EFFICIENT GRAPH BASED SPATIAL FACE CONTEXT REPRESENTATION AND MATCHING

Yi-I Chiu[1], *Congcong Li*[2], *Chun-Rong Huang*[3], *Pau-Choo Chung*[1], *Tsuhan Chen*[4]

[1]Department of Electrical Engineering, National Cheng Kung University, Taiwan
[2]Google Inc., USA
[3]Department of Computer Science and Engineering, National Chung Hsing University, Taiwan
[4]Department of Electrical and Computer Engineering, Cornell University, USA

ABSTRACT

In this paper, we propose a novel orientation-aware Urquhart graph based spatial face context representation method to efficiently describe the spatial relationship among faces in group photos. We combine graph matching with orientations of graph edges to assess the similarity of spatial face contexts from different group photos. The experimental results show that our method can find more structurally similar group photos compared to the state-of-the-art spatial face context representation methods.

Index Terms— spatial face context, graph matching, social event analysis.

1. INTRODUCTION

Group photos record daily events and moments for reminiscence. The positions of people in group photos naturally rely on the relationship of the attendee. For example, a group photo shown in Fig. 1(a) illustrates a souvenir photo of a family reunion. Elders are at the central focal point, children sit in front of them, and grown-ups stand behind them. In contrast, colleagues and classmates often stand orderly in rows as shown in Fig. 1(b) and (c), respectively, for souvenir photos. Fig 1(d) shows a group photo captured during dining scene. Unlike previous group photos, people are surrounding the table. Thus, relative face positions provide meaningful spatial face contexts among different social events and reflect the semantic facts in group photos.

Recently, the spatial face context has been shown its effectiveness for pairwise social relationship analysis [1][2], demographical estimation of individuals [3][4], and photo organization [5][6]. In most approaches, spatial face contexts can be represented by either pairwise relationship [1][2][4][5][6], and nearest neighbors [3]. These works take advantage of *local* spatial face context to improve the experimental results.

This work was supported in part by the National Science Council of Taiwan, R.O.C., under NSC-101-2221-E-005-086-MY3 and NSC-101-2221-E-006-262-MY3.

This work was done when Congcong Li was with Cornell University.

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

Fig. 1. (a) A family photo. (b) A colleague photo. (c) A classmate photo. (d) A dining photo.

However, because of lacking global distributions of face positions, semantic understanding of relations of people in group photos is hard to be achieved.

When seeing a group photo, we observe not only the information of individuals but also their group activity. Thus, if the *global* spatial face context can be extracted, the circumstantial settings and the group activity of the photography can be represented.

To analyze the global spatial face context, Gallagher et al. [7] utilize iterative graph cuts with pre-training edges to detect rows in a group photo. However, as shown in Fig. 1, the spatial face context will change according to different social events. Modeling faces in rows remains insufficient to represent global spatial face context. To solve the above-mentioned problem, minimum spanning tree (*MST*) is considered to model the spatial face context for the evaluation of centrality in [3]. Then, the spatial face context is used as an evidence for dining event understanding. Chen et al. [8] model the spatial face context as a complete graph. Informa-

tive face subgraphs are retrieved to indicate different social relationships for bag of subgraphs training. These approaches indicate that using relative positions of faces provides significant improvement for semantic analysis compared to using absolute positions and distances of faces. However, they are still limited to considering spatial face contexts from two faces and cannot be applied for the comparison of spatial face contexts from different group photos. Moreover, the scaling problem [8] during comparing graphs in different photos remains unsolved.

To model the global spatial face context, we propose using the Urquhart graph (*UG*) [9] which can efficiently represent relative face positions. A graph matching procedure [10] is employed to estimate the similarity between two *UG*s of two group photos. Because *UG*s capture the human perceptions of the shape [11][12], graphs with similar structures can be matched under different numbers of vertices. The orientation differences between matched edges are also considered when evaluating the similarity between two graphs to further increase the matching accuracy.

2. SPATIAL FACE CONTEXT REPRESENTATION

As indicated by [3][7][8], relative positions of faces in group photos contain the relationship of the attendee. Thus, one can model the spatial face context as the connections of each face pair. Then, the spatial face context among N faces in a group photo I can be represented as a complete graph $G = \{V, E\}$, where $V = \{v_1, \ldots, v_N\}$ and $E = \{e_{ij} | 1 \leq i, j \leq N, i \neq j\}$. Each vertex v_n in V represents a face in the photo, where the position of v_n is denoted as $\mathbf{v}_n$. Each edge e_{ij} in E is the spatial relation between v_i and v_j. The complete graph directly links two faraway faces to represent the face relationship. Such connection is redundant if the relationship between two faces can be connected by a path through other faces. Edges between spatially nearby vertices are sufficient to represent spatial relations among faces, i.e. people who are in neighboring positions can represent spatial relations via graph edges. In this paper, we propose using a relative neighborhood graph (*RNG*) to replace G to reduce the complexity of modeling the neighboring face positions. As shown in [11][12], *RNG* provides human perceptions of the shape of G. It is then a cognitive representation to describe relative face positions in group photos.

A relative neighborhood graph G^R of G can be defined as $G^R = \{V, E^R\}$. G^R connects two face vertices v_i and v_j by an edge whenever there does not exist a third point v_k which is closer to both v_i and v_j than they are to each other. An edge e_{ij}^R belongs to E^R is defined as follows:

$$\{e_{ij}^R = 1 | \|\mathbf{v}_i - \mathbf{v}_j\| < \max\{\|\mathbf{v}_i - \mathbf{v}_k\|, \|\mathbf{v}_j - \mathbf{v}_k\|\}, \forall v_k \in V, v_k \neq v_i, v_k \neq v_j\}, \quad (1)$$

where $\|v_i - v_j\|$ is the 2-D image distance between v_i and v_j. As a result, *RNG* indicates that two faces do not have directed relationship to each other if their relation is intervened by another vertex.

In our approach, we use Urquhart graph (*UG*) [9] to approximate *RNG* for its computational efficiency. Although *UG* is a supergraph of *RNG*, it serves nearly equally well for computational morphology comparisons. *UG* is obtained by removing the longest edge from each triangle in Delaunay triangulation (*DT*) of V and its complexity is O($N \log N$). Detailed comparison between *RNG* and *UG* can be found in [11]. Given a $DT = \{T_1, \ldots, T_C\}$ of V where T_c is a triangle containing three vertices (v_i, v_j, v_k). Then, the Urquhart graph $U = \{V, E^U\}$, and the edge e_{ij}^U belongs to *UG* are defined as follows:

$$\{e_{ij}^U = 1 | \|\mathbf{v}_i - \mathbf{v}_j\| < \max\{\|\mathbf{v}_i - \mathbf{v}_k\|, \|\mathbf{v}_j - \mathbf{v}_k\|\} \& \{v_i, v_j, v_k\} \in T_c\}. \quad (2)$$

As a result, faces appearing in a group photo are modeled as vertices of *UG*. Each face (vertex) is correlated to its neighbor faces (vertices) and linked by e_{ij}^U. Such representation provides an efficient way to describe the spatial face context.

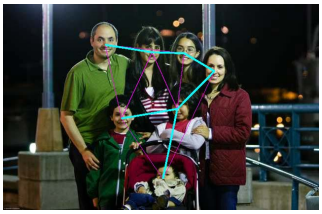
3. SPATIAL FACE CONTEXT MATCHING

Given two group photos I^x and I^y, two *UG*s $U^x = \{V^x, E^x\}$ and $U^y = \{V^y, E^y\}$ are retrieved to represent their spatial face contexts. To assess the similarity between U^x and U^y, a graph matching based measurement is proposed. Graph matching using the path following algorithm [10] has been shown the effectiveness for matching the correspondence between vertices of two graphs. A permutation matrix P is defined where the element P_{ij} of P equals to 1 if the i-th vertex of U^x is matched to the j-th vertex of U^y. Otherwise, P_{ij} equals to 0. With the permutation matrix P, U^y is transformed to its isomorphic, which is denoted by $P(y)$. After applying P to U^y, the edge matrix $E^{P(y)}$ of the permuted graph is obtained from E^y as $E^{P(y)} = PE^yP^T$. If P can be found, U^x and U^y are matched, i.e. they have similar spatial face context. The measurement $F_0(P)$ of P between U^x and U^y after matching is defined as follows:

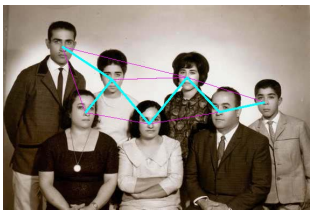
$$F_0(P) = \|E^x - E^{P(y)}\|_F^2 = \|E^x - PE^yP^T\|_F^2, \quad (3)$$

where $\|.\|_F$ is the Frobenius matrix norm defined by $\|E\|_F^2 = tr(EE^T)$ which is the trace of EE^T. If $F_0(P)$ is large, U^x and U^y are dissimilar with P. As a result, the problem of graph matching between U^x and U^y can be formulated as the problem of minimizing $F_0(P)$ with respect to the permutation matrices.

In practice, the numbers of vertices N^x and N^y of U^x and U^y will be different, i.e. the numbers of faces in I^x and I^y are different. To match graphs of different sizes, dummy isolated vertices are added to the smaller graph [10], which means zeros rows to the vertices set, and zero rows and zero



(a)



(b)

Fig. 2. (a) and (b) show two photos with similar *UG*s, but the spatial face locations are dissimilar. Red edges are created by Delaunay Triangulation and blue edges belong to the UG.

columns to the edge set. Assume that $N^x > N^y$, then P will become a square $N^x \times N^x$ matrix.

As shown in [10], this graph matching problem can be generalized to the problem of labeled graph matching which fits graph labels and graph structures at the same time. Let C_{ij} denote the cost of fitness between the i-th vertex of U^x and the j-th vertex of U^y. The label comparison between U^x and U^y under the permutation matrix P is formulated as follows:

$$\min_{P_{ij} \in P} \text{tr}(C^T P) = \sum_{i=1}^{N^x} \sum_{i=1}^{N^x} C_{ij} P_{ij}. \tag{4}$$

To consider both of the graph structure and the labels of vertices, a convex combination can be formulated as follows:

$$\min_{P_{ij} \in P} (1-\alpha) F_0(P) + \alpha \text{tr}(C^T P), \tag{5}$$

where $\alpha \in [0, 1]$. A small α value favors to assess the similarity of graph structures and a large α favors to assess the similarity of labels of vertices. For detailed optimization algorithms and implementation, please refer to [10].

In practice, matching graphs of the spatial face context is not exactly the same as matching general graphs. As shown in Fig. 2, these two group photos have similar *UG*s, but the relative face positions are different. Thus, not only the vertices, but also the orientations of corresponding edges should be matched. After finding correspondence between V^x and V^y using (5), the orientations of edges are used to further assess the similarity between U^x and U^y. Given an edge e^y_{ij} in E_y, we use the permutation matrix P to transform the end vertices (v^y_i, v^y_j) of e^y_{ij} to a new graph edge $e^{x'}_{ij}$, which is corresponding to e^x_{ij} in U^x. Because U^y may not be the exact isomorphism of U^x, $e^{x'}_{ij}$ is not always equal to e^x_{ij}. Here, $e^{x'}_{ij}$ is defined by the transformed vertices as follows:

$$e^{x'}_{ij} = \{v^{x'}_i, v^{x'}_j\}, \tag{6}$$

where $v^x_i = P^T v^y_i$. If U^y has similar spatial face context compared to U^x, the orientation of $e^{x'}_{ij}$ should be similar to the orientation of exij. Thus, the similarity $O(e^x_{ij}, e^{x'}_{ij})$ of orientations between e^x_{ij} and $e^{x'}_{ij}$ is defined as follows:

$$O(e^x_{ij}, e^{x'}_{ij}) = \| \frac{\mathbf{v}^x_j - \mathbf{v}^x_i}{\|\mathbf{v}^x_j - \mathbf{v}^x_i\|} - \frac{P^T\mathbf{v}^y_j - P^T\mathbf{v}^y_i}{\|P^T\mathbf{v}^y_j - P^T\mathbf{v}^y_i\|} \|. \tag{7}$$

The characteristics of $O(e^x_{ij}, e^{x'}_{ij})$ can be summarized as follows:

$$\begin{cases} O(e^x_{ij}, e^{x'}_{ij}) > 1, \text{if ang}(e^x_{ij}, e^{x'}_{ij}) > 90^\circ \\ O(e^x_{ij}, e^{x'}_{ij}) = 1, \text{if ang}(e^x_{ij}, e^{x'}_{ij}) = 90^\circ \text{ or } e^x_{ij} = 0 \text{ or } e^{x'}_{ij} = 0 \\ O(e^x_{ij}, e^{x'}_{ij}) < 1, \text{if ang}(e^x_{ij}, e^{x'}_{ij}) < 90^\circ \end{cases} \tag{8}$$

where ang$(,)$ is the angle between two edges. The similarity O of orientations between U^x and $U^{x'}$ is then defined as the summation of all correspondent edges as follows:

$$O = \sum_{ij} O(e^x_{ij}, e^{x'}_{ij}). \tag{9}$$

If O is small, the spatial face contexts of U^x and U^y are similar, i.e. I^x and I^y have similar relative face positions.

4. EXPERIMENTS

We perform our experiments on the Family and Group dataset [3]. Each face is treated as a vertex for generating *DT*. Then, *UG* is obtained from *DT* results by removing the longest edges. For comparison, *MST* based spatial face context representation [3] is implemented. Given a randomly selected group photo I^x as the reference image, we aim to find another group photo I^y, which has the most similar spatial face context as I^x, to evaluate the correctness of the spatial face context representation. Fig. 3 (a) shows 5 reference images. Fig. 3 (b) and (c) show *MST*s of reference images and the retrieved images which contains the most similar spatial face contexts under *MST*. Fig. 3 (d) and (e) show *UG*s of reference images and the retrieved image which contains the most similar spatial face context under *UG*. The dark blue dots represent the faces which are the vertices of the graph. Light blue lines represent the linked edges of *MST*. Yellow lines represent the linked edges of *UG*.

I_1 in Fig. 3 shows a group photo in a restaurant. People sit surrounding the table and face to the camera to take the souvenir photo. In I_1, *MST* and *UG* have the same structure because *MST* is a subgraph of *UG*. Another dining photo with similar spatial face context can be retrieved for both graphs. From I_2 to I_5, the *MST* and *UG*s of face structures are different because *UG*s include edges between perceptually nearby vertices and then have cycles, which can represent the closer relations of neighbor faces. I_2 demonstrates a family photo in two rows. *MST* in I_2 matches a row of people with larger height variation, while *UG* matches another family photo in two rows. The same situation can be found in I_3, a larger

Fig. 3. The comparisons between *MST* and *UG*. I_1 to I_5 (first row to fifth row) represent five randomly selected images from the dataset [3]. (a) The original images. (b) The *MST* of the spatial face context. (c) The image with the most similar spatial face context using *MST*. (d) The *UG* of the spatial face context. (e) The image with the most similar spatial face context using *UG*.

group photo with two main rows. The perceptually informative edges added by *UG* contribute greatly for matching photos of rows of people. I_4 and I_5 show people stagger closely. Because of lacking the information about the original structures, *MST* tends to match a line with branches to these complicated photos. *UG*, on the other hand, finds better matches since it includes not only the closest vertices (faces) but also all the approximate vertex pairs to construct the graph structure which captures human perceptions of the spatial distributions.

The experimental results show that the proposed approach successfully retrieve photos with similar spatial face contexts. In addition, our method is implemented using Matlab on an Intel i7 computer with a 3.4G CPU and the average computation time for computing *UG* and matching two *UG*s is 0.04 and 0.05 seconds, respectively.

5. CONCLUSION

We propose an Urquhart graph based spatial face context representation to describe the spatial relationship of faces in a photo. To evaluate the similarity between the face contexts of two group photos, an orientation-aware graph matching method is proposed. Such representation and matching can be useful for applications such as social relationship analysis and photo organization.

6. REFERENCES

[1] P. Singla, H. Kautz, J. Luo, and A. Gallagher, "Discovery of social relationships in consumer photo collections using markov logic," in *Proc. of IEEE Conference of Comptuer Vision and Pattern Recognition Workshops (CVPRW)*, 2008.

[2] S. Xia, M. Shao, J. Luo, and Y. Fu, "Understanding kin relationships in a photo," *IEEE Transactions on Multimedia*, vol. 14, no. 4, pp. 1046–1056, 2012.

[3] A. Gallagher and T. Chen, "Understanding images of groups of people," in *Proc. of IEEE Conference of Comptuer Vision and Pattern Recognition (CVPR)*, 2009.

[4] G. Wang, A. Gallagher, J. Luo, and D. Forsyth, "Seeing people in social context: recognizing people and social relationships," in *Proc. of European conference on Computer vision (ECCV)*, 2010, pp. 169–182.

[5] P. Wu, W. Ding, Z. Mao, and D. Tretter, "Close & closer: discover social relationship from photo collections," in *Proc.*

of IEEE international conference on Multimedia and Expo (ICME), 2009, pp. 1652–1655.

[6] T. Zhang, H. Chao, C. Willis, and D. Tretter, "Consumer image retrieval by estimating relation tree from family photo collections," in *Proc. of International Conference on Image and Video Retrieval (CIVR)*, 2010, pp. 143–150.

[7] A. C. Gallagher and T. Chen, "Finding rows of people in group images," in *Proc. of the 2009 IEEE International Conference on Multimedia and Expo (ICME)*, 2009, pp. 602–605.

[8] Y.-Y. Chen, W. H. Hsu, and H.-Y. M. Liao, "Discovering informative social subgraphs and predicting pairwise relationships from group photos," in *Proc. of the 20th ACM international conference on Multimedia*, 2012, pp. 669–678.

[9] R. B. Urquhart, "Algorithms for computation of relative neighborhood graph," *Electronics Letters*, vol. 16, no. 14, pp. 556–557, 1980.

[10] M. Zaslavskiy, F. Bach, and J.-P. Vert, "A path following algorithm for the graph matching problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2227–2242, Dec. 2009.

[11] D. V. Andrade and L. H. de Figueiredo, "Good approximations for the relative neighbourhood graph," in *Canadian Conference on Computational Geometry (CCCG)*, 2001, pp. 25–28.

[12] J. W. Jaromczyk and G. T. Toussaint, "Relative neighborhood graphs and their relatives," in *Proc of the IEEE*, 1992, pp. 1502–1517.