Common Visual Pattern Recognition Using Hierarchical Clustering Methods with Asymmetric Networks

Magdalena Bouza[1] and Bruno Cernuschi Fías[2]

[1] Facultad de Ingeniería, Universidad de Buenos Aires
`bouza.magdalena@gmail.com`,
[2] Instituto Argenitno de Matemáticas (IAM), CONICET
`bcf@ieee.com`

Abstract. In this paper we propose a new method for common visual pattern identification via Directed Graphs. For this we match common feature points between two images and then apply hierarchical clustering methods to one of them to discriminate between different visual patterns. In order to achieve this last task we introduce a technique to obtain an asymmetric dissimilarity function $A_X(x, x')$ between the nodes X of the network $N = (X, A_X)$. For each node, the method weighs the distance between each node and the distance with all the other neighbours. A dendrogram is later obtained as the output of the hierarchical clustering method. Finally we show a criteria to select one of the multiple partitions that conform the dendrogram.

Keywords: clustering, pattern recognition, computer vision, machine learning, image processing, directed graph, hierarchical clustering.

1 Introduction

Visual pattern recognition focuses on identifying visual objects that are commonly presented in two images under comparison. Among the several possible approaches to achieve this end, such as statistical classification and neural networks, etc. [4, 5, 1], a very popular one involves clustering and graph theory.

In pattern recognition problems, certain features are extracted from the image, and some similarity (or dissimilarity) function is computed between all pairs. In the classical approach, this similarity functions need not be a metric but are required, however, to be symmetric.

The use of asymmetric similarity (dissimilarity) functions in visual pattern recognition is quite unexploited. By introducing asymmetries, we are incorporating more information to our problem. In [6] a method for asymmetrical clustering is presented, making use of the *scale invariant feature transform* (SIFT) to obtain the image feature points.

In [7] an axiomatic construction of hierarchical clustering in asymmetric networks is developed, while in [8] an algorithmic approach is presented.

In this paper we propose a method for visual pattern recognition via hierarchical clustering, using asymmetric dissimilarity functions. Two issues will be addressed in this work: first, the construction of the link weights of the graph. Second, the procedure to obtain the dendrogram associated to the problem, and how to select a partition from it.

Throughout this paper matrices will be uppercase italic boldface letters and vectors lowercase italic boldface. For any matrix $\boldsymbol{Z} \in \mathbb{R}^{p \times q}$ (or vector $\boldsymbol{z} \in \mathbb{R}^q$), $\boldsymbol{Z}^T$ ($\boldsymbol{z}^T$) denotes its transpose.

2   Hierarchical Clustering

In this section we will give general notions of hierarchical clustering, as well as present the notation to be used throughout the paper.

Let $N = (X, A_X)$ be a network, with X being the set of n nodes, and $A_X : X \times X \to \mathbb{R}_+$ the dissimilarity function defined for all pairs $x, x' \in X$. $A(x, x')$ must be nonnegative, and $A(x, x') = 0$ if and only if $x = x'$, however it does not need to satisfy the triangle inequality and may be asymmetric. Note that the dissimilarity function can be written in matrix form $\boldsymbol{A}_X$, where i, jth element of such matrix is $A_X(x_i, x_j)$.

Consider the partition $P_X = (B_1, \ldots, B_l)$, where $B_i \cap Bj = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{l} B_i = X$ is a clustering of the set X. In hierarchical clustering, the output is called a dendrogram, D_X, and consists of the collection of nested partitions $D_X(\delta)$, where $\delta \geq 0$ represents the resolution. If nodes x and x' are in the same cluster at resolution δ, we say that they are equivalent at such resolution ($x \sim_{D_X(\delta)} x'$). The dendrogram must satisfy:

(D1) *Boundary Conditions* $D_X(0) = \{\{x\}, x \in X\}$ and $D_X(\delta_0) = \{X\}$ for some δ_0 sufficiently large.

(D2) *Resoltion* For any $\delta_1 < \delta_2$, if we have $x \sim_{D_X(\delta_1)} x'$ for some x, x', then $x \sim_{D_X(\delta_2)} x'$.

Depending on the clustering method used, the dendrogram obtained is different. If $\mathcal{D}$ is the space of all dendrograms and $\mathcal{N}$ is the space of networks, one can define the clustering method as the function $\mathcal{H} : \mathcal{N} \to \mathcal{D}$. For the network $N = (X, A_X)$, $D_X = \mathcal{H}(X, A_X)$ corresponds to the output of the clustering method $\mathcal{H}$.

As shown in [7], there exists an equivalence between dendrograms and ultrametrics, which are mathematically easier to work with. An ultrametric u_X is a function $u_X : X \times X \to \mathbb{R}_+$ that satisfies $u_X(x, x') = u_X(x', x)$, $u_X(x, x') = 0 \Leftrightarrow x = x'$ and $u_X(x, x') \leq \max(u_X(x, x''), u_X(x'', x')) \, \forall \, x, x', x'' \in X$.

The dendrogram D_X defines the ultrametric $u(x, x') := \min \left(\delta \geq 0, x \sim_{D_X(\delta)} x' \right)$. Conversely, for a given u_X defining the relation $\sim_{U_X(\delta)}$ as $x \sim_{U_X(\delta)} x' \Leftrightarrow u_X(x, x') \leq \delta$ we obtain an equivalence relation. The collection of partitions induced by $\sim_{U_X(\delta)} := \left(X \mod \sim_{U_X(\delta)} \right)$ is a dendrogram.

For the construction of hierarchical clustering methods, [7] proposes two axioms: Axiom of Value (A1) and Axiom of Transformation (A2).

Axiom 1. (*Axiom of value*) Given a two-node network $N = (X, A_X)$ with $X = (p, q)$, $A_X(p, q) = \alpha$, and $A_X(q, p) = \beta$, the ultrametric $u_X = \mathcal{H}(X, A_X)$ produced by $\mathcal{H}$ is $u_N(p, q) = \max(\alpha, \beta)$.

Axiom 2. (*Axiom of Transformation*) Consider two networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$, and a dissimilarity reducing map $\Phi : X \to Y$ (for all $x, x' \in X$ $A_X(x, x') \geq A_Y(\Phi(x), \Phi(x')))$. The output ultrametrics satisfy $u_X(x, x') \geq u_Y(\Phi(x), \Phi(x'))$

A hierarchical clustering method $\mathcal{H}$ is said to be admissible if it satisfies both Axioms 1 and 2.

In [7] and [8] two hierarchical clustering methods are proposed: reciprocal (u_X^R) and non-reciprocal (u_X^{NR}) clustering. These two have the property of being extremal ultrametrics, which means that for any admissible clustering method $\mathcal{H}$ and an arbitrary network $N = (X, A_X)$, the resulting ultrametric $u_X = \mathcal{H}(X, A_X)$ satisfies that $u_X^{NR}(x, x') \leq u_X(x, x') \leq u_X^R(x, x')$. In this work we shall focus only on the non-reciprocal method.

To obtain the non-reciprocal clustering method one first needs to define the unidirectional minimum chain cost as $\tilde{u}_X^{NR}(x, x') = \min_{C(x,x')} \max_{i | x_i \in C(x,x')} A_X(x, x')$, where $C(x, x')$ is a chain linking nodes x and x'. The non-reciprocal clustering method $\mathcal{H}$ with ultrametric output $u_X^{NR} = \mathcal{H}(X, A_X)$ is given by

$$u_X^{NR}(x, x') = \max(\tilde{u}_X^{NR}(x, x'), \tilde{u}_X^{NR}(x', x)) \tag{1}$$

From 1 it can be seen that in non-reciprocal clustering, we allow for x and x' to be joined if there exist (possibly) different chains $C_1(x, x')$ and $C_2(x', x)$ whose costs are less than δ.

3 Wang-Ma Method

We shall first explain the method proposed in [6], which, up to our knowledge, is the only other approach to the use of directed graphs in visual pattern recognition.

3.1 Extraction of the dissimilarity matrix

In order to obtain the asymmetric link weights, they first find a symmetric graph for the problem. Given two images I_1 and I_2, the scale invariant feature transform (SIFT) technique [10] is used to extract feature points from them. The output of this technique consists of two sets of vectors: the feature points ($\boldsymbol{f}$) and the feature descriptors ($\boldsymbol{d}$). The first is a 4-dimension vector containing orientation, magnitude and position of each feature. The second is a 128-dimension vector, one for each feature point, that contains the local feature representation. One of the main advantages of this method is that the features are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint [10].

Let $(\boldsymbol{f}_1(s), \boldsymbol{d}_1(s))$, $s = 1, \ldots, S$ and $(\boldsymbol{f}_2(t), \boldsymbol{d}_2(t))$, $t = 1, \ldots, T$ be the feature points and feature descriptor vectors for images I_1 and I_2 respectively. A point-to-point correspondence is obtained for all points from one set to the other, computing the euclidean distance between each pair $\boldsymbol{f}_1(i)$, $\boldsymbol{f}_2(j)$, with i being a point from I_1 and j from I_2. To reduce the number of points and in order to eliminate incorrect point-to-point matches, a pruning process is applied based on a thresholding process. This threshold η is empirically determined. Each remaining point-to-point correspondence will be considered a node in the graph, and the connection between any two nodes is considered a link.

Given the two sets A with M elements, and B with N, obtained from I_1 and I_2 respectively, each vertex is a duple (a, b), where $a \in A$ and $b \in B$. For a pair of vertices $v_i = (a_i, b_k)$ and $v_j = (a_j, b_l)$ we introduce the notation $d_{i,j} = \|a_i, a_j\|_2$ and $d_{k,l} = \|b_k, b_l\|_2$.

Based on the Pairwise Spatial Consistency (PSC) principle (see [6],p. 1410) the weights between nodes v_i and v_j for the symmetric graph are computed as $w_{i,j} = |d_{k,l} - \alpha d_{i,j}|$. The value of α is unknown, but for the PSC principle to be valid one must have $\alpha = C_n = \frac{d_{k,l}}{d_{i,j}}$.

If there are K vertices, then the weights wi, j can be represented by a $K \times K$ matrix, $\boldsymbol{W}$. The element in row i and column j is simply $w_{i,j}$. This matrix is symmetric $(\boldsymbol{W} = \boldsymbol{W}^T)$ and the diagonal elements are zero.

Once the undirected graph's weights are obtained, [6] proposes an intuitive approach to map it into an asymmetrical graph. The basic idea is to compare each link weight with the link weights of all direct neighbours. For example, consider the graph in Fig. 1. For node v_1, the weight $w_{1,2}$ can be regarded small with respect to $w_{1,3}$ and $w_{1,4}$. However, for node v_2 this same weight is quite large compared to the link weights shared with other neighbours. It is intuitive that node v_1 should favour joining with v_2 more than v_2 with v_1.
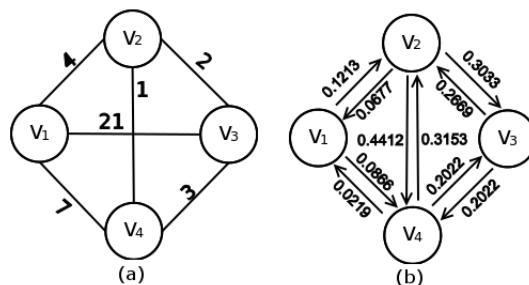

Fig. 1: Example: (a) Undirected graph, (b) Directed graph.

This mapping of the weights $w_{i,j}$ into the directed graph weights $e_{i,j}$ is done through the *Gaussian Link-Weight Mapping Function* [6] . To reflect the asymmetries in the graph, the *n-ranking* process is proposed. Calling γ_i the ith smallest weight $w_{i,j}$ between node v_i and its neighbours

$$e_{i,j} = \frac{1}{\gamma_i} \exp\left(-\frac{w_{i,j}^2}{2\gamma_i^2}\right), \qquad e_{j,i} = \frac{1}{\gamma_j} \exp\left(-\frac{w_{i,j}^2}{2\gamma_j^2}\right) \qquad (2)$$

This weights can be expressed in matrix form, $\boldsymbol{E}$, which is asymmetrical.

3.2 Common visual pattern extraction

Once the asymmetric weights $e_{i,j}$ are obtained, for each α considered a new digraph is constructed. A subset of the digraph that yields the largest average link weight will be identified and denoted as the subgraph. The average link weight is computed as $C = (\boldsymbol{x}^T \boldsymbol{E} \boldsymbol{x})/L^2$, where $\boldsymbol{x}$ is a vector of length K that has $x_i = 1$ if v_i belongs to the subgraph, $x_i = 0$ otherwise. Thus, the maximization problem $\text{argmax}_{\boldsymbol{x}} C$, subject to $\sum_{i=1}^{K} = L$ and $x_i = \{0, 1\}$ must be solved. Note that since $\boldsymbol{E}$ is asymmetrical, traditional optimization methods such as the spectral method can not be used. Each identified subgraph is then subjected to a thresholding process: if $C \geq \nu$ then the subgraph is considered a strongly-associated subgraph and is identified as one common visual pattern. This threshold is also determined empirically.

If two distinct objects are present in both images with the same scale-change factor α, they will be identified in the same subgraph. In order to discriminate patterns in this situation some extra processing is needed.

4 Our Proposed Method

4.1 Asymmetric dissimilarity matrix extraction

As in the method proposed in [6], we will also apply the SIFT technique on each of the pictures to extract the corresponding features. If R and Q feature points are detected in I_1 and I_2 respectively, the process will yield R feature points and their corresponding feature descriptors $(\boldsymbol{f}_1, \boldsymbol{d}_1)$ for I_1 and Q for I_2 $(\boldsymbol{f}_2, \boldsymbol{d}_2)$.

We propose to first find one-to-one correspondences between points from image 1 and image 2 and apply a pruning process, just like in [6], to discard wrong matches. Once correct matches between I_1 and I_2 are found, we can simply detect objects in one of the images, and use the matched points to locate them in the other.

The first task is achieved by computing $d_{i,j} = \|\boldsymbol{d}_{1,i} - \boldsymbol{d}_{2,j}\|_2 \ \forall i = 1, \ldots, R, \ j = 1, \ldots, Q$, where indexes i, j indicate the ith and jth features in $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ respectively. If $d_{i,j} < \xi$ then we consider it as a correct match. In this paper we empirically determined the threshold ξ to be 0.5.

The nodes for the graph will be those points in one image, say I1, which found a match in the other. To compute the weights between nodes, we use feature point vectors. If two feature points are close to each other, and belong to the same pattern, their distance is expected to be small. Furthermore, since no two vectors $\boldsymbol{f}_{1,i}, \boldsymbol{f}_{1,j}, j \neq i$ can be equal (it contains the position), if the distance between two feature points is zero, then they must be the same. Thus, we define the distance between two nodes v_i, v_j

$$\text{dist}_{i,j} = \|\boldsymbol{f}_{1,i} - \boldsymbol{f}_{1,j}\|_2 \tag{3}$$

In order to obtain our asymmetric dissimilarity matrix, $\boldsymbol{A}$ we shall use an approach similar to the one proposed in [6]. To avoid outlier problems, we will use robust error measure, in particular the Forsyth error function: $e(r) = \frac{r^2}{a^2+r^2}$. We shall define a different a_i for the each node v_i. The goal is that those nodes which have some distance significantly smaller compared to the remaining neighbours, give more importance to those links. This implies a steeper error function around zero, which, occurs for larger values of a. We then define a_i for the ith node as $a_i = 2\max(\tilde{a}_k) - \tilde{a}_i$, where $\tilde{a}_k$ is the minimum distance between nodes for node v_k. The link weights, and thus the dissimilarity matrix, are computed as:

$$A_{i,j} = \frac{\text{dist}_{i,j}^2}{a_i^2 + \text{dist}_{i,j}^2}; \qquad A_{j,i} = \frac{\text{dist}_{i,j}^2}{a_j^2 + \text{dist}_{i,j}^2}; \tag{4}$$

Theorem 1. *The matrix whose coefficients are obtained from 4 is in fact a dissimilarity matrix, i.e. satisfies non-negativity ($\boldsymbol{A}_{i,j} \geq 0$) and $\boldsymbol{A}_{i,j} = 0$ iff $i = j$.*

Proof. First, note that $\text{dist}_{i,j} = \|\boldsymbol{f}_{1,i} - \boldsymbol{f}_{1,j}\|_2 \geq 0$ (5), and $dist_{i,j} = 0 \Leftrightarrow i = j \equiv v_i = v_j$. (6).
    Second, $a_i = 2\max(\tilde{a}_k) - \tilde{a}_i \geq 2\max(\tilde{a}_k) - \max(\tilde{a}_k) = \max(\tilde{a}_k) > 0$. (7) $A(v_i, v_j) = \boldsymbol{A}_{i,j} = \frac{\text{dist}_{i,j}}{a_i^2 + \text{dist}_{i,j}}$. From **??** and 7, we have that $a_i^2 + \text{dist}_{i,j} > 0$, thus $A(v_i, v_j) \geq 0$ and $A(v_i, v_j) = 0$ only when $\text{dist}_{i,j} = 0$, which, from **??**, occurs if and only if $v_i = v_j$.

4.2   Common Visual Pattern Extraction

Once we have our dissimilarity matrix $\boldsymbol{A}$, we can apply the theory presented in Section 2 to obtain the corresponding ultrametric $u_X(x, x')$ for a given hierarchical clustering method $\mathcal{H}$.
    As seen from 1, non-reciprocal clustering searches for directed chains of minimum infinity norm cost in $\boldsymbol{A}_X$ to construct $\tilde{u}_X$ and selects the maximum of $\tilde{u}_X, \tilde{u}_X^T$. In [8], they show an algorithm for performing such operations using matrix powers in the dioid algebra $(\mathbb{R}^+ \cup \{+\infty\}, \min, \max)$[3]. Denoting the operations min and max with $\oplus$ and $\otimes$ respectively we have that the matrix product is given by

$$[\boldsymbol{A} \otimes \boldsymbol{B}]_{i,j} = \bigoplus_{k=1}^{n} (\boldsymbol{A}_{i,k} \otimes \boldsymbol{B}_{k,j}) = \min_{k \in [1,n]} \max (\boldsymbol{A}_{i,k}, \boldsymbol{B}_{k,j}) \tag{8}$$

Proposition 1. *For a given network $N = (X, \boldsymbol{A}_X)$ with K nodes, the non-reciprocal ultramentric can be computed as $u_X^{NR} = \max\left(\boldsymbol{A}_X^{(K-1)}, (\boldsymbol{A}_X^T)^{(K-1)}\right)$*

Proof. See [9]

    Once we obtain the output ultrametric for our hierarchical clustering method, we use the joining algorithm [2] to extract the dendrogram D_X.

As stated in Section 2, a dendrogram D_X is a collection of nested partitions. The question is which of this partitions better identifies each visual pattern presented in both images. Ideally, we would like a partition in which all different patterns conform different clusters and no node belonging to one pattern is clustered with some other object.

Let $\boldsymbol{\delta} = \delta_1, \ldots, \delta_j$ be de resolutions at which a new cluster is formed, in ascending order. In order to select a partition, we propose to compute the difference between consecutive elements of $\boldsymbol{\delta}$ and select de largest, i.e. $\max_{t=\{2,\ldots,j\}} (\delta_t - \delta_{t-1})$. If this maximum is achieved between δ_k and δ_{k+1}, we will select partitions that cluster together at resolution $\delta < \delta_{k+1}$. While simulations suggest this method works, further research will be done in order to obtain some other criteria that satisfies some optimality condition.

5   Simulations

In this section we will analyse the performance of the method. To simulate practical situations, we shall use a synthetic image. For this we will randomly generate P feature points corresponding to l different visual objects. The same feature points will be present in both images, but a predefined rotation factor will be added to the features in I2. Also, some outliers will be randomly generated in both images. An outlier is a feature point present in one image, but not in the other. In order to simulate geometric perturbations, a Gaussian white noise $\mathcal{N}(0, \sigma^2)$ is also added to features in I2 to deviate the computed spatial coordinate. In the simulations the pattern recognition technique will be applied to I2, since it represents the worst case scenario (using I1, rotation and noise wouldn't affect the pattern identification).

We ran trials with P=12 with the number of outliers ranging from 0 to 60 with no noise and some predefined rotation applied. In all cases, a correct point to point detection as well as a correct object classification were obtained. This shows that our method is immune to this sort of outliers. We also tested the performance under the effect of Gaussian noise for $\sigma = \{1, 2, \ldots, 10\}$. As expected from the previous experiment, the performance with and without outliers was the same for each value of σ. For $\sigma < 10$ all matches were correct, and could perform a correct pattern classification. For $\sigma = 10$ we found all point to point correspondences were correct, however we lost the ability to identify different objects.

The results presented in [6] for the Wang-Ma method show loss of performance when both outliers and gaussian noise are present is significant for values of $\sigma \geq 4$. Also, for a fixed σ there is significant performance loss as the number of outliers increases. Comparing the results of both methods suggests our presented algorithm has better performance.

Secondly, we would like to evaluate how well the method perform with natural images. This is shown in figure 2. It can be seen that both objects are correctly identified. It also appears to have no one to one correspondence errors between images. Finally, in the dendrogram we see how the hierarchical clustering method satisfies Axioms 1 and 2.
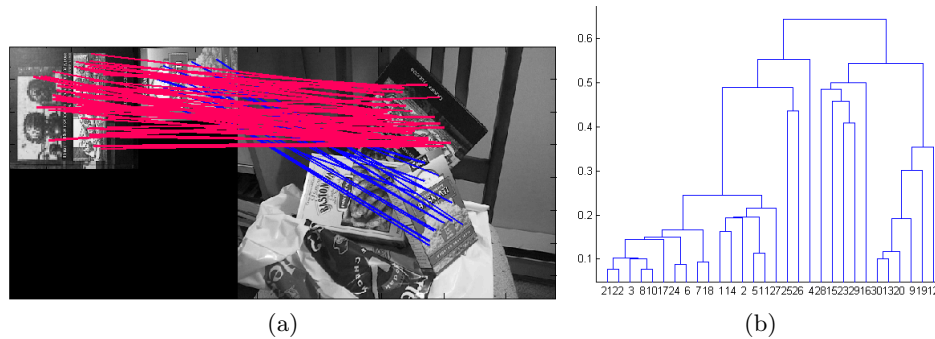
Fig. 2: Simulation result, (a) Final matching, (b) Dendrogram

6 Conclusions

In this paper, a new method is proposed to detect all common visual patterns presented in two images using directed graphs and hierarchical clustering methods. Experimental results suggest improvement with respect to other similar methods. Future work will be done in the optimization of the criteria to select partitions. Finally, it is worth to mention that while this paper is focused on the application in image processing, the approach could be applied to other fields.

References

1. Duda, R. O., Hart, P. E.: Pattern Classification and Scene Analysis. Wiley (1973)
2. Hartigan, J.: Clustering Algorithms. Wiley (1975)
3. Gondran, M., Minoux, M.: Graphs, dioids and semi rings: New models and algorithms. Springer (2008).
4. Bishop, C. M.: Pattern Recognition and Machine Learning. Springer (2007)
5. Marques de Sa, J. P.: Pattern Recognition: Concepts, Methods and Applications. Springer (2001)
6. Wang, C., Ma, K.K.: Common Visual Pattern Discovery via Directed Graph. IEEE Transactions on Image Processing, Vol. 23, No. 3, 1408–1418 (2014)
7. Carlsson,G., Mémoli, F., Ribeiro, A., Segarra, S.: Alternative Axiomatic Consuctions for Hierarchical Clustering of Assymetric Networks. IEEE GlobalSIP 2013, 791–794 (2013)
8. Carlsson,G., Mémoli, F., Ribeiro, A., Segarra, S.: Hierarchical Clustering Methods and Algorithms for Assymetric Networks. IEEE, Asilomar 2013, 1773–1777 (2013)
9. Carlsson, G., Mémoli, F., Ribeiro, A., and Segarra, S.: Axiomatic construction of hierarchical clustering in asymmetric networks. https://ing.seas.upenn.edu/ssegarra/wiki/index.php?n=Research.Publications, (2012).
10. Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints International Journal of Computer Vision, 60, 2 , 91–110 (2004).