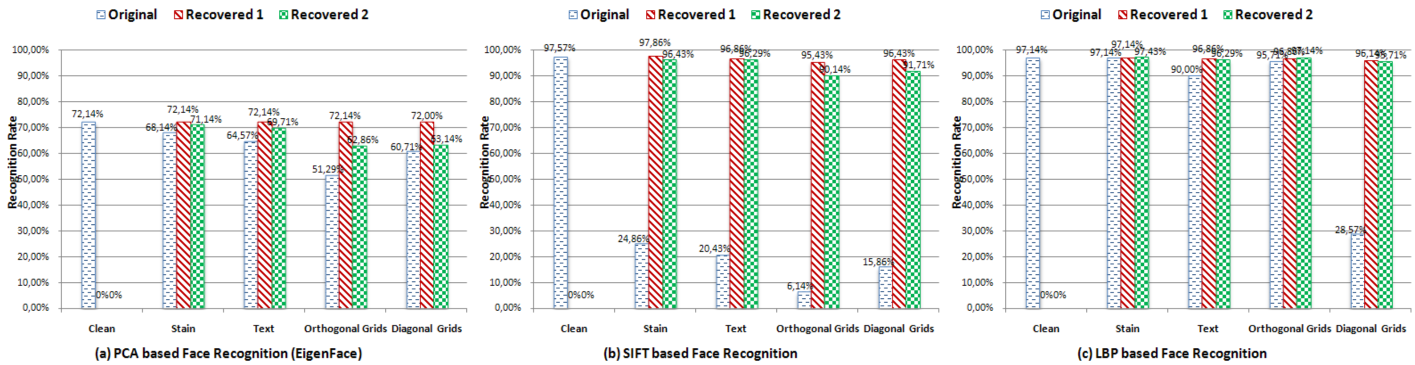

Fig. 4: Recognition rates of (a) PCA, (b) SIFT and (c) LBP based face recognition on different face sets -- the clean face set, as well as the face sets with different sparse occlusions(stain, text, orthogonal grid, and diagonal grid). The results marked as "original" are from the standard face recognition algorithms; the "Recovered 1" shows the results of inpainted faces based on FoE method using the ground truth masks; the "Recovered 2" shows the results from inpainted faces using the proposed sparse occlusion detection and inpainting algorithms

different types of sparse occlusions, respectively. It is clear that without explicit treatment, sparse occlusions can greatly deteriorate the results of those face recognition algorithms (the results marked as "Original"). In the figure, SIFT based method achieves very accurate recognition rate (97.57%) for non-occluded faces, however it is very sensitive to the sparse occlusion distortions (less than 25% in all cases, since the descriptor summarizes the edge-like features which correspond to the sparse occlusions located on the probe faces). LBP based method are somewhat robust to certain types of sparse occlusions (stain, text and orthogonal grid, because those occluded parts are located at the boarder of 8x8 LBP blocks); however when the occluded part is located inside the LBP blocks (e.g. the diagonal grid case) its recognition rate decreases drastically (down to 28.57%). Those results illustrate that even if local-feature based methods are known to be somewhat robust to conventional partial occlusions (such as scarf and sunglasses); they are still fragile to sparse occlusions.

In the figure, the inpainted faces using FoE prior based on the ground truth masks ("Recovered 1") can achieve recognition rates as good as the non-occluded faces (with the deviations less than 2%). In addition, using the proposed sparse occlusion detection and inpainting ("Recovered 2") can significantly promote the recognition rates in comparison to the results from the occluded faces. For LBP based method, the proposed approach can even achieve recognition results very close to the non-occluded faces (97.14%), because the small distortions (due to the detection errors) in few LBP blocks do not impair the discriminative power of the overall representation. In conclusion, explicit occlusion detection and inpainting can greatly improve the results of those face recognition algorithms when dealing with sparse occlusions.

4. CONCLUSIONS

This paper presents the first solution to a newly identified facial occlusion problem: sparse occlusion in the context of face biometrics in video surveillance. The proposed system automatically detects sparse occlusion on faces (using R-PCA) and then inpaints the occluded part (using FoE prior) to improve face recognition. We have demonstrated the significant improvements of various face recognition systems based on the proposed approach via extensive experiments.

5. REFERENCES

[1] A. M. Martínez, "Recognizing Imprecisely Localized, Partially Occluded, and Expression Variant Faces from a Single Sample per Class," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 6, pp. 748-763, Jun. 2002.

[2] J. Wright, A. Ganesh, A. Yang, and Y. Ma, "Robust face recognition via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.31, no.2, pp. 210-227, 2009.

[3] H. J. Oh, K. M. Lee, and S. U. Lee, "Occlusion invariant face recognition using selective local non-negative matrix factorization basis images," Image Vision Comput., vol. 26, no. 11, pp. 1515-1523, Nov. 2008.

[4] R. Min, A. Hadid, J.-L. Dugelay, "Improving the recognition of faces occluded by facial accessories," 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG 2011), pp.442-447, 21-25 March 2011

[5] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, Y. Ma , "Face recognition with contiguous occlusion using markov random fields," IEEE 12th International Conference on Computer Vision (ICCV 2009), pp.1050-1057, Sept. 29 2009-Oct. 2 2009

[6] R. Diestel, (2005), Graph Theory, Graduate Texts in Mathematics, Springer-Verlag

[7] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. 2000, "Image inpainting." In Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH '00).

[8] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis?" Journal of the ACM, volume 58, no. 3, May 2011.

[9] S. Roth and M. J. Black, "Fields of Experts." International Journal of Computer Vision (IJCV), 82(2):205-229, April 2009.

[10] M. Turk, and A. Pentland, 1991. "Eigenfaces for recognition," J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, Jan. 1991.

[11] D.G. Lowe, "Object recognition from local scale-invariant features," The Proceedings of the Seventh IEEE International Conference on Computer Vision, (ICCV 1999), vol.2, no., pp.1150-1157 vol.2, 1999

[12] T . Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," Computer Vision, ECCV 2004 Proceedings, Lecture Notes in Computer Science 3021, Springer, 469-481.

[13] D. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," AMS Math Challenges Lecture, 2000.

[14] Z. Lin, M. Chen, and Y. Ma, "The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices," UIUC Technical Report UILU-ENG-09-2214, October 2010.

[15] A.M. Martinez and R. Benavente, " The AR face database," Technical report, CVC Technical report, no. 24, 1998.