VIDEO FACE CLUSTERING VIA CONSTRAINED SPARSE REPRESENTATION

Chengju Zhou[1]*,Changqing Zhang*[1]*,Xuewei Li*[1]*,Gaotao Shi*[1]*,Xiaochun Cao*[1,2*]

[1]School of Computer Science and Technology, Tianjin Universtiy, Tianjin 300072, China
[2]SKL Of Information Security Institute of Information Engineering, CAS, Beijing 100093, China
{zhoucj, zhangchangqing, lixuewei,shgt}@tju.edu.cn,caoxiaochun@iie.ac.cn

ABSTRACT

In this paper, we focus on the problem of clustering faces in videos. Different from traditional clustering on a collection of facial images, a video provides some inherent benefits: faces from a face track must belong to the same person and faces from a video frame can not be the same person. These benefits can be used to enhance the clustering performance. More precisely, we convert the above benefits into must-link and cannot-link constraints. These constraints are further effectively incorporated into our novel algorithm, Video Face Clustering via Constrained Sparse Representation (CS-VFC). The CS-VFC utilizes the constraints in two stages, including sparse representation and spectral clustering. Experiments on real-world videos show the improvements of our algorithm over the state-of-the-art methods.

Index Terms— video face clustering, constrained sparse representation

1. INTRODUCTION

Face clustering aims to divide the facial images into different subsets according to different persons. This technique can be used in many applications [1, 2, 3], including content-based retrieval, automatic cast listing in feature-length films, rapid browsing and organization of video collections. However, the task of face clustering is still very challenging. In real-world videos, lighting conditions especially light angle, facial expressions and head poses drastically change the appearance of faces. Moreover, partial occlusions caused by objects in front of a face and hair style changes also increase the difficulties.

General face clustering often distinguishes the different individuals based only on the facial similarities. In video face clustering case, there is some additional information that can be used to improve the performance. In [4], the scripts and subtitles are used to obtain cues as to which characters are present. These weak cues for character presence are then combined with facial similarities to help face clustering. However, such text-based information is not always available. Fortunately, there are some inherent benefits in videos: faces in a face track must be the same person while faces can not be the same person if they appear together in a video frame. These facts can be converted into must-link and cannot-link constraints. Then these constraints are combined with facial similarities to improve face clustering.

While a large body of work has been conducted on face recognition, the face clustering is a relatively less unattentioned topic with few publications in the literature. [5, 6] focus on how to induce invariance for pose, lighting and expression. In [1, 2, 7], the above constraints are used as a weak prior. In order to utilize the inherent benefits of video more effectively, we propose a novel method, Video Face Clustering via Constrained Sparse Representation, which incorporates the must-link and cannot-link constraints into sparse representation as strong prior and spectral clustering as weak prior respectively.

The contributions of this paper include: (1)A fully automatic end-to-end system for video face clustering is developed, which includes face tracking, face alignment and face clustering. (2) By using the inherent benefits of videos, we derive a novel algorithm, CS-VFC, which performs video face clustering via 2-step manner: constrained sparse representation and constrained spectral clustering. (3) We compare the proposed method with the state-of-the-art methods and show its improvements on the real-world data.

The rest of this paper is organized as follows: Section 2 discusses the related work on video face clustering. Then section 3 describes our entire framework for video face clustering from detection to clustering. Next, Section 4 evaluates our method on two datasets. Finally, we present our conclusions in Section 5.

2. RELATED WORK

Most existing methods on face clustering focus on obtaining a good representation for the structure of the interpersonal dissimilarities from the unlabeled faces. An affine invariant distance metric is proposed in [5, 6], which is robust to a desired group of transformations. Wang et al. [8] use a metric called Manifold-Manifold Distance (MMD) to

* Corresponding Author

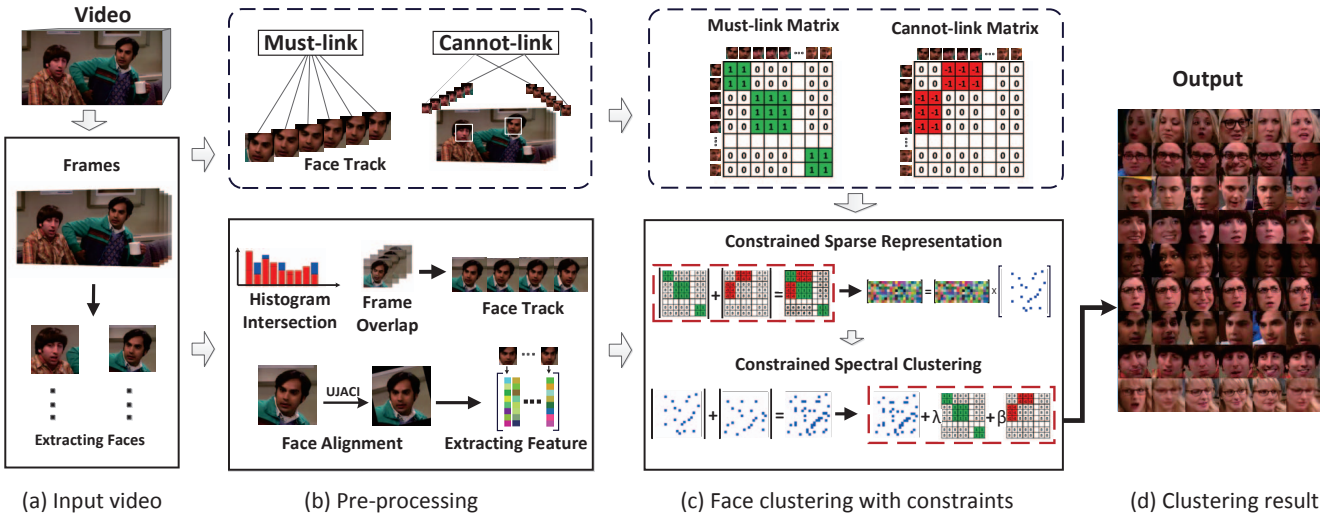


Fig. 1. The framework of Video Face Clustering via Constrained Sparse Representation. With the input video (a), we perform pre-processing to obtain facial features and constraints (b). Then, we build up the must-link and cannot-link matrix (c) respectively. Based on these constraints, we perform our novel algorithm CS-VFC in two steps, constrained sparse representation and constrained spectral clustering (c), to get the clustering result (d).

calculate the distance between manifolds each of which represents one image set. In fact, a video provides the temporal cues above mentioned which can be utilized in clustering. [2] presents a method called unsupervised logistic discriminative metric learning (ULDML). A metric is learned that must-linked faces are close, while cannot-linked faces are far from each other. The latest work on face clustering with constraints is presented in [7]. The clustering model is based on Hidden Markov Random Fields (HMRFs), in which the pairwise constraints are augmented by label-level and constraint-level local smoothness to conduct the clustering process. However, such methods either focus on obtaining distance metric or how to use the constraints to help clustering. In our method, we not only use sparse representation to obtain a robust metric, but also effectively utilize the constraints by two steps, constrained sparse representation and constrained spectral clustering.

In the literature of clustering, spectral clustering is one of the most popular clustering algorithms. It is simple to implement and can be solved efficiently by standard linear algebra methods. To incorporate the constraints in spectral clustering, Lu et al. [9] propose a Gaussian based method to propagate the pairwise constraints over the original affinity matrix. Wang et al. [10] propose a principled and flexible framework for constrained spectral clustering that explicitly encodes the constraints as a part of a constrained optimization problem. Nevertheless, there are some limitations of their approaches. In [9], there would be a must-link between face **A** and face **B** if **A** and **B** are cannot-linked to face **C**. This might lead to many false clustering considering the face track cases. The approach [10] works well only when the number of must-links and cannot-links is approximately balanced. In real-world videos, the overlapped tracks are often much less than the whole tracks, which results in a huge difference between the number of must-links and cannot-links. Hence, their approaches are limited in the video case. Different from utilizing the constraints as hard manner [9], we use the constraints as hard manner and soft manner in two stages respectively, which avoiding the limitation of [9]. Compared with [10], our method provides coherent performance for different number of the must-links and cannot-links. In our method, the constraints are utilized in the whole process, which includes the constrained sparse representation and the constrained spectral clustering. With these two steps, our method obtains a robust relationship metric as well as an efficient solution of constrained spectral clustering problem.

3. THE APPROACH

We start with a detailed description of our system as shown in Fig. 1. First, we extract the frames and faces from the input video as input (a). Then, by pre-processing (b), the constraint cues and features are obtained. Based on the constraints, the must-link and cannot-link information, we build up constrained matrix by integrating the two constraints. We perform sparse representation with constraints, obtaining the sparse coefficient matrix and finally, apply spectral clustering with constraints on the similarity matrix (c) to get the final clustering result (d).

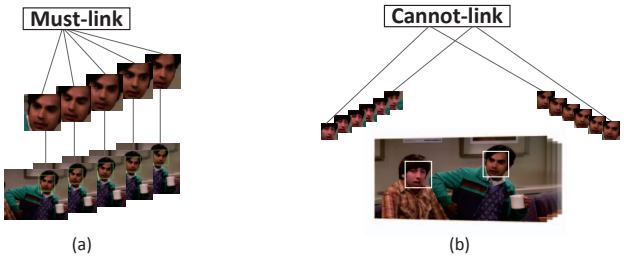


Fig. 2. The inherent constraints used in our method: (a) the must-link faces from the same tracks; (b) the cannot-link faces from the overlapped tracks.

3.1. Pre-processing

Generally, face tracking is the first step in video face clustering algorithms. In this paper, the face detector in [11] is used to get an initial set of detections. In order to link the detected faces into face tracks, the approach of [7] is employed, which consists of two metrics: Histogram Intersection and Frame Overlap as shown in bottom of Fig. 1 (b). Many of the false positives of the face detectors do not have temporal supports. Therefore, such false detections can be easily eliminated by only selecting the tracks with a sufficiently large number of faces. In our implementation, we filter out the tracks with less than 7 faces. After extracting face tracks, we employ [12] to align the faces, and extract the gray scale value as feature, which is simple and fast to implement.

3.2. Constraints

There are two inherent constraints in videos. Fig. 2 illustrates the constraints. To utilize the inherent benefits of a video, we build up the must-link matrix $\mathbf{M} \in \mathbb{R}^{N\mathbf{x}N}$ and the cannot-link matrix $\mathbf{C} \in \mathbb{R}^{N\mathbf{x}N}$, where N is the number of the total faces of the video. The matrix $\mathbf{M}$ represents the must-link constraints, where the indices of faces in the same track are set to 1 while others are 0, i.e., M_{ij} (*i* and *j* are respectively the index of the face *i* and face *j* from the same face track) is set to 1. The matrix $\mathbf{C}$ represents the cannot-link constraints, where the indices of faces belong to the overlapped tracks are set to -1 while others are 0, i.e., C_{ij} (*i* and *j* are respectively the index of the face *i* and face *j* from the overlapped face tracks) is set to -1. We also define $\mathcal{M}$ and $\mathcal{C}$ as the set of the must-link and cannot-link constraints, which correspond to the must-link matrix $\mathbf{M}$ and the cannot-link matrix $\mathbf{C}$ respectively. And we define $\mathcal{I}$ as the set of indices corresponding to the elements with value 1 in the identity matrix $\mathbf{I} \in \mathbb{R}^{N\mathbf{x}N}$.

3.3. Clustering Method

In this section, we describe our constrained video face clustering approach, CS-VFC, in two steps, including constrained sparse representation and constrained spectral clustering.

3.3.1. Constrained Sparse Representation

Ideally, given a set of facial images, which is represented as $\mathbf{Y} \triangleq \begin{bmatrix}\mathbf{y}_1 & \mathbf{y}_2 & \dots & \mathbf{y}_N\end{bmatrix}$, the ***i***-th face $\mathbf{y}_i$ can be sparsely represented by a small subset of faces from the same person [13, 14] in the dataset. The relationship can be written as:

$$\mathbf{y}_i = \mathbf{Y}\mathbf{a}_i, \quad a_{ii} = 0, \tag{1}$$

where $\mathbf{a}_i \triangleq \begin{bmatrix}a_{1i} & a_{2i} & .. & a_{Ni}\end{bmatrix}^\mathsf{T}$ and the constraint $a_{ii} = 0$ eliminates the trivial solution of representing a face as a linear combination of itself. Therefore, the coefficient vector $\mathbf{a}_i$ should only have non-zero entries for these few facial images from the same person and zeros from the rest. In other words, the matrix of faces $\mathbf{Y}$ is a *self-expressive* [13] dictionary in which each face can be rewritten as a linear combination of other faces in $\mathbf{Y}$.

Note that some relationships among faces have been known from the must-link and the cannot-link constraints, so we pour attention into exploring the unknown relationships. Hence, the above formulation is written as:

$$\mathbf{y}_i = \mathbf{Y}\mathbf{a}_i, \quad a_{ji} = 0, \quad (j,i) \in (\mathcal{M} \cup \mathcal{C} \cup \mathcal{I}), \tag{2}$$

where $a_{ji} = 0, (j,i) \in (\mathcal{M} \cup \mathcal{C})$ eliminates the solutions of must-links and the cannot-links. However, the representation of $\mathbf{y}_i$ in the dictionary $\mathbf{Y}$ is *not unique* in general. Since we are interested in efficiently finding a non-trivial sparse representation of $\mathbf{y}_i$ in the dataset $\mathbf{Y}$, we use the tightest convex relaxation of the ℓ^1-norm, i.e.,

$$\min \left\|\mathbf{a}_i\right\|_1 \text{ s.t. } \mathbf{y}_i = \mathbf{Y}\mathbf{a}_i, \ a_{ji} = 0, \ (j,i) \in (\mathcal{M} \cup \mathcal{C} \cup \mathcal{I}), \tag{3}$$

which can be solved efficiently using convex programming tools [15, 16] and is known as the sparse solutions [17, 18]. Without loss of the generality, we can rewrite the sparse optimization program (3) for all faces $i = 1, ..., N$ in matrix form as:

$$\min \left\|\mathbf{A}\right\|_1 \text{ s.t. } \mathbf{Y} = \mathbf{Y}\mathbf{A}, \ A_{ji} = 0, \ (j,i) \in (\mathcal{M} \cup \mathcal{C} \cup \mathcal{I}), \tag{4}$$

where $\mathbf{A} \triangleq \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & .. & \mathbf{a}_N\end{bmatrix} \in \mathbb{R}^{N\mathbf{x}N}$ is the matrix, and the ***i***-th column of which corresponds to the sparse representation of $\mathbf{y}_i$, and $\mathbf{a}_i \in \mathbb{R}^{\mathbb{N}}$ is the vector of elements of $\mathbf{A}$.

Ideally, the solution of equation (4) corresponds to the sparse representation of the facial images, which can be used to infer the clustering process. Next, we will illustrate our clustering method based on the sparse coefficient matrix obtained in this step.

3.3.2. Constrained Spectral Clustering

After solving the proposed optimization program in equation (4), we build a weight graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V}$

donates the *N* nodes in graph corresponding to the *N* faces in facial set, and the $\mathcal{E}$ donates the edges between nodes. $\mathbf{W} \in \mathbb{R}^{N\mathbf{x}N}$ is a symmetric non-negative similarity matrix representing the weights of the edges. An ideal similarity graph $\mathcal{G}$ should only has connected faces from the same persons and has no connections among faces from different persons.

In the sparse representation solution $\mathbf{A}$, nonzero elements correspond to faces from the same person and the coefficients can be regarded as a measurement of relationships between faces. This provides an choice of constructing similarity matrix as:

$$\mathbf{W} = |\mathbf{A}| + |\mathbf{A}|^{\mathsf{T}}, \tag{5}$$

where the nodes *i* and *j* can get connected to each other with the same weight score in the weight graph. Note that, we normalize $\mathbf{A}$ as $\mathbf{a}_i \leftarrow \mathbf{a}_i / \|\mathbf{a}_i\|_\infty$ to lead the weights in similarity graph to be of a same scale. Moreover, the elements corresponding to the constraints are 0 in the sparse coefficient matrix $\mathbf{A}$ and so is the similarity matrix $\mathbf{W}$. Therefore, we use a straightforward combination approach to incorporate the must-link and cannot-link constraints into the similarity matrix. It can be written as:

$$\mathbf{W}^{const} = \mathbf{W} + \lambda\mathbf{M} + \beta\mathbf{C}, \tag{6}$$

where λ and β are trade-off parameters. The parameter λ is user-defined, and we suggest it is larger than the max value in $\mathbf{W}$. Specially, the cannot-link constraints mean that they should have no connections in weight graph $\mathcal{G}$. Therefore, the elements corresponding to cannot-links in similarity matrix $\mathbf{W}^{const}$ are set to 0 in our implementation. Then, the spectral clustering method is conducted on the weight graph $\mathcal{G}$ with the new similarity matrix $\mathbf{W}^{const}$ to get the final clustering results. The CS-VFC method is summarized in Algorithm 1.

Algorithm 1 Video Face Clustering via Constrained Sparse Representation

Input: the video ***V***, cluster number *k*.
Output: *k* disjoint face clusters Y_1 ... Y_k.

1: Perform pre-processing on ***V***, obtaining the face matrix $\mathbf{Y}$, must-link matrix $\mathbf{M}$ and cannot-link matrix $\mathbf{C}$;
2: Solve Eq. 4, obtaining the sparse coefficient matrix $\mathbf{A}$, and normalize $\mathbf{A}$ as $\mathbf{a}_i \leftarrow \mathbf{a}_i / \|\mathbf{a}_i\|_\infty$;
3: Construct similarity matrix $\mathbf{W}$ using Eq. 5, and then build up the final similarity matrix $\mathbf{W}^{const}$ by Eq. 6;
4: Apply spectral clustering on weight graph $\mathcal{G}$ with cluster number *k*.

4. EXPERIMENTS

In this section, we evaluate our face clustering method on two datasets and compare it to several methods, including COP-KMEANS [19], SSC [13] and HMRFs [7].

4.1. Experimental Setting

Datasets: The dataset Notting-Hill [7, 3] is derived from the movie "Notting-Hill". Faces of 5 main casts are used, including 4660 faces in 76 tracks. The original dataset consists of 120x150 facial images. 6 faces are uniformly sampled from each track and then a subset of 456 faces is obtained. To reduce the computational cost and the memory requirements, we downsample each facial image to 40x50 and get the 2000-dimensional features. We use this feature in COP-KMEANS [19], SSC [13] and our method. In HMRFs [7], we use PCA to project the feature's dimension to which is equal to the number of casts.

We build up the dataset TBBTS06E12 from the Season 6 Episodes 12 of TV series "The Big Bang Theory". The detected faces of 9 main casts are used, including 17168 faces in 385 tracks. To reduce the computational cost and the memory requirements, we downsample facial images to 50x50 and use the 2500-dimensional vector as features. This feature is used in COP-KMEANS, SSC and our method. In HMRFs, PCA is used to project the original gray scale feature space to a lower dimensional space which is equal to the number of casts.

Evaluation Criteria: We evaluate the performance based on confusion matrix, which is derived from the match between the predicted labels of all faces and the ground-truth labels. We compare four algorithms: CS-VFC (ours), HMRFs, COP-KMEANS and SSC. More importantly, we test these algorithms except SSC in four cases: with all-links, with only cannot-links, with only must-links and with no links. All the experiments are repeated 10 times, and the mean value is used as the average accuracy except the HMRFs in Table 2 which is reported from [7].

4.2. Quantitative Results

The results are shown in Table 1 and Table 2. Our method outperforms the latest best method, HMRFs, in both datasets. In Table 1, the CS-VFC outperforms all other method in all cases: at least 5.45% increase by introducing cannot-links, 17.95% by adding must-links and 17.64% in all-links. The result of CS-VFC in all-link case is higher than that without links about 18.47%. Our method in cannot-link case is slightly better than HMRFs in all-link case despite the number of cannot-links is much less than the number of all-links. Table 2 shows the clustering result on Notting-Hill. The CS-VFC$_{19}$ means that we sample 19 faces from each track while other methods sample 6 faces. Our method still achieve better results than all others in all cases: at least 2.37% ,11.66% and 7.56% in three cases, respectively. Especially the CS-VFC$_{19}$ in all-links case, our method achieves 96.05% accuracy. The results of CS-VFC in all-link cases are much better than the no-link cases. Note that HMRFs [7] in no-link cases are not presented duo to the HMRFs needs constraints to build up the neighborhood system which takes the key role in HMRFs. On the other hand, the clustering results without links on both

Table 1. Constrained Face Clustering on TBBTS06E12

Methods	No-link	Cannot-link	Must-link	All-link
COP-KMEANS	52.67	37.40	38.96	47.01
HMRFs	-	58.18	60.51	63.37
CS-VFC	**62.54**	**63.63**	**78.46**	**81.01**

Table 2. Constrained Face Clustering on Notting-Hill

Methods	No-link	Cannot-link	Must-link	All-link
COP-KMEANS	55.26	46.05	52.63	53.95
HMRFs	-	80.52	81.76	85.86
CS-VFC	**76.31**	**82.89**	**93.42**	**93.42**
CS-VFC$_{19}$	**70.26**	**92.10**	**94.73**	**96.05**

TBBTS06E12 and Notting-Hill are coherently much lower than all-link cases. This demonstrates that the utilized constraints can significantly improve the clustering performance.

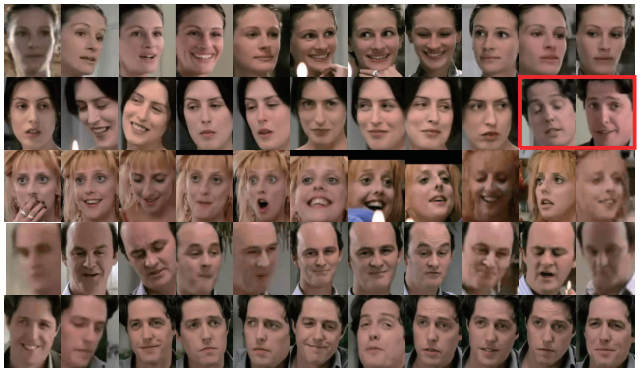
4.3. Qualitative Results

The clustering examples of HMRFs and CS-VFC on Notting-Hill are shown in Fig. 3, including 5 clusters in 5 rows respectively. 11 faces are randomly chosen from the final clusters. In the clustering result of HMRFs, all clusters occur incorrect faces except the 3-th. Especially the 5-th, about half clustering faces are incorrect. Our method achieves a fairly better clustering result. The clustering accuracies of 5 clusters (top-down) are: 100%, 78.58%, 100%, 100% and 100% respectively. Only the 2-th cluster occurs incorrect faces and the incorrect ratio is 21.42%, so the whole clustering accuracy is up to 96.05%. One main reason of the incorrect clustering may be the very similiar face and hair style in vision. As shown in Fig. 3, the average accuracy 96.05% is an encouraging result for the difficult conditions where the facial images have different poses, facial expressions and occlusions.

4.4. Effects of Varying Face Sampling Number and Parameter λ

The goal of face clustering is to obtain a precise accuracy with limited computation. The face sampling number from tracks often affects both the clustering accuracy and the computational cost. We conduct it on the two datasets with all-link constraints, and test the influence of different sampling number using four algorithms: CS-VFC (ours), HMRFs [7], SSC [13] and COP-KMEANS [19]. Note that in TBBTS06E12, a slightly wider range of sampling number is taken for a better measurement of influence of sampling number. The average clustering accuracies are shown in Fig. 4 (a) and (b). All of the clustering methods, except the SSC, achieve a stable performance in different sampling numbers. Though SSC obtains capable clustering accuracies in some cases, its best performance is still lower than ours in both

(a) HMRFs



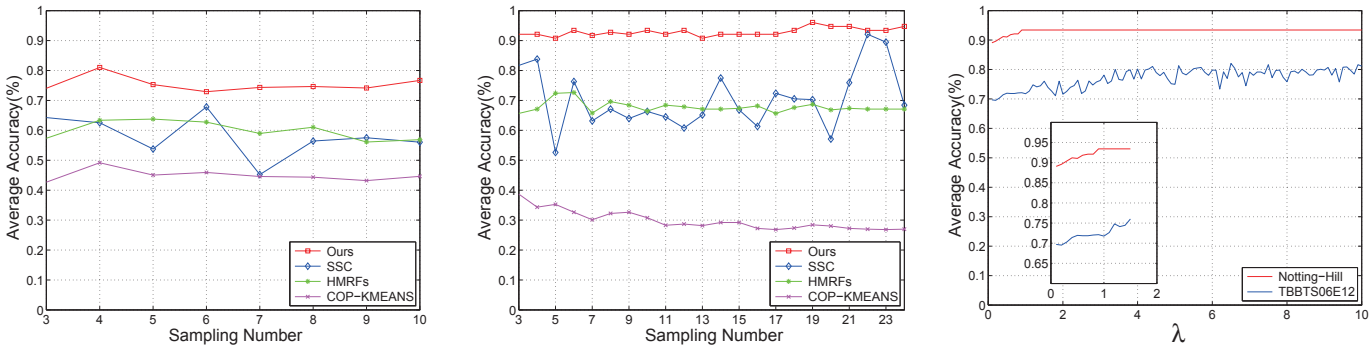
(b) CS-VFC

Fig. 3. The clustering results of HMRFs and CS-VFC on Notting-Hill. The false clustering faces are highlighted by the red rectangles and the incorrect rate in each row is approximately equal to its proportion in the clusters.

datasets. This indicates that the must-link and cannot-link constraints can significantly improve performance and stability of the clustering method. The performance of HMRFs is much more stable than SSC. Nevertheless, compared with our method, its clustering accuracy is about 10% lower on TBBTS06E12 and 20% lower on Notting-Hill. Fig. 4 (c) shows the clustering accuracies on different parameter λ. Our method can achieve a pleasurable performance as long as the parameter λ is large enough. All above demonstrates that our method not only is robust to the sampling number, but also obtains a much better clustering result than the existing methods. As a result, our method can achieve a fairly respectable performance with limited computational cost.

5. CONCLUSION

This paper shows how to utilize the inherent benefits of video to help face clustering. Together with the inherent benefits, we propose a novel algorithm, Video Face Clustering via Constrained Sparse Representation (CS-VFC), in which the inherent benefits are used as must-link and cannot-



(a) TBBTS06E12 (b) Notting-Hill (c) Parameter λ tuning of the datasets

Fig. 4. The performance of clustering with respect to different sampling numbers and different λ

link constraints. We use the must-links and cannot-links in the whole process, including constrained sparse representation and constrained spectral clustering. In sparse representation, the constraints are utilized as hard manner, which makes sparse representation focus on exploring the unknown relationships among faces. The constraints are used as soft manner in spectral clustering, and the weight of constraints can be easily adjusted. We also test the influence of different sampling number and show our method is capable to obtain a superior result with limited computational cost. Experiments on two face datasets from real-world videos have demonstrated the improved performance of our algorithm.

Acknowledgements. This work was supported by National Natural Science Foundation of China (No. 61332012, 61272264), R&D Program of China (2012BAH07B01), National Basic Research Program of China (2013CB329305), 100 Talents Programme of The Chinese Academy of Sciences, and the Opening Project of State Key Laboratory of Digital Publishing Technology.

6. REFERENCES

[1] N. Vretos, V. Solachidis, and I. Pitas, "A mutual information based face clustering algorithm for movie content analysis," *Journal of Image and Vision Computing*, vol. 29, no. 10, pp. 693–705, 2011.

[2] R. G. Cinbis, J. Verbeek, and C. Schmid, "Unsupervised metric learning for face identification in tv video," in *ICCV*, 2011, pp. 1559–1566.

[3] Y. F. Zhang, C. S. Xu, H. Lu, and Y. Huang, "Character identification in feature-length films using global face-name matching," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pp. 1276–1288, 2009.

[4] T. Cour, B. Sapp, A. Nagle, and B. Taskar, "Talking pictures: Temporal grouping and dialog-supervised person recognition," in *CVPR*, 2010, pp. 1014–1021.

[5] A. Fitzgibbon and A Zisserman, "On affine invariant clustering and automatic cast listing in movies," in *ECCV*, pp. 304–320. 2002.

[6] A. W. Fitzgibbon and A Zisserman, "Joint manifold distance: A new approach to appearance based clustering," in *CVPR*, 2003, vol. 1, pp. 26–33.

[7] B. Y. Wu, Y. F. Zhang, B. G. Hu, and Q. Ji, "Constrained clustering and its application to face clustering in videos," in *CVPR*, 2013, pp. 3507–3514.

[8] R. Wang, S. G. Shan, X. L. Chen, and W. Gao, "Manifold-manifold distance with application to face recognition based on image set," in *CVPR*, 2008.

[9] Z. D. Lu and M. A. Carreira-Perpinán, "Constrained spectral clustering through affinity propagation," in *CVPR*, 2008.

[10] X. Wang, B. Y. Qian, and I. Davidson, "On constrained spectral clustering and its applications," *Journal of Data Mining and Knowledge Discovery*, pp. 1–30, 2012.

[11] P. Viola and M. Jones, "Robust real-time face detection," *IJCV*, vol. 57, no. 2, pp. 137–154, 2004.

[12] G. B. Huang, V. Jain, and E. Learned-Miller, "Unsupervised joint alignment of complex images," in *ICCV*, 2007.

[13] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *Arxiv preprint arXiv:1203.1005*, 2012.

[14] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *PAMI*, vol. 31, no. 2, pp. 210–227, 2009.

[15] S. P. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[16] S. J. Kim, K. Koh, S. Lustig, M. Byod, and D. Gorinevsky, "An interior-point method for large-scale l1-regularized logistic regression.," *Journal of Machine learning research*, vol. 8, no. 8, pp. 1519–1555, 2007.

[17] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[18] G. L. Chen and G. Lerman, "Spectral curvature clustering (scc)," *IJCV*, vol. 81, no. 3, pp. 317–330, 2009.

[19] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl, "Constrained k-means clustering with background knowledge," in *ICML*, 2001, vol. 1, pp. 577–584.