A Topology-Independent Similarity Measure for High-Dimensional Feature Spaces

Jochen Kerdels[1] and Gabriele Peters[2]

[1] DFKI - German Research Center for Artificial Intelligence,
Robotics Lab, Robert Hooke Str. 5, D-28359 Bremen, Germany
[2] University of Dortmund, Department of Computer Science,
Computer Graphics, Otto-Hahn-Str. 16, D-44221 Dortmund, Germany

Abstract. In the field of computer vision feature matching in high dimensional feature spaces is a commonly used technique for object recognition. One major problem is to find an adequate similarity measure for the particular feature space, as there is usually only little knowledge about the structure of that space. As a possible solution to this problem this paper presents a method to obtain a similarity measure suitable for the task of feature matching without the need for structural information of the particular feature space. As the described similarity measure is based on the topology of the feature space and the topology is generated by a growing neural gas, no knowledge about the particular structure of the feature space is needed. In addition, the used neural gas quantizes the feature vectors and thus reduces the amount of data which has to be stored and retrieved for the purpose of object recognition.

1 Introduction

In the field of computer vision objects are often represented by feature vectors describing local areas of them (e.g., [1,2,3]). These local descriptors often are vectors of high-dimensional feature spaces. To identify equal or similar objects, for example for the purpose of object recognition, feature matching techniques are common, and for these matching techniques similarity measures for the feature vectors are needed. One major problem when choosing the similarity measure is often the lack of knowledge about the structure of the feature space. For example the features in the SIFT feature space as described by Lowe [1] are not uniformly distributed. Using the Euclidean distance – as Lowe does – leads to the problem, that the direct distance between two features cannot be used as an absolute measure of their similarity. Accordingly, Lowe uses a workaround for this problem with the drawback that it requires each object to have at least one unique (i.e., identifying) feature. This is a general problem of non-uniform feature spaces. Wrongly presumed uniformity can result in a classification of unsimilar features as similar and vice versa. Some approaches try to improve the matching of features in the non-uniform feature space by using dimensionality reduction techniques such as Principal Component Analyses (PCA) [4]. For example Ke and Sukthankar showed in [5] that using PCA can improve

J. Marques de Sá et al. (Eds.): ICANN 2007, Part II, LNCS 4669, pp. 331–340, 2007.

the matching of features in SIFT space. But they still rely on the Euclidian distance and the proposed workaround of Lowe. This paper describes how to obtain a similarity measure which is suitable for the task of feature matching without knowledge of the particular structure of the high-dimensional feature space. Using a growing neural gas as described by Fritzke in [6], the topology of the feature space is first learned and then used as a basis for the similarity measure. The similarity between two feature vectors will incorporate the length of the shortest path between those two nodes of the neural gas the feature vectors are mapped on. Besides the ability to adapt to non-uniformly distributed feature spaces the neural gas also quantizes the feature vectors. On the one hand, this can be accompanied by a possible loss of information. But on the other hand, it also vastly reduces the amount of data which has to be stored and retrieved for feature matching purposes. In section 2 we recall the functionality of growing neural gas on which our similarity measure will be based. The proposed measure is derived in section 3, after which the description of the experiments (section 4), the summary of the results (section 5), and a conclusion (section 6) follow.

2 Growing Neural Gas Revisited

The growing neural gas (GNG) used for the similarity measure is described in detail by Fritzke in [6]. The GNG is similar to the self organizing maps of Kohonen [7]. In contrast to the self organizing maps the GNG does not have a fixed number of nodes (often also called "neurons" or "units"). It is subject to a data driven growing process which ends when a halting criterion (e.g., a minimal quantization error or a maximum number of nodes) is complied with. Figure 1 depicts the growing process of a GNG on a non-uniformly distributed, two-dimensional feature space Ω. The set $\mathcal{S}$ of nodes of the GNG is initialized with two nodes. Both nodes are associated with different random vectors $w \in \Omega$, called *reference vectors*. That are random positions in the high-dimensional feature space Ω. In addition, every node $s \in \mathcal{S}$ has an accumulated error E_s initialized with 0. The edges of the GNG, which connect the nodes, have an attribute "age". During the growing process this attribute makes it possible to detect edges which are not needed any more and thus to delete them. At first, the set of edges $\mathcal{C}, \mathcal{C} \subseteq \mathcal{S} \times \mathcal{S}$, between the nodes is empty. A new feature vector $\xi \in \Omega$ is processed as follows: Those two nodes s_1 and s_2 the reference vectors w_{s_1} and w_{s_2} of which are closest (in terms of Euclidean distance) to the feature vector ξ are selected. If there is no connecting edge between s_1 and s_2 in $\mathcal{C}$, an edge between s_1 and s_2 is added to $\mathcal{C}$. By setting the age of the edge to 0 the edge (if it already existed) is refreshed. The accumulated error E_{s_1} of the nearest node s_1 is increased by the square of the distance between the feature vector ξ and the reference vector w_{s_1}. Next, the reference vector of s_1 and all reference vectors of the direct neighbors of s_1 are adapted. The age is increased by one for those edges the endpoints of which have been moved in the previous step. If the age of an edge reaches a threshold $a_{\max}$, the edge will be removed. When this leads to an isolated node it is also removed. The value of $a_{\max}$ defines the

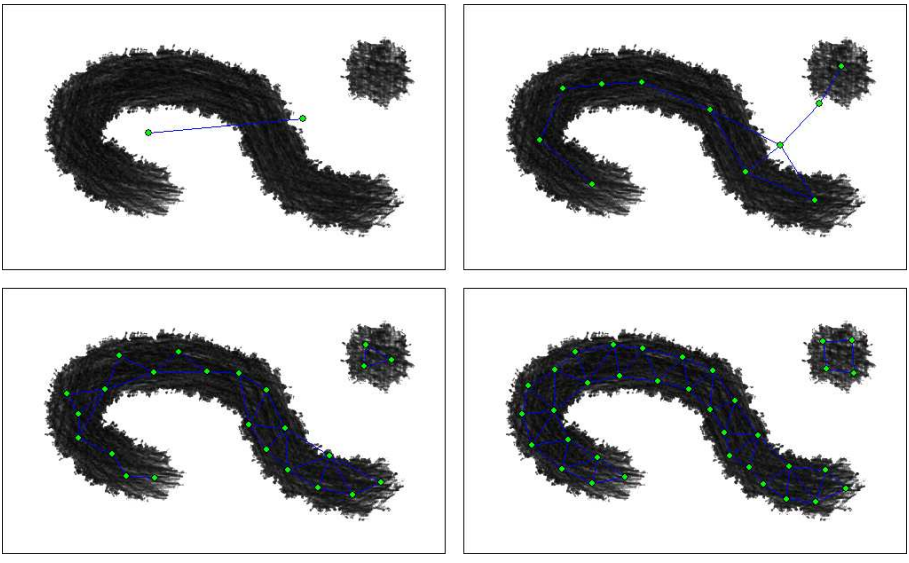


Fig. 1. A growing neural gas at different points in time. The dark shaded areas represent the values of a non-uniformly distributed, two-dimensional feature space Ω which should be characterized by the GNG.

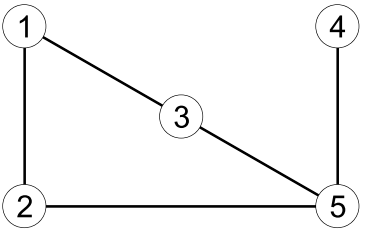
stiffness of the generated topology. A small value leads to an unstable topology, a very high value leads to an only slow detachment of isolated areas in the feature space. A value of 100 for $a_{\max}$ turned out to be a good medium choice. After a number λ of input feature vectors (e.g., $\lambda = 300$) a new node r is added to the GNG if the halting criterion is not met yet. For this purpose two nodes of $\mathcal{S}$ are selected: first, the node q with the highest accumulated error E_q and secondly, that node f of all nodes adjacent to q that has the highest accumulated error E_f. The new node r is added to $\mathcal{S}$ and obtains a reference vector w_r which is the average between w_q and w_f. The accumulated error E_r of r is interpolated between the accumulated errors E_q and E_f, which were reduced by a fraction in a preceding step. Next, the set of edges $\mathcal{C}$ is extended by an edge between r and q and an edge between r and f. The edge between q and f is removed from $\mathcal{C}$. In a last step the accumulated errors of all nodes are reduced by a fraction β. This last step simulates a kind of aging on the accumulated errors, thus giving newer errors more weight and avoiding a build-up of small errors over time. With respect to the behaviour of the neural gas the parameter β influences how good the neural gas can adapt to fine structures in the feature space. Summarizing, the algorithm produces a graph with nodes explicitly linked to their closest neighbors. The graph is a subset of the Delaunay triangulation, a property we refer to later in subsection 3.3 as Delaunay property. Using the GNG to quantize vectors of high-dimensional feature spaces, the return value for an input feature vector $\xi \in \Omega$ could be the reference vector or just the the nearest node (which is the definition we will apply in subsection 3.3).

3 Defining a Similarity Measure on a Feature Space

The neural gas described in the previous section generates a topology of the feature space which can be used for a similarity measure. The generated topology is represented by a graph the nodes of which are the nodes of the GNG and the edges of which connect neighboring and thus similar nodes. Accordingly, we can describe the distance between two nodes (and later between two feature vectors of the high-dimensional space) by the number of edges on the shortest path between them. By doing so we utilize the ability of the neural gas to reflect the structure of the feature space. To develop our similarity measure we need a distance matrix for the GNG graph. This distance matrix is derived by the calculation of paths of length n, where n is the number of edges connecting two nodes. How to determine nodes that are reachable on paths with a distinct length is described in subsection 3.1. The derivation of the distance matrix is then described in subsection 3.2, after which we are able to define our topology-independent similarity measure in subsection 3.3.

3.1 Paths of Distinct Length

Once the growing neural gas has learned the topology of the feature space using N nodes, an $N \times N$ distance matrix D can be generated that contains for every node the shortest distance to all other nodes. The distance matrix D can be calculated using the fact that the adjacency matrix of a graph to the power of n codes for every node the subsequent nodes which are n edges away. This is explained in the following. First we give an example of a simple graph:



The adjacency matrix A of this graph describes in every column $\boldsymbol{a}_i$ the direct neighbors of node $i, i = 1, \ldots N$:

$$A = \begin{pmatrix} 0\,1\,1\,0\,0 \\ 1\,0\,0\,0\,1 \\ 1\,0\,0\,0\,1 \\ 0\,0\,0\,0\,1 \\ 0\,1\,1\,1\,0 \end{pmatrix}$$

The multiplication of a column vector $\boldsymbol{b} := (b_1, b_2, \ldots, b_N)^T$ and an $N \times N$ matrix M can be seen as a linear combination of the columns $\boldsymbol{m}_i$ of M:

$M\boldsymbol{b} = \ b_1\boldsymbol{m}_1 + b_2\boldsymbol{m}_2 + \ldots + b_N\boldsymbol{m}_N$. The multiplication of two $N \times N$ matrices G and H can be done column by column as multiplication of columns $\boldsymbol{h}_i$ of H and matrix G:

$$GH = \begin{pmatrix} G\boldsymbol{h}_1 & G\boldsymbol{h}_2 & \ldots & G\boldsymbol{h}_N \end{pmatrix}.$$

Thus, the square $A' = AA$ of adjacency matrix A contains in every column $\boldsymbol{a'}_i$ a linear combination of the columns $\boldsymbol{a}_j$ of A:

$$A' = \begin{pmatrix} \boldsymbol{a'}_1 & \boldsymbol{a'}_2 & \ldots & \boldsymbol{a'}_N \end{pmatrix} = \begin{pmatrix} A\boldsymbol{a}_1 & A\boldsymbol{a}_2 & \ldots & A\boldsymbol{a}_N \end{pmatrix}.$$

As every column $\boldsymbol{a}_i$ of adjacency matrix A describes the adjacent nodes of node i, every column $\boldsymbol{a'}_i$ of A' describes all adjacent nodes of the adjacent nodes of node i or in other words, it describes those nodes which are reachable from node i on paths of length 2. The values of the entries a'_{ij} of matrix A' describe, how many paths of length 2 between node i and node j exist. Accordingly, another multiplication of A' with A results in a matrix $A'' = A'A = AAA$ the columns $\boldsymbol{a''}_i$ of which describe the nodes that are reachable from node i on paths of length 3. In general, the adjacency matrix of a graph to the power of n codes the nodes that are connected via paths of length n.

3.2 Distance Matrix D

For the task of calculating the distance matrix D, i.e., the matrix that contains for every node the shortest distance to all other nodes, the precise values of the exponentiated adjacency matrix are not needed. The information whether or not there is a path of length n between two nodes is satisfactory. Thus it is sufficient to use boolean values 0 and 1 and to replace the addition by the disjunction and the multiplication by the conjunction when exponentiating the adjacency matrix. Then the distance matrix D is calculated as follows:

$$D = D_0 - \sum_{i=0}^{N-1} \bigvee_{j=0}^{i} A^j \qquad with$$

$$D_0 = \begin{pmatrix} N & \cdots & N \\ \vdots & \ddots & \vdots \\ N & \cdots & N \end{pmatrix}, \qquad A^0 = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix},$$

and $A^1 = A$ the adjacency matrix, $A^2 = A'$, $A^3 = A''$, etc. D is symmetric and its entries are either positive or zero, zero if and only if they are elements of the diagonal. The complete computation of distance matrix D requires N matrix multiplications. For each matrix multiplication N^2 matrix elements have to be computed, which requires N conjunctions and $N - 1$ disjunctions for every

element. Therefore the distance matrix D can be calculated in $\mathcal{O}\left(N^4\right)$. But the computing time can be further reduced to $\mathcal{O}\left(kN^3\right)$ if a maximum depth k for a path between two nodes is used instead of a complete computation of the distance matrix with depth N. As normally one is not interested in the similarity of features beyond a certain threshold, the exact distance between those very unsimilar features can be disregarded without posing too many restrictions to possible applications. Thus, for most applications at stake the computational complexity of $\mathcal{O}\left(N^3\right)$ of the proposed similarity measure compares to other all-pairs shortest path algorithms such as Floyd-Warshall [8].

3.3 Topology-Independent Similarity Measure *d*

Having defined the distance matrix D for the nodes of the GNG, we will now derive our topology-independent similarity measure d. (As we will define d as a pseudometric, we should properly speak about a *dissimilarity measure* rather than a *similarity measure*, but we will adhere to the more colloquial term.)

Let Ω be the high-dimensional feature space. Features ξ are represented as points in this metric space: $\xi \in \Omega$. Furthermore, let $\mathcal{S} := \{s_1, s_2, \ldots, s_N\}$ be the set of nodes of the GNG as introduced in section 2. The quantization operation induced by the growing neural gas is a mapping $Q : \Omega \to \mathcal{S}, Q(\xi) = s$, with s the node ξ is assigned to by the GNG. Now we first can define a metric $\widetilde{d}$ on $\mathcal{S}$:

$$\widetilde{d} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}, \qquad \widetilde{d}(s_i, s_j) := \mathtt{d}_{ij}$$

with $\mathtt{d}_{ij}$ the entries of the distance matrix D: $D = \left(\mathtt{d}_{ij}\right)_{i,j=1,\ldots,N}$. The metric axioms (i) non-negativity, (ii) identity of indiscernibles, (iii) symmetry, and (iv) triangle inequality obviously hold true for $\widetilde{d}$ because of the properties of D mentioned in subsection 3.2. Given $\widetilde{d}$, we can define a topology independent pseudometric d on Ω now:

$$d : \Omega \times \Omega \to \mathbb{R}, \qquad d(\xi, \eta) := \widetilde{d}\left(Q(\xi), Q(\eta)\right).$$

As different feature vectors can be mapped on the same node of the GNG the second metric axiom is not necessarily fulfilled. This means d is a pseudometric only, i.e., only the following axioms hold true:

(i) $d(\xi, \eta) \geq 0$,
(iii) $d(\xi, \eta) = d(\eta, \xi)$,
(iv) $d(\xi, \eta) \leq d(\xi, \rho) + d(\rho, \eta)$.

Properties (i) and (iii) are induced by the corresponding properties of $\widetilde{d}$. Property (iv) holds true because of the Delaunay property of the growing neural gas mentioned in section 2. The concrete values of d depend on the granularity of the similarity measure. This granularity is determined by the halting criterion of the GNG. Thus, the precision of the similarity measure can be controlled by adjusting the halting criterion. Figure 2 shows an example of our similarity measure in a schematical way.

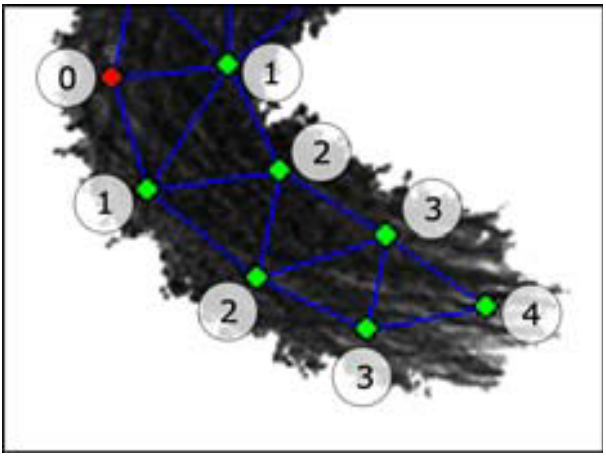


Fig. 2. Topology-independent similarity measure. The gray areas represent the feature vectors ξ of the high-dimensional space Ω. Some nodes of the growing neural gas, i.e., some elements of set $\mathcal{S}$, are depicted by red and green dots. The numbers they are labeled with are those values of the metric $\widetilde{d}$ which express the distance between the red node and each of the green nodes. For example, the shortest distance between the red node and the upper green node labeled with "2" is a path of 2 edges.

4 Experiments

We carried out our experiments on a database of 798 gray value images, a few of which are shown in figure 3. As features we consider patches of 18×18 pixels, thus our high-dimensional feature space Ω has 324 dimensions. These features are not optimal descriptors for the purpose of object recognition. Nevertheless, we chose them for the evaluation of the similarity measure because they can be evaluated more easily by visual inspection than more advanced feature descriptors such as the SIFT vectors, for which a visual interpretation is much harder. Per sample image we extracted about 250 feature vectors, the positions of which have

Fig. 3. Database. A selection of 798 sample images on which we carried out our experiments.

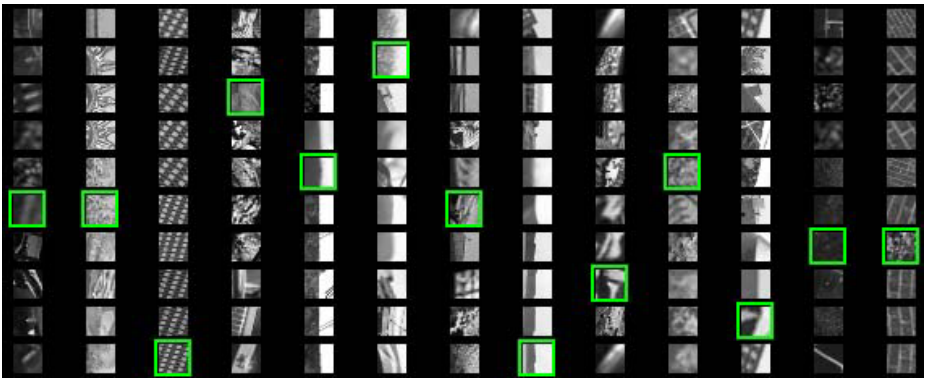


Fig. 4. Quantization of feature vectors by the growing neural gas. Each column stands for one node of the GNG, e.g., the first column represents $s_1 \in \mathcal{S}$, the second column $s_2 \in \mathcal{S}$, etc. Each entry (i.e., row) of a column shows one feature vector $\xi \in \Omega$ in form of a 18×18 gray value patch, which has been assigned to this node. In each case a column shows the last 10 feature vectors which have been mapped onto it. The features marked by a green frame are those which are the last assigned.

been determined with a KLT detector [9], resulting in a total of about 200,000 features. We ran the growing neural gas algorithm as described in section 2. After the processing of $\lambda = 300$ feature vectors we added one node to the GNG and stopped the growing procedure after it consisted of 300 neurons. (All of the 200,000 features have been used for the generation of the GNG according to the algorithm described in section 2. After 300 nodes have been incorporated into the gas not many changes of the topolopy of the GNG were caused by the remaining features. Thus, those remaining features contributed to the stabilization of the GNG only, rather than to its overall topoloy.) Figure 4 shows exemplarily how some feature vectors have been quantized.

5 Results

The final purpose of defining feature vectors and endowing their space with a suitable similarity measure, is the adequate encoding of the characteristics we intend to measure. In this case the application is an encoding of the visual characteristics of objects for purposes such as storage, classification, or recognition. Therefore, we have to analyze whether the features classified as being (mathematically) similar are also assessed by humans as being (visually) similar. In other words, the similarity (or difference) we determine by the proposed method must bear some correlation with the perceptual similarity (or difference) of two feature vectors. As Santini and Jain point out in [10], if our systems have to respond in an "intelligent" manner, they must use a similarity model resembling the perceptual similarity model of humans. Having these considerations in mind, we decided to assess the quality of the similarity measure d by a visual inspection

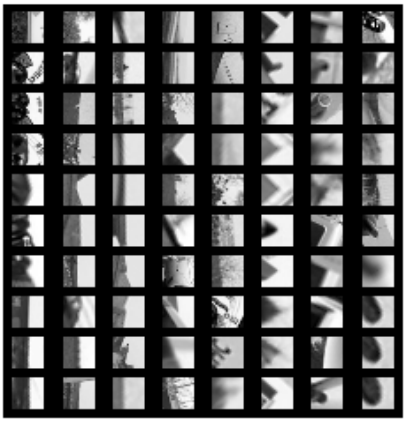


Fig. 5. Classification of feature vectors into 8 neighboring nodes of the growing neural gas. The columns represent the nodes. The entries of the columns are the last 10 feature vectors which were mapped onto them. The distance between the nodes is equivalent to the number of columns between them.

of the feature classification. Figure 5 shows again nodes of the growing neural gas, represented by 8 columns of the 10 last assigned feature vectors each. This time neighboring columns show *adjacent* nodes of the GNG, thus the number of columns between two nodes in the diagram is proportional to the distance between the nodes in the GNG. For example, the node represented by the first column and the node represented by the last column have a distance of 7 edges. We can summarize the results of the visual inspection as follows: Firstly, the similarity between features belonging to one node (i.e., features within one column) is, in general, larger than between features of different nodes. Secondly, one can observe a gradual decrease in similarity from the left to the right node for most of their assigned features. For example, the second column displays a larger overall similarity to the first column than the last column. Summarizing, one can say that the classification of features emerged from the proposed similarity measure corresponds to the assessment of the perceptual similarity by humans. Object recognition experiments remain to be done.

6 Conclusion

We considered the problem of endowing a feature space with an adequate similarity measure. Often researchers make unwarranted assumptions about the metric of the space. Usually it is assumed to be Euclidean. In this paper we presented a similarity measure for high-dimensional feature vectors which is independent from the actual structure of the feature space in the sense that no a priori knowledge on the topology of the feature space is necessary. The similarity measure

is based on the advantageous distribution of the nodes in a growing neual gas. In addition, the use of a growing neural gas provides a quantization of the high-dimensional feature vectors. Despite a possible loss of information, this reduces the amount of data which has to be stored and searched for in further processing. The described similarity measure is particularly useful for object recognition tasks where an object is represented by a set of feature vectors, as it seems to correspond to human perceptual similarity assessment.

Acknowledgements

This research was funded by the German Research Association (DFG) under Grant PE 887/3-1.

References

1. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
2. Schmid, C., Mohr, R.: Local Grayvalue Invariants for Image Retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence 19(5), 530–535 (1997)
3. Kadir, T., Brady, M.: Saliency, Scale and Image Description. Int. J. of Computer Vision 45(2), 83–105 (2001)
4. Kessler, W.: Multivariate Datenanalyse. Wiley-VCH, Weinheim (2006)
5. Ke, Y., Sukthankar, R.: PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In: Proc. IEEE Conf. Comp. Vis. and Pat. Rec (CVPR) (2004)
6. Fritzke, B.: A Growing Neural Gas Network Learns Topologies. In: Advances in Neural Information Processing Systems, 7th edn., pp. 625–632. MIT Press, Redmond, Washington (1995)
7. Kohonen, T.: Self-Organizing Maps. Springer Series in Information Sciences, vol. 30. Springer, Berlin, Heidelberg, New York (2001)
8. Cormen, T.H., Leiserson, C.E., Rivest, R.L.: Introduction to Algorithms, pp. 558–565. MIT Press and McGraw-Hill, Cambridge (1990)
9. Tomasi, C., Kanade, T.: Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132 (1991)
10. Santini, S., Jain, R.: Similarity Measures. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(9), 871–883 (1999)