ISSN:0975-9646

Ashish R Tulshi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5041-5046

Automatic Face Detection & Identification Using Robust Face Name Graph Matching In Video & Live Streaming

Ashish R Tulshi, Prof.Praveen Bhanodiya.

Computer Engineering Department
P.C.S.T. Indore, R.G.P.V. Bhopal, Madhya Pradesh, India.

ABSTRACT: **Automatic face identification of characters in movies has drawn significant research interests and led to many interesting the applications. And it is the challenging problem due to the huge variation in the appearance of each character. Although these existing methods are demonstrate promising results in this the clean environment that the performances are limited in the complex movie scenes due to the noises can be generated during the face tracking and face clustering process. In this paper we can present two schemes of global face-name matching and based on the framework for robust character identification. In this the contributions of this work can be include: 1) A noise insensitive character and the relationship representation is incorporated. 2) We can introduce an edit operation based on the graph matching algorithm. 3) Complex character changes are can be handled by simultaneously graph partition and graph matching. 4) Beyond the existing character identification can be approaches; we can further perform an in-depth sensitivity analysis by introducing two types of simulated noises. There are the proposed schemes demonstrate state-of-the-art performance on this the movie character identification in various genres of movies.**

***Keywords*—face name matching, graph matching, character identification.**

I. INTRODUCTION

A. Objective and Motivation

When the proliferation of movie and TV provides the large amount of the digital video data and effective techniques for video content understanding and the objective is to identify the faces of the characters in the video and label them with the corresponding names in the cast. The textual cues, in the movie, characters are the focus center of interests for the audience.

Textual cues [6]. There are ambiguity problem in establishing the correspondence between names and faces: ambiguity can arise from a reaction shot where the person speaking may not be shown in the frames1; ambiguity can also arise in partially labeled frames when there are multiple speakers in the same scene2. 2) Face identification in videos is more difficult than that in images [5].Low resolution, occlusion, non rigid the character identification. 3) Then the same character appears quite differently during the movie [3]. There may be huge pose, expression and illumination variation, wearing, clothing, even makeup actors playing

Fig.1. Examples of character identification from movie "Notting Hill".

B. Related Work

Then the crux of the character identification problem is to be exploit the relations between videos and the associated texts in order to identifying faces in the names of characters are seldom directly shown in the subtitle names have no time stamps to align we roughly divide the identification methods into three categories.

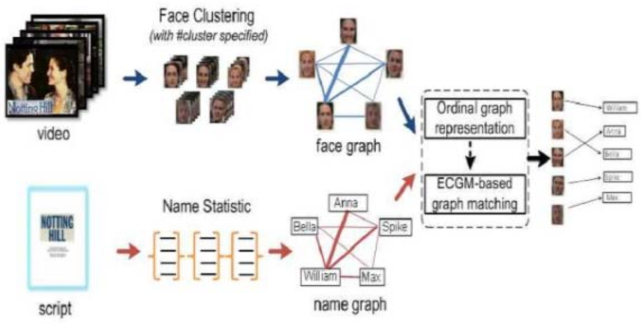

Fig.2. Framework of scheme 1: Face-name graph matching with #cluster pre specified.

1) Category 1: Cast list based:

These methods only utilize the case faces are clustered by appearance and faces of a particular character are expected to be collected in a few selected from the cast list. Raman an et al. proposed to manually label an initial set of face clusters and further cluster the rest face instances based on clothing within scenes [4]. In [5], the authors have addressed the problem of finding particular characters by building a model/classifier of the character's appearance from user-provided identification with web image retrieval is proposed in [6]. The character names in the cast are used as queries to search face images and identified as one of the characters by multi-task joint sparse representation and classification. Newly, metric learning is introduced into character identification in uncontrolled videos [5]. Cast specific metrics are adapted to the people appearing in a particular video in. The clustering as well as identification performance are demonstrated however, without other textual cues, they either need manual labeling or guarantee no robust clustering and classification performance due to the large intra-class variances.

2) Category 2: Subtitle or Closed caption and local matching based:

Subtitle and closed caption provide the proposed to combine the film script with the subtitle for local face-name matching. The rest of the faces were then classified into these exemplars for identification. They further extended their work in [1], by replacing the nearest neighbor classifier by multiple kernels learning utilized the readily available more reliable than OCR-based subtitles [2], [5]. They investigated on the ambiguity issues in the local alignment multiclass classification character identification problem without the use of screenplay [1]. The reference cues in the closed captions are employed as multiple instance constraints and face tracks grouping as well as face-name association are solved in a the time-stamped information, which is either extracted by OCR (i.e., subtitle) or unavailable for the majority of movies and TV series (i.e., Closed caption). Besides, the ambiguous and partial annotation makes local matching based methods more sensitive to the face detection and tracking noises.

3) Category 3: Script and Screenplay, the Global matching based:

Global matching based methods open the possibility of character identification without OCR-based subtitle or closed caption. Since it is not easy to get formulated as a global matching problem in [2], [2], and [4]. Our method belongs to this category and can be considered as an extension to Zhang's work [2]. In movies, the names of characters seldom directly appear the task of character identification is formulated as a global matching problem between the faces detected from the video and the names extracted from are used for

• Regarding the fact that characters may show can be affected by the noise introduced by some statistic properties are preserved in spite for character relationship and introduce a name-face matching method which can accommodate a certain noise.

• Face track clustering serves as an important step in movie character are utilized to determine the number of target clusters prior to face clustering, e.g., in [2], the number of clusters is the same as the number of distinct speakers appearing results sometimes. In this paper, we lose the restriction of one face cluster corresponding to one are jointly optimized and conducted in a unique framework.

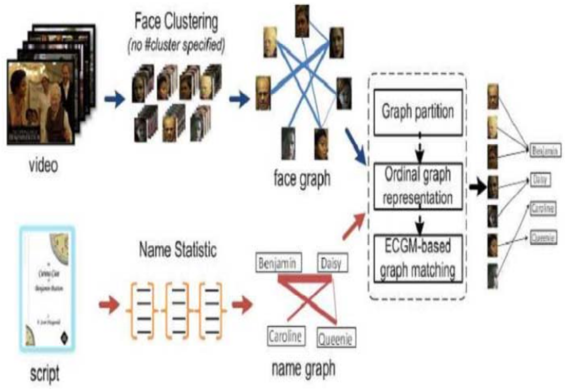

Fig.3. Framework of scheme 2: Face-name graph matching without #cluster pre-specified.

• Sensitivity analysis is common in financial area where Models are developed [3], [4]. Good modeling practice requires that the modeler provides an evaluation of the confidence associated with the modeling process and with the outcome of the analysis offers valid tools for characterizing the robustness to been no efforts directed at the sensitivity analysis for movie character introducing two types of simulated noises. A preliminary version of this work was introduced by [1]. We provide additional algorithmic and pre-specification for the number of face clusters. Improved performances as well as robustness are demonstrated in movies with large character appearance changes.

C. Overview of Our Approach

In this paper, we propose a global face-name graph matching based framework for robust movie character identification two schemes are measured. There are connections as well as differences between them. About the connections, initially, the proposed two schemes both belong to the global matching name graph representation and a novel graph matching algorithm called Error Correcting Graph Matching (ECGM) the number of clusters when performing face clustering (e.g., K-means, number of vertexes with required and face tracks are clustered based on their intrinsic data structure (e.g., mean shift, affinity propagation). Moreover, as partition compared as an extension to scheme 1.

1) Scheme 1:

The proposed framework for scheme 1 is shown in Fig.2. It is similar to the framework of [2]. Face tracks are clustered using number of clusters in video constitutes the corresponding face graph framework by using ordinal graphs for robust representation and introducing an ECGM-based graph matching method. For face and name graph in rank ordinal Level [5], which scores the strength of the relationships in a rank order from the weakest and thus are less sensitive global matching is interval measures of the co-occurrence relationship of relationship holds. For name-face graph matching, we utilize the ECGM algorithm. In ECGM, then the difference between two graphs is measured by edit distance which is a sequence of graph appropriate graph edit operations and adapts the distance functions to obtain improved name-face matching performance.

2) Scheme 2:

The proposed framework for scheme 2 is the face tracks may have different ordinal graph representation. The basic premise behind the scheme 2 is that appearances of the same character vary significantly and it is difficult to group them in a unique cluster. Take the movie "The Curious Case of Benjamin than the face clusters as the number of characters will disturb the character is grouped together. In scheme 2, we utilize affinity propagation for the face of clusters, through appearance-based similarity transmit to several (i.e., divided into the same sub graph) is determined by whether the partitioned face graph achieves an optimal graph matching with two steps: coarse clustering by appearance and further are optimized and noises. In advance and the face tracks are clustered that can be based on their intrinsic robustness to change appearance significantly or go through a pedestrians whose face is detected and added into the same as that of name from movie cast will deteriorate Chance that movie cast does not cover all face clusters is risky: face tracks from different characters will be mixed together and graph matching tends to fail.

3) The Sensitivity Analysis:

In the Sensitivity analysis play an important role in characterizing the uncertainties associated and the performance of name-face matching with respect to the simulated noises. We can introduce the ordinal graph representation and ECGM-based graph matching of scheme 1 graph noises for conducted and discussed.

II. SCHEME 1: FACE-NAME GRAPH MATCHING WITH NUMBER OF CLUSTER SPECIFIED

In this section we first briefly review the framework of traditional global graph matching based character identification. Based on this investigations of the noises generated during the affinity graph, graph in rank ordinal level and employ ECGM with specially designed edit cost function for face-name matching.

I. Review of Global Face-name Matching Framework

In a movie, the interactions among characters resemble them into the relationship network. When the co-occurrence of names in script and faces in videos can be represent such interactions. Similarity graph is built according to the co-occurrence status can be represented as the weighted graph G = {V, E} where vertex V denotes the characters and edge E denotes edge weights script analysis and a face affinity graph from video analysis can be corresponding to the name and face affinity graphs from the movie then the interval [0, 1]. We can see that some of the face affinity values differ much from the corresponding name affinity values (e.g. {WIL, SPI} and {Face1, Face2}, {WIL, BEL} and {Face1, Face5}) due to formulated as the problem of finding optimal vertex to vertex matching to be find the optimal name-face correspondence. Technical details can be referred to [2].

B. Ordinal Graph Representation

The name affinity graph and face affinity graph are built based on the co-occurrence transform from the name affinity graph by in the generated properties of the characters are relatively stable and insensitive with character B than C, character D has never co-occurred with the absolute Quantitative affinity values of these characters

	WIL	SPI	ANN	MAX	BEL
WIL	0.173	0.024	0.129	0.009	0.013
SPI	0.024	0.017	0.007	0.001	0.002
ANN	0.129	0.007	0.144	0	0
MAX	0.009	0.001	0	0.009	0.006
BEL	0.013	0.002	0	0.006	0.011

(a)

	Face1	Face2	Face3	Face4	Face5
Face1	0.186	0.041	0.147	0.008	0.021
Face2	0.041	0.012	0.005	0.002	0.004
Face3	0.147	0.005	0.157	0	0.003
Face4	0.008	0.002	0	0.005	0.007
Face5	0.021	0.004	0.003	0.007	0.009

(b)

Fig.4. Example of affinity matrices from movie "Notting Hill": (a) Name affinity matrix R^{name} (b) Face affinity matrix R^{face}

(E.g. A is more closer to B than to C) and the qualitative affinity values (e.g. whether D has co-occurred with A) usually the statistic properties and propose to represent the character co-occurrence in rank order. We denote the original affinity matrix as R = {rij} N×N, where N is the number of characters. Initial we look at the cells along the main diagonal (e.g. A co-occur with A, B co-occurs with B). We rank the diagonal affinity values rii in ascending order, then the corresponding diagonal cells ˜ rii in the rank ordinal affinity matrix ˜R: The Zero-cell represents that no co-occurrence relationship for this row and column for with changing the graph we can be remain the zero-cell unchanged. Then the number of zero-cells in the i^{th} row is recorded as nulli. Other than the diagonal cell i.e., for the i^{th} row, the corresponding cells ˜rij in the ith row of ordinal affinity matrix: matrices Note that the zero-cells are not as the initial rank order. The ordinal matrix is not necessarily symmetric. The scale reflects variances in degree of affinity matrices corresponding to the affinity matrices in Fig. 4. Then, it is shown that although there are major differences between original name and face affinity Matrices, the derived ordinal affinity are basically the same.

	WIL	SPI	ANN	MAX	BEL
WIL	5	3	4	1	2
SPI	4	3	3	1	2
ANN	4	3	4	0	0
MAX	4	2	0	1	3
BEL	4	2	0	3	2

(a)

	Face1	Face2	Face3	Face4	Face5
Face1	5	3	4	1	2
Face2	4	3	3	1	2
Face3	4	3	4	0	2
Face4	4	2	0	1	3
Face5	4	2	1	3	2

(b)

Fig.5. Example of ordinal affinity matrices corresponding to figure 4: (a) Name ordinal affinity matrix ˜R name (b) Face ordinal affinity matrix ˜R face

The differences are generated due to the changes matrix is less sensitive to the noises than the original affinity matrix. We will further validate the advantages of ordinal graph representation in the experiment section

C. ECGM-based Graph Matching

ECGM is a powerful tool for graph matching with and computer vision [6]. In order to measure the similarity of a usually reflect the likelihood of according to the noise investigation and design the edit cost function to improve the performance. For explanation convenience, we provide some notations and definitions taken from [2]. Let L be a finite alphabet of labels for vertexes and edges.

Notation: A graph is a triple g = (V, α, β), where V is the finite set of vertexes, α: V → L is vertex labeling function, and β: E → L is edge labeling function. Then the set of edges E is implicitly given by assuming that graphs are fully connected, i.e., E = V × V. the alphabet.

Definition1. Let g 1 = (V1, α1, β1) and g2 = (V2, α2, β2) be two graphs. An ECGM from g1to g2 is a objective function f:ˆV1 →ˆV2, where ˆV1 ⊆ V1 andˆV2 ⊆ V2.We say that vertex x ∈ˆV1is substituted by vertex y ∈ˆV2if f (x) = y. If α1(x) =α2 (f (x)), the substitution is called an identical is usually is greater than zero.

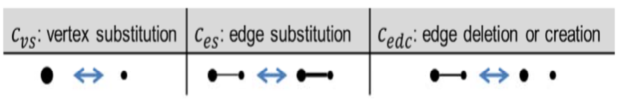

Fig. 6. Three basic graph operators for editing graphs.

Definition2. The cost of an ECGM f:ˆV1 →ˆV2 from graph g1 = (V1, α1, β1) to g2 = (V2, α2, β2) is given by

$$\gamma(f, g1, g2) = \sum_{x\epsilon v1-v^1} C_{vd}(x) + \sum_{x\epsilon v1-v^2} C_{vi}(x) + \sum_{x\epsilon^{v1}} C_{vs}(x) + \sum_{e\epsilon\varepsilon^{\wedge}1} C_{es}(e) \qquad \textbf{(1)}$$

Where cvd (x) is the cost of deleting a vertex x ∈ V1 − ˆV1 from g1, cvi(x) is the cost of inserting a vertex ∈ V2 −ˆV2in g2, cvs (x) is the cost of substituting a vertex x ∈ˆV1byf (x) ∈ˆV2, and ces (e) is the cost of substituting an edge e = (x, y) ∈ˆV1 ×ˆV1by e′= (f (x), f (y)) ∈ˆV2 ×ˆV2.

Definition3. Let f be an ECGM from g1to g2, C a cost function. We call f an optimal ECGM under C if there is no other ECGM f′ from g1to g2 with γ C (f′, g1, g2) <γ C(f, g1, g2). In our cases, if we set the number of face track clusters as the same with the number of same sub graph isomorphism in scheme 1. We have |ˆV1| = |V1| = |ˆV2| = |V2|. Also, as no vertex deletion or insertion operation is involved, we can directly assign $c_{vd} = c_{vi} = \infty$. According to the investigation on noises, we establish cedc (e) for the cost of destroying an edge e ∈ˆV1 ×ˆV1or creating an edge e ∈ˆV2 ×ˆV2. The edit operation of destroying an edge means certain cell in the name ordinal affinity matrix is nonzero while the corresponding cell in the face ordinal affinity ordinal affinity graphs matching application as: Where that we can measure the degree of vertex substitution and edge substitution, correspondingly. According to the likelihood of graph distortions during the graph of vertex it is defined as

$$C = (C_{vd}, C_{vi}, C_{vs}, C_{es}, C_{edc}) = (\infty, \infty, \lambda 1, 1, \lambda 2) \qquad \textbf{(2)}$$

Where λ1and λ2 embody the likelihood of we can be perform experiments on a training set with various values of λ1 and λ2 and select those which maximize the average matching accuracy. Recalling the operation is involved. Moreover no vertex substitution or edge substitution operations happen. There involves two edge insertion operations (edge {Face3, Face5}, {Face5, Face3}) and one edge substitution operation. The cost of this ECGM under our designed cost function C is:
$\gamma\, C(f, \tilde{R}^{face}, \tilde{R}^{name}) = 2\lambda 2 + \lambda 1$.
Consider N face clusters and character names, the number of feasible states for the solution space is the permutation of N,
i.e.,
N ! = N × (N − 1) ×· · ·× 2 × 1.
A general algorithm to obtain the optimal ECGM is based on the A ∗ method [2]. By applying A ∗, we are able to find the best matching by exploring only the most promising avenues, which guarantee a global optimal.

III. SCHEME 2: FACE- NAME GRAPH MATCHING WITHOUT NUMBER OF CLUSTER SPECIFIED

Scheme 2 requires no specification for the face cluster number. Standard affinity propagation [29] is utilized for face. Earth Mover's Distance (EMD) [30] between face tracks, which is same as introduced in [2]. All face tracks are equally suitable as exemplars and the preferences s(k, k) are set as the median of the input and "responsibility", changed between face tracks. With "availability" a(i, k) initialized to be zero, the "responsibilities" r(i, k) are updated and computed using the rule

$$r(i,k) \leftarrow s(i,k) - \max_{k', s.t. k' \neq k} \{a(i,k') + s(i,k')\} \qquad \textbf{(3)}$$

While, a (i , k) is updated using the rule

$$a(i,k) \leftarrow \min\{0, \mathrm{r}(k,k) + \textstyle\sum_{i', s.t. i' \notin i,k} \max(0, r(i',k)\} \textbf{(4)}$$

The message-passing procedure converges when the local decisions remain constant for certain number of iterations. Since no restriction is set on the one-to-one face-name to cope with the situations where several face clusters correspond to the same is conducted before a graph into disjoint sub graphs of the same size [31]. In this paper, graph partition is only used to denote the process of dividing this original face graphs.

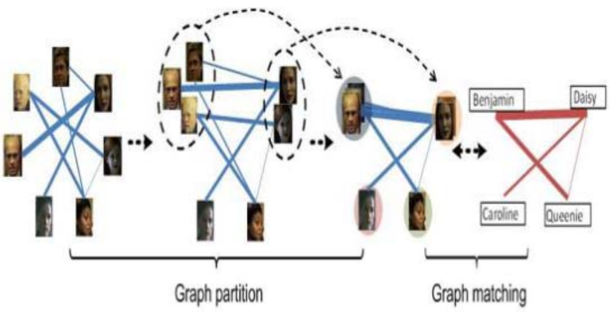

Fig.7. simultaneously graph partition and matching for scheme 2.

	Face1	Face2	Face3	Face4	Face5	Face6	Face7
Face1	0.124	0	0	0.125	0	0	0
Face2	0	0.133	0	0	0.031	0.036	0
Face3	0	0	0.024	0.011	0	0.004	0
Face4	0.125	0	0.011	0.175	0	0	0.063
Face5	0	0.031	0	0	0.081	0.002	0
Face6	0	0.036	0.004	0	0.002	0.096	0
Face7	0	0	0	0.063	0	0	0.054

(a) Original face affinity matrix R^{face}

	Partition 1	Partition 2	Partition 3	Partition 4
Partition 1	0.281	0.167	0.040	0
Partition 2	0.167	0.256	0.002	0.063
Partition 3	0.040	0.002	0.096	0
Partition 4	0	0.063	0	0.054

(b) Partitioned face affinity matrix $R^{face}(p)$

	Benja	Daisy	Queen	Carol
Benja	0.241	0.137	0.053	0
Daisy	0.137	0.276	0.001	0.061
Queen	0.053	0.001	0.099	0
Carol	0	0.061	0	0.044

(c) Name affinity matrix R^{name}

Fig.8. The example affinity matrices from the movie "The Curious Case of Benjamin Button".

When we do not set separate metrics for an optimal with the in partition and input for in a unique framework. We first define the graph partition p with respect to the original face clusters, it divides G face into N disjoint sub graphs:

$$p = \{g1^{face}, g2^{face}, \ldots\ldots, gn^{face}\} \qquad \textbf{(5)}$$

Each sub graph g_k^{face} is a sub-layer of G^{face} with vertex set v_k^{face} , and

$$U_{k=0}^{N} v_k^{face} = v^{face}\ v_i^{face} \neq \emptyset, \forall i, v_i^{face} \cap v_i^{face} = \emptyset, \forall i \neq j \qquad \textbf{(6)}$$

Where v^{face} denotes the vertex set of face graph G^{face}. In this way, the number of vertexes for each sub graph $g_k^{face}, |g_k^{face}| \in \{1,2,\ldots., M-N+1\} k = 1,2,\ldots..N$.

The vertexes in the same sub graph are considered from the same character and their co occurrence statistics are integrated. Then the partitioned face graph has the same vertex number with this name graph. Then this partitioned face affinity is given in a matrix.

When the affinity matrices are shown in Fig.8 are in from the movie Then this is the curious Case of graph by p = {(Face1, Face2, Face3), (Face4, Face5), (Face6), (Face7)} from the original face graph for this Fig.8 (a). Then the partitioned face affinity matrix is then transformed to the corresponding ordinal affinity matrix $\acute{R}^{face}$ (p) according. When the optimal solution for this graph partition and graph matching Θ∗= (p∗, f∗) under cost function C is obtained by:

$$\ominus^{*} = \arg min_{p,f}^{min} \gamma C\,(f, \tilde{_R}fcae(p), \tilde{_R}name) \qquad \textbf{(7)}$$

Consider N character names and M face track clusters, there are $P_M^N\, N^{M-N}$ possible partitions, where P_M^N = M × (M −1) · · ·× (M −N +1) denotes the N - permutations of M .Therefore, the solution space for the joint graph partition and graph matching has $P_M^N\, N^{M-N}$.N! .Possible states. A simple preprocess is used to filter the candidate partitions. While there is a very small chance that different faces of the same character appear in the same time, that the face clusters having the large affinities in the original face matrix are unlikely from the same character. That therefore, we can be add the following constraint to the graph partition:

$$if \gamma_{ij}^{face} > T_{p,}(v_i, v_j) \notin v_k . i \neq j, k = 1,2,\ldots,N. \qquad \textbf{(8)}$$

Where, Tp is the threshold to allow certain noises. We set Tp = 0.005 in this experiments. Then this filtering process will be significantly reducing the solution space. For the typical case of 20 face clusters and then10 character names, the inventive solution space has a huge O (10^{18}) possible states. After filtering, the result space is reduced to about O (10^{12}).

IV. CONCLUSION

We have shown that the proposed two schemes are useful to improve results for clustering and identification of the face tracks extracted from uncontrolled by the movies and then videos. From this the sensitivity analysis, that we have also shown that to some degree, such that these schemes are have better robustness to the noises in that the constructing affinity graphs than the traditional methods. A third conclusion is a principle for this developing robust character identification method: intensity as like noises must be emphasized more than the coverage as like noises. In the future, we will extend our work to investigate the optimal functions for the different movie genres. When another goal of future work is to exploit more than this is the character relationships, for e.g., then this is the sequential statistics for the speakers, to be build affinity graphs and then improve the robustness.

REFERENCES

[1] P. Sang, G. Liang, D. Xu, and A. Cheng,“ Robust movie character identification and the sensitivity analysis”.

[2] N. Zhang, D. Xu, T. Lu, and M. Huang, “Character identification in feature-length films using global face-name matching.”.

[3] N. Everingham, S. Sivic, and V. Zissserman,“ Taking the bite out of automated naming of characters in TV video.

[4] E. Liang, U. Xu, L. Cheng, and K. Lu, TV parser: An automatic TV video parsing method.

[5] G. Sang and M. Xu, “Character-based movie summarization”

[6] T . Hong, N. Wang, P. Xu, K. Yan, and E.- K. Chua, Dynamic captioning: