MatchDR: Image Correspondence by Leveraging Distance Ratio Constraint

Rui Wang, Dong Liang, Wei Zhang*, Xiaochun Cao
State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
wangrui@iie.ac.cn, dongliang.ucas@gmail.com, wzhang.cu@gmail.com, caoxiaochun@iie.ac.cn

ABSTRACT

Image correspondence is to establish the connections between coherent images, which can be quite challenging due to the visual and geometric deformations. This paper proposes a robust image correspondence technique from the perspective of spatial regularity. Specifically, the visual deformation is addressed by introducing the spatial information by enforcing the distance ratio constrain. At the same time, the geometric deformation is tolerated by adopting a smoothness term. Subsequently, image correspondence is formulated as permutation problem, for which, we propose a Gradient Guided Simulated Annealing method for robust optimization. Furthermore, our method is much more memory efficient, where the storage complexity is reduced from $O(n^4)$ to $O(n^2)$. The experiments on several datasets indicate that our proposed formulation and optimization significantly improve the baselines for both visually-similar and semantically-similar images, where both visual and geometric deformations are present.

Keywords

Image Correspondence; Spatial Regularity; Distance Ratios; Discrete Optimization

1. INTRODUCTION

Image correspondence is to establish the connections between similar points across different images. Due to its wide applications in many other high-level visual problems, image correspondence has been extensively studied in the past decade as rigid scene matching, image registration / alignment, optical / SIFT flow, and dense correspondences. Traditional correspondence problems mainly focus on images that are spatially overlap or temporally adjacent, where images are closely related by similar visual appearances, e.g., the stereo correspondence problem.

Image correspondence is challenging as pictures taken by different persons for different purposes may look differently, the image pairs can be semantically similar images, images with blur / quality changes, images with occlusions, images taken from different views of the same 3D object, to name a few. For example, two cars with different models can be visually dissimilar, and thus difficult to relate. However, matching images with diverse appearances is a critical technique for various applications.

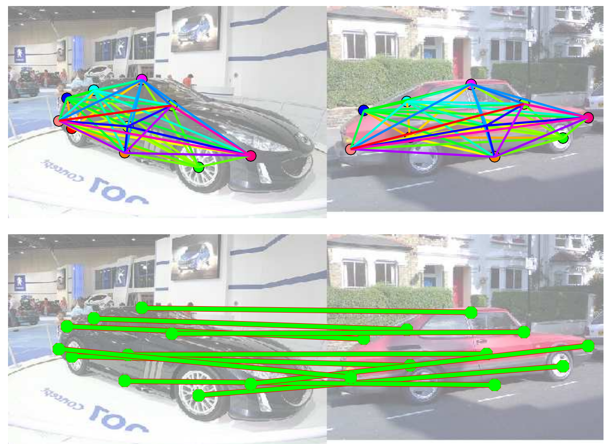

Figure 1: The distance ratio gives clues for the spatial regularity in image correspondence. The top image pair plots the distances between all landmark points. The lines with the same color indicate corresponding edges. Note that distances are more stable than angles for complex cases such as semantically similar cars. The bottom pair shows the result by our method MatchDR by enforcing the distance ratio constraint.

The challenges of this problem can be summarized into two types of deformations. The first one is the visual deformation. The image pairs could exhibit diverse visual appearances in different background, colors or models, as shown in Fig 1. Matching based on visual similarities can be very difficult. The second one comes from the geometric deformation. The same object could show different sizes, shapes and poses, where the Planar Homography does not hold anymore. For example, cars viewed from different viewpoints will also deform in shape.

The problem is traditionally formulated as a quadratic assignment problem (QAP), which is known to be NP-hard, and it incorporate both *node-to-node* and *edge-to-edge* similarity as structural information in a bulky way. The memory cost is huge for coping with large images. This paper proposes a new formulation which requires much less memory, and a geometric based method addressing the aforementioned challenges. We define the problem using the Distance Ratio (DR) matrix. And the proposed method is named “Match via Distance Ratio" (MatchDR). Based on the

*Corresponding author

MM '16, October 15-19, 2016, Amsterdam, Netherlands
© 2016 ACM. ISBN 978-1-4503-3603-1/16/10...$15.00
DOI: http://dx.doi.org/10.1145/2964284.2967293

observation that the locations of the corresponding points on pairwise images should be structurally similar, we use the constant that the distance ratios between correspondences should be a constant. Specifically, the objective is to find the permutation matrix to minimize the variance of the distance ratios. Note that we discard the absolute locations and only use the relative distances between points, This kind of information is actually strong enough to recover the original locations of points as stated in [25].

On one hand, we introduce the spatial regularity to the correspondence problem by enforcing the DR matrix, which improves the quality of visual correspondence where elements are with small variance. On the other hand, by introducing a smoothness term, the geometric deformation is modeled by tolerating local spatial jitters. Compared to other methods using spatial information, our method is much more memory efficient, due to the DR matrix encoding all the weight information is of $n \times n$, rather than $n^2 \times n^2$, where n is the number of landmarks on each image.

Our contributions can be summarized as follows: First, we propose an effective geometric method by enforcing consistent distance rations. The problem can be modeled as a permutation problem. Second, we achieve the state-of-the-art performance on both visually similar and semantically similar datasets.Third, we proposed a Gradient Guided Simulation Annealing method for robust optimization of the proposed method. The discovery of new solutions is directed by the gradient, making the optimization more accurate and effective.

2. RELATED WORKS

Image correspondence is one of the fundamental problems in computer vision and multimedia, which has been widely studied through the past decades [8]. According to the information exacted from the images, previous literatures can be divided into visual based and geometric based methods.

2.1 Visual-based methods

Visual-based correspondence is a reliable method for image correspondence and has already been used in many applications. A tremendous number of research efforts have been conducted in various forms. In computer vision, early correspondence method is often called the stereo correspondences [20, 23]. There are two different ways dealing with the problem in the literature. Many algorithms on image correspondence are based on comparing the differences of the appearance of the landmark points, leading to area-based matching (ABM) [2, 16] or feature-based matching (F-BM) [7, 9, 12], which are classical and well-developed.

Many works have been done since then and there are quite a lot new techniques for detecting landmarks and generating local features [17]. For example, the invention of SIFT for detecting and describing the local features of the landmarks in the images. Visual-based methods of both pairwise matching [19, 24] and multi-matching [14, 26] are based on the feature detection. Recently, the Convolutional Neural Network is used to detect and describe features and gets a better result. Furthermore, [3] trains the LDA classifier for every pixel to exclude the false matches. [27] trains a convolutional neural network to predict how well two image patches match. [30] introduces the universe to simplify the calculation.

2.2 Geometric-based methods

Though visual based methods make up a large proportion in the literature, the geometric-based methods are widely found in various computer vision problems, e.g., image retrieval [10, 28, 29]. In the literature, the spatial information is used as a kind of feature the same as visual feature ponderously [1]. Spatial rigidity is usually preferred, i.e., selecting the correct corresponding pair of edges by QAP, which is well known as a graph matching problem. While QAP is NP-hard, many efficient algorithms have been proposed to solve it approximately. The earliest approach for QAP is the spectral methods [11]. [6] introduces a random walk view on the problem, and the result is relatively noise-robust. Another weakness for QAP is that the dimension of the affinity matrix equals the number of the possible point pairs. Usually, the landmark of an image is several hundreds, making the affinity matrix very large.

3. METHODOLOGY

3.1 Problem formulation

The image correspondence problem is defined on two related images. For each image $k \in \{1, 2\}$, we define an undirected graph as $\mathcal{G}^k = (\mathcal{V}^k, \mathcal{E}^k, \mathcal{W}^k)$, where $\mathcal{V}^k, \mathcal{E}^k, \mathcal{W}^k$ denote the nodes, edge, and attributes of edge respectively.

The pairwise match problem between a graph pair $\mathcal{G}^1$ and $\mathcal{G}^2$ is to find a one-to-one correspondence τ between $\mathcal{V}^1$ and $\mathcal{V}^2$ which minimizes some costs between them. p_k is the number of nodes in the graph $\mathcal{G}^k$. We represent the correspondence τ by a partial permutation matrix $\mathbf{X} \in \{0, 1\}^{p_1 \times p_2}$. "One-to-one" constraint is assumed, that is, one node in graph $\mathcal{G}^1$ can match at most one node in graph $\mathcal{G}^2$ and vice versa. The assumption can be formulated as the doubly stochastic constraints:

$$0 \leq \mathbf{X}\mathbf{1} \leq 1, \ \ 0 \leq \mathbf{X}^\top \mathbf{1} \leq 1. \tag{1}$$

3.2 Approach overview and algorithm details

The geometric information of images can be encoded as a graph. For each image k, we extract p_k features, and let $\mathcal{V}^k$ denote the set of feature points. The coordinate of the centre point of the ith($i \leq p_k$) feature in image k is (x_i^k, y_i^k). $\mathcal{E}^k$ denotes each pair of nodes. As we want to extract the distance information in every images, $\mathcal{W}^k(v_i^k, v_j^k)$ is regarded as the Euclidian distance on image k between two nodes i and j, i.e. $\mathcal{W}^k(v_i^k, v_j^k) = \sqrt{(x_i^k - x_j^k)^2 + (y_i^k - y_j^k)^2}$.

The adjacency matrix $D_k \in \mathbb{R}^{p_k \times p_k}$ encodes the geometric information of graph $\mathcal{G}^k$,

$$\mathbf{D}_k = [d_{i,j}^k] = \begin{cases} d_{i,j}^k = \mathcal{W}^k(\mathcal{V}_i^k, \mathcal{V}_j^k) & i \neq j; \\ d_{i,j}^k = \frac{1}{p_k} \sum\limits_{l=1, l \neq i}^{p_k} \mathcal{W}^k(\mathcal{V}_i^k, \mathcal{V}_l^k) & i = j. \end{cases} \tag{2}$$

To avoid zeros in $\mathbf{D}_k$, we replace the zeros in the original affinity matrix with the mean distance.

For image correspondence, the main problem is to find the correct correspondence between two graphs $\mathcal{G}^k$. As shown in Fig. 1, for the graphs with the same spatial structures, the ratio of distances should be roughly consistent, i.e.

$$\frac{d_{1,2}^1}{d_{1,2}^2} \approx \frac{d_{1,3}^1}{d_{1,3}^2} \approx \cdots \approx \frac{d_{i,j}^1}{d_{i,j}^2} \approx \cdots \approx \frac{d_{p_1,p_1}^1}{d_{p_2,p_2}^2} \approx \text{C}. \tag{3}$$

To represent the ratios efficiently, we fill the ratios in a matrix:

$$\mathbf{Q} = \mathbf{X}^\top \mathbf{D}_1 \mathbf{X} \circ \mathbf{D}_2^{\circ -1}, \tag{4}$$

where $\circ$ denotes the Hadamard product. The matrix $\mathbf{Q}$ is called the distance ratio matrix.

Ideally, the elements in $\mathbf{Q}$ should be the same for a pair of graphs. So it is clear that the variance of all elements in $\mathbf{Q}$ should be minimized:

$$J_1(\mathbf{X}) = \text{tr}((\mathbf{Q} - E(\mathbf{Q}))^\top (\mathbf{Q} - E(\mathbf{Q}))) \tag{5}$$

To compute the variance, $E(\mathbf{Q})$ is a matrix with the mean of all elements in $\mathbf{Q}$. Particularly, all the elements in $E(\mathbf{Q})$ equal to $\frac{\langle \mathbf{1},\mathbf{Q}\rangle}{\langle \mathbf{1},\mathbf{X}\rangle^2}$, where $\langle\cdot,\cdot\rangle$ denotes inner product of matrices.

Moreover, the estimated $\mathbf{X}$ should be sparse since at most one value in each row of $\mathbf{X}$ can be nonzero. To introduce sparsity, we minimize the summation of values in $\mathbf{X}$. The problem can be formulated as:

$$\begin{aligned} \min_{\mathbf{X}} J(\mathbf{X}) &= J_1(\mathbf{X}) + J_2(\mathbf{X}) \\ s.t. \mathbf{X} &\in \mathcal{C} \end{aligned} \tag{6}$$

where $J_2(\mathbf{X}) = \alpha\langle \mathbf{1},\mathbf{X}\rangle$, α is the trade-off weight for sparsity, $\mathcal{C}$ denotes the set of matrices satisfying the constraints in Eqn. (1).

The affinity matrix defined here is much smaller comparing with the the formulation of QAP. The storage complexity of the affinity matrix of the proposed formula is reduced from $O(n^4)$ to $O(n^2)$.

3.3 Optimization

3.3.1 Gradient Guided Simulated Annealing (GGSA)

The solution space of the problem is discrete (see Eqn. (1)) and the problem is NP-hard. Usually, such problem is relaxed to real space by treating $\mathbf{X}$ as a real matrix. But the methods in real space e.g. Gradient Descent is slow to converge, and always stucks to inferior local minimum solution in our case. In this paper, we propose a Gradient Guided Simulated Annealing algorithm to find solutions in large discrete space. Basically, we follow the framework of the Simulated Annealing algorithm [4] and improve neighbor generation using the gradient information.

To optimize Eqn. (6), we first compute the gradient as follow:

$$\begin{aligned} \frac{dJ}{d\mathbf{X}} = &\frac{4}{\langle \mathbf{1},\mathbf{X}\rangle^2}(\mathbf{D}_1\mathbf{X})(\mathbf{X}^\top\mathbf{D}_1\mathbf{X})\circ\mathbf{D}_2^{\circ-2} \\ &+ 4\frac{\langle \mathbf{1},\mathbf{Q}\rangle}{\langle \mathbf{1},\mathbf{X}\rangle^6}(p^2 - 2\langle \mathbf{1},\mathbf{X}\rangle^2)\mathbf{D}_1\mathbf{X}\circ\mathbf{D}_2^{\circ-1} \\ &+ 2\frac{\langle \mathbf{1},\mathbf{Q}\rangle^2}{\langle \mathbf{1},\mathbf{X}\rangle^7}(4\langle \mathbf{1},\mathbf{X}\rangle^2 - 3p^2)\mathbf{X} - 2\frac{\langle \mathbf{1},\mathbf{Q}^{\circ 2}\rangle}{\langle \mathbf{1},\mathbf{X}\rangle^3}\mathbf{X} + \alpha\mathbf{X} \end{aligned} \tag{7}$$

Although the problem can not be trivially optimized with gradient descent, $\frac{dJ}{d\mathbf{X}}$ gives clues on generating neighbors:

- According to the gradient, the minimum element of the gradient matrix infers that the corresponding element of $\mathbf{X}$ is most likely to be 1. To keep $\mathbf{X}$ a valid permutation matrix, a 0 changes to 1 at the same time. This new permutation is included in the neighborhood set $\mathcal{X}$.

- Meanwhile, the maximum element of the gradient matrix infers $1 \rightarrow 0$ in the expected solution $\mathbf{X}$. While the corresponding element change $0 \rightarrow 1$ has $\max(p_1, p_2)$ possible choices. We collect them in the neighborhood set $\mathcal{X}$.

The $\mathbf{X} \in \mathcal{X}$ minimizing the Eqn. (6) is an approach to get the neighbor of the simulated annealing algorithm.

Another approach for the neighbor is random permutation. As the non-convexity of the problem, the gradient does not always point to the right direction. To avoid being caught in a local minimum, we randomly generate a permutation matrix as the neighbor with a probability ρ. Upon our experimental observation, we set $\rho = 0.2$ in the experiment. Because, in most conditions, the gradient guided neighbor is trustable. In our paper, the Gradient Guided neighbor finding method is denoted $GGN(\mathbf{D}_1, \mathbf{D}_2, \mathbf{X})$.

3.3.2 Initialization

Many algorithms have been proposed based on visual matching or spatial matching. It is hard to match two images with a random initialization due to several reasons. First, the optimization of Eqn. (6) is discrete and NP-hard. The solution-space is essentially $O(n!)$ where n is the number of nodes in the graph. Second, even the relaxed version (discrete to continuous space) of Eqn.(6) is non-convex, such that the initialization impacts the result prominently. To solve the problem, it is necessary to make full use of both visual and spatial information jointly. Therefore in practice, we mostly initialize our algorithm with the result of another algorithm to ease the optimization difficulty. In the experiments, we observe significant improvements over a range of state-of-the-arts based on this strategy.

3.3.3 Smoothness term

As the existence of very short edges, the error can be magnified even with very small geometric deformation, which makes the optimization difficult. In such case, the gradient becomes unstable due to the division by a small (e.g., close to zero) value in Eqn. (4). The cost of $J(\mathbf{X})$ will change a lot when the denominator slightly changes. To avoid these errors, we add a smoothness term λ for both sides of the fraction in the distance ratio matrix $\mathbf{Q}$. That equals to adding λ to each element of the affinity matrices $\mathbf{D}_1$ and $\mathbf{D}_2$. Moreover, the smoothness term also tolerates some geometric jitters for the correspondences.

Above all, the proposed algorithm can be written as Algorithm. 1, where $P(E(\mathbf{X}), E(\mathbf{X}^{new}), t) = \exp((J(\mathbf{X}) - J(\mathbf{X}^{new}))/t)$.

Algorithm 1 GGSA: Gradient Guided Simulated Annealing

```
Input: D_1, D_2, X^0: initialization, T : maximum iterations.
Output: Solution X
X ← X^0;
for t = T ··· 0 do
    //generate a neighbor based on the gradient in Eqn. (7).
    X^new ← GGN(D_1, D_2, X);
    if P(E(X), E(X^new), t) ≥ rand(0, 1) then
        X ← X^new; //update solution
        if converge then
            break;
        end if
    end if
end for
```

4. EXPERIMENTAL EVALUATION

4.1 Datasets and Evaluation Metric

In the real-data experiment we apply our method on image correspondence problems using local feature detectors. The dataset used is listed as bellow.

The Building dataset [18] contains images taken from the the frame of a sequence, where the view of a building gradually changes. The image pairs are visually similar for temporally adjacent frames. The landmark are points marked by [15][1].

Willow Object class dataset [5] It includes images of five object classes. The appearance of images varies significantly, since the instances are only loosely related to the same semantic object, e.g., different models of car. We use SIFT [13] to extract interesting points and represent the node and edge attributes.

[1] http://pages.cs.wisc.edu/~pachauri/perm-sync/

Table 1: Performance Comparison with Other Methods in Error-Rates(%)

		K-M [15]	MatchALS [30]	SM [11]	RRWM [6]	SMCM [22]	SMCP [21]
The Building	w/o, MatchDR	81.44	79.72	85.47	78.83	86.02	76.36
	MatchDR	**30.49**	74.95	47.52	39.30	47.33	35.78
Willow Object	w/o, MatchDR	88.75	91.63	80.17	68.02	81.08	67.55
	MatchDR	47.70	41.21	39.48	35.01	45.26	**32.37**

(a) Kuhn-Munkres(0.89) (b) Kuhn-Munkres(0.95)
(c) MatchALS(0.89) (d) MatchALS(0.89)
(e) SM(0.89) (f) SM(0.95)
(g) RRWM(0.75) (h) RRWM(0.89)
(i) SMCM(0.67) (j) SMCM(0.95)
(k) SMCP(0.75) (l) SMCP(0.89)
(m) rnd-init MatchDR(0.0) (n) rnd-init MatchDR(0.0)

Figure 2: Example results motocycle and car in the Willow. Our approach is randomly initialized. The numbers in parentheses are the corresponding Error-Rates(%).

We adopt the evaluation metric used in [30]. Given a solution $\mathbf{X}^{alg}$ and the ground-truth $\mathbf{X}^{gt}$: We measure the error rate by intersection over union:

$$\text{Error-Rate} = 1 - \frac{|\tau(\mathbf{X}^{alg}) \cap \tau(\mathbf{X}^{gt})|}{|\tau(\mathbf{X}^{alg}) \cup \tau(\mathbf{X}^{gt})|}, \qquad (8)$$

where τ denotes the correspondence defined by the permutation matrix $\mathbf{X}^{alg}$ or $\mathbf{X}^{gt}$, and $|\cdot|$ is the cardinality of the set.

4.2 Performance Comparison

The methods compared are released, and the codes used in the comparison are from the corresponding websites. Table. 1 shows the error-rates of the image correspondence in both datasets, with both visual based methods (MatchALS [30][2], Kuhn-Munkres algorithm [15][1]) and geometric based methods (SM [11][3], RRWM [6][4], SMCM [22][5], SMCP [21])[6] in the first row, while the error-rates of our method (MatchDR) based on the corresponding initialization are shown in the second row.

The Building dataset is designed for homographic image correspondence. The local feature of images provides a good initialization, and Kuhn-Munkres with MatchDR achieves the best performance. Though their error rate seems high, their results implicit more visual correspondence information. The geometric-based methods cannot leverage the visual information and fail in most run.

The Willow Object class dataset is mainly for semantic correspondence, where visual-based correspondence is almost invalid. In table. 1, the geometric based methods offer better initialization, and our method contributes an obvious improvement for them. As shown in the example correspondence in Fig. 2, random initialization is qualified for the proposed method.

5. CONCLUSIONS

In this paper, we have introduced a novel technique addressing the visual and geometric deformation problems in image correspondence, by leveraging the consistent distance ratio constraint. For the visual deformation, we add the spatial regularity into the correspondence and achieved significant improvement. For the geometric deformation, a smooth term is introduced to adapt the spatial perturbation. Our experiments on several datasets, show the effectiveness of the proposed method. However, the permutation problem is essentially difficult to optimize, and sensitive to initialization. Our future direction is to find a more robust optimization technique and more reliable initialization.

6. ACKNOWLEDGMENTS

This work was supported by National Key Research and Development Plan(No. 2016YFB0800603), National Natural Science Foundation of China (No. 61422213), "Strategic Priority Research Program" of the Chinese Academy of Sciences (XDA06010701), and National Program for Support of Top-notch Young Professionals.

7. REFERENCES

[1] A. C. Berg, T. L. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In

[2] https://fling.seas.upenn.edu/~xiaowz/dynamic/wordpress/my-uploads/codes/MatchALS.zip

[3] The code of this work can be found in the project of RRWM[4]

[4] http://cv.snu.ac.kr/research/~RRWM/

[5] http://cv.snu.ac.kr/research/~SMCM/

[6] http://cv.snu.ac.kr/research/~SMCP/

2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA, pages 26–33, 2005.

[2] R. Berthilsson. Affine correlation. In *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, volume 2, pages 1458–1460 vol.2, Aug 1998.

[3] H. Bristow, J. Valmadre, and S. Lucey. Dense semantic correspondence where every pixel is a classifier. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 4024–4031, 2015.

[4] V. Černý. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, 45(1):41–51.

[5] M. Cho, K. Alahari, and J. Ponce. Learning graphs to match. In *Proceedings of the IEEE Interational Conference on Computer Vision*, 2013.

[6] M. Cho, J. Lee, and K. M. Lee. Reweighted random walks for graph matching. In *Computer Vision - ECCV 2010 - 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part V*, pages 492–505, 2010.

[7] W. Eric and L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 7(1):17–34, 1985.

[8] P. Foggia, G. Percannella, and M. Vento. Graph matching and learning in pattern recognition in the last 10 years. *IJPRAI*, 28(1), 2014.

[9] W. Förstner. A feature based correspondence algorithm for image matching. *International Archives of Photogrammetry and Remote Sensing*, 26(3):150–166, 1986.

[10] H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *Computer Vision - ECCV 2008, 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I*, pages 304–317, 2008.

[11] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *10th IEEE International Conference on Computer Vision (ICCV 2005), 17-20 October 2005, Beijing, China*, pages 1482–1489, 2005.

[12] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Processing*, 11(4):467–476, 2002.

[13] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157, 1999.

[14] A. Nguyen, M. Ben-Chen, K. Welnicka, Y. Ye, and L. J. Guibas. An optimization approach to improving collections of shape maps. *Comput. Graph. Forum*, 30(5):1481–1491, 2011.

[15] D. Pachauri, R. Kondor, and V. Singh. Solving the multi-way matching problem by permutation synchronization. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1860–1868. Curran Associates, Inc., 2013.

[16] W. K. Pratt. *Digital Image Processing (2Nd Ed.)*. John Wiley & Sons, Inc., New York, NY, USA, 1991.

[17] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN features off-the-shelf: An astounding baseline for recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2014, Columbus, OH, USA, June 23-28, 2014*, pages 512–519, 2014.

[18] R. Roberts, S. N. Sinha, R. Szeliski, and D. Steedly. Structure from motion for scenes with large duplicate structures. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, pages 3137–3144, 2011.

[19] H. Sarnel and Y. Senol. Accurate and robust image registration based on radial basis neural networks. *Neural Computing and Applications*, 20(8):1255–1262, 2011.

[20] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *Int. J. Comput. Vision*, 80(2):189–210, Nov. 2008.

[21] Y. Suh, K. Adamczewski, and K. M. Lee. Subgraph matching using compactness prior for robust feature correspondence. In *Conference on Computer Vision and Pattern Recognition*, 2015.

[22] Y. Suh, M. Cho, and K. M. Lee. Graph matching via sequential monte carlo. In *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III*, pages 624–637, 2012.

[23] R. Szeliski. Computer vision: Algorithms and applications, 2010.

[24] A. Taimori and A. Behrad. A new deformable mesh model for face tracking using edge based features and novel sets of energy functions. *Multimedia Tools Appl.*, 74(23):10735–10759, Dec. 2015.

[25] Y. Terada and U. V. Luxburg. Local ordinal embedding. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 847–855, 2014.

[26] C. Zach, M. Klopschitz, and M. Pollefeys. Disambiguating visual relations using loop constraints. In *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010*, pages 1426–1433, 2010.

[27] J. Zbontar and Y. LeCun. Computing the stereo matching cost with a convolutional neural network. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 1592–1599, 2015.

[28] W. Zhang and C. Ngo. Topological spatial verification for instance search. *IEEE Trans. Multimedia*, 17(8):1236–1247, 2015.

[29] W. Zhang, L. Pang, and C. Ngo. Snap-and-ask: answering multimodal question by naming visual instance. In *Proceedings of the 20th ACM Multimedia Conference, MM '12, Nara, Japan, October 29 - November 02, 2012*, pages 609–618, 2012.

[30] X. Zhou, M. Zhu, and K. Daniilidis. Multi-image matching via fast alternating minimization. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 4032–4040, 2015.