Figure 11.4. Grotesque medieval monsters composed of human body parts are an example of how the human perception of partial similarity can be misleading (woodcuts from folio XIIr of Hartman Schedel's 1493 *Liber Chronicarum* [337], Morse Library, Beloit College. Reproduced by courtesy of Constantine T. Hadavas).

Note that the object Y is used entirely, whereas parts are cropped only from X; this makes the distance d_{OPS} non-symmetric. Latecki *et al.* give an analogy with text search according to keywords: the large object X can be thought of as text and a small query object Y as a keyword. The optimal partial similarity approach relies on a tacit assumption that the query object Y is carefully selected in order to be sufficiently representative, very much like text search is sensitive to the selection of keywords.

However, in many situations, the knowledge of parts similarity does not allow us to infer information about the similarity of the whole objects – we may be comparing small parts that happen to be similar, yet, belong to objects that are completely different. For example, a leg is supposed to be a representative part according to which we would recognize an object as a human body; this is a prior information we have acquired during our lives and use in our judgment of similarity. Yet, applied to grotesque medieval monsters depicted in Figure 11.4, this judgment could be wrong: the shapes, though including legs, are not of human beings. Normally, one would label such objects as "weird" or "strange," implying that they look differently from what we are used to.[5]

11.2 Paretian approach to partial similarity

The above example teaches us that often two dissimilar shapes have many common parts, yet, using these parts to conclude that the shapes are similar would be wrong. The existence of common similar parts appears to be insufficient *per se*: such a criterion does not describe how *significant* such parts are. It is clear that some parts carry more information that allows us to recognize the entire object, hence they are more significant than other parts.

The most straightforward way to quantitatively define significance is by making it proportional to the "size" of the parts: the larger is the part, the more it is significant. For the following formulation, it is convenient to define *partiality* $\lambda(X', Y')$, which represents how small X' and Y' are compared with the entire surfaces X and Y (the larger is the partiality, the smaller are the parts). In shape comparison problems, it is natural to use

$$\begin{aligned}\lambda(X', Y') &= \mu_X(X'^c) + \mu_Y(Y'^c) \\ &= (\mu_X(X) + \mu_Y(Y)) - (\mu_X(X') + \mu_Y(Y'))\end{aligned} \tag{11.1}$$

as the partiality, where

$$\mu_X(X') = \int_{X'} da \tag{11.2}$$

is the measure of area derived from the Riemannian structure of X. The partiality can also be interpreted as a measure of "reliability" of our judgment of similarity: if there exist two parts of the objects that are similar, but these parts are small, our conclusion about the similarity of the entire objects is unreliable.

Following this logic, in order for two objects to be partially similar, they must have *significant similar parts*. The computation of partial similarity between X and Y can be therefore formulated as a problem of finding a pair of parts (X', Y') with minimum dissimilarity $d_{\mathrm{F}}(X', Y')$ and minimum partiality $\lambda(X', Y')$. More formally, we define a *multicriterion optimization problem*,[6] in which a *vector objective* function $\Phi(X', Y') = (\lambda(X', Y'), d_{\mathrm{F}}(X', Y'))$ is minimized over $\Omega = \Sigma_X \times \Sigma_Y$. Because we optimize over all the possible combinations of parts, the headache of finding a meaningful shape partition is avoided – we obtain it as a by-product of our solution, like in the method by Latecki *et al.* [241].

It is important to understand that partiality and dissimilarity are competing, such that no solution simultaneously optimal for both can be found, unless X and Y are fully similar. One consequence is that the notion of optimality used in traditional scalar optimization must be replaced by a new one, adapted to the multicriterion problem. Recall that for a scalar objective function $f : \mathbb{R}^N \to \mathbb{R}$, we define a global minimizer as a point x^*, for which the value of the objective is "the best," or, said differently, there does not exist another x such that $f(x^*) > f(x)$. In the vector case, we cannot straightforwardly apply this definition, as there does not exist a *total order* relation between vectors: we cannot say, for example, whether $(0.5, 1)$ is better than $(1, 0.5)$ or vice versa. At the same time, there is no doubt that $(0.5, 0.5)$ is better than $(1, 1)$, because it has both criteria smaller. We can therefore define *partial order* between vectors, saying that $(\lambda_1, \epsilon_1) < (\lambda_2, \epsilon_2)$ if $\lambda_1 < \lambda_2$ and $\epsilon_1 < \epsilon_2$ simultaneously.

Using this relation, we define a minimizer of our vector objective Φ as a pair (X^*, Y^*), such that there is no other pair $(X', Y') \in \Omega$ for which $\Phi(X^*, Y^*) > \Phi(X', Y')$, where the inequality is understood in the vector sense. Such a point is called *Pareto optimal*,[7] after the Italian economist Vilfredo Pareto

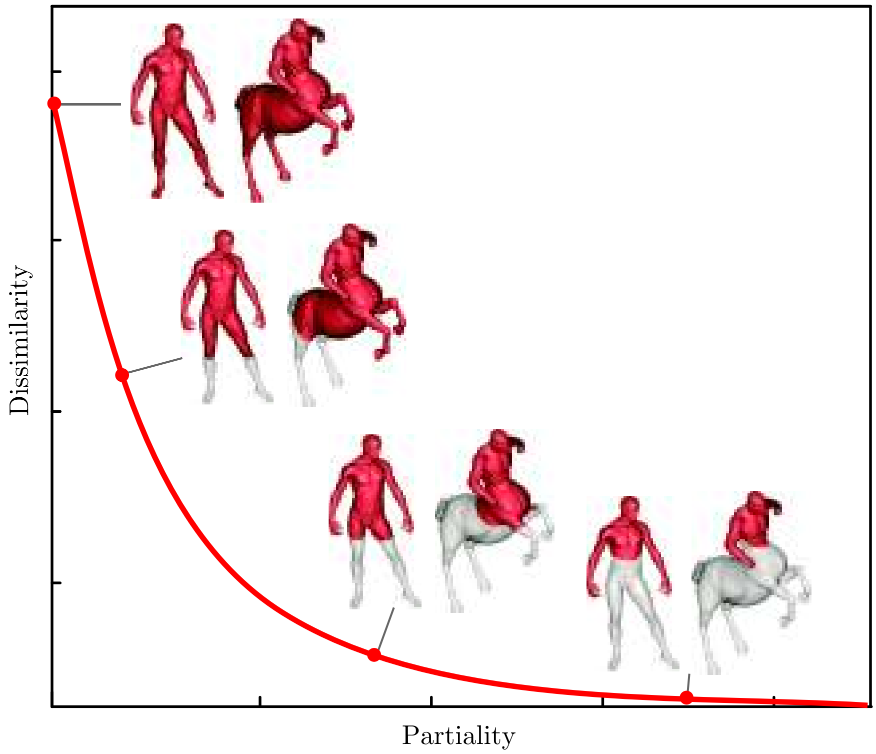

Figure 11.5. Illustration of the multicriterion optimization problem used for the computation of partial similarity. The Pareto frontier is shown as a curve.

(1842–1923), who first introduced this notion [307]. An intuitive explanation of Pareto optimality is that no criterion can be improved without compromising the other: if we try to reduce the dissimilarity, we necessarily increase the partiality, and vice versa. Note that Pareto optimum is not unique. If we denote by $\Omega^* \subseteq \Omega$ the set of all pairs (X^*, Y^*) satisfying the above condition, the set of the corresponding criteria values $\Phi(\Omega^*)$ can be represented as a planar curve referred to as the *Pareto frontier* (see Figure 11.5).

The notion of Pareto optimality brings us to a somewhat unorthodox concept of partial similarity: the *entire* Pareto frontier is used as generalized, *set-valued distance*. We denote $d_{\mathrm{P}}(X, Y) = \Phi(\Omega^*)$ and call it the *Pareto distance*. The dissimilarity value at the point $\lambda = 0$ on $d_{\mathrm{P}}(X, Y)$ coincides with the value of the full similarity $d_{\mathrm{F}}(X, Y)$. Consequently, the information contained in the set-valued Pareto distance is a superset of that contained in the traditional full similarity. The following example accentuates this difference.

Example 11.1 (set-valued distances). Consider the shapes depicted in Figure 11.6: a man, a spear-bearer, and a centaur. We would like to find their partial intrinsic similarity. The Gromov-Hausdorff distance (acting as d_{F} in this example and corresponding with the point with $\lambda = 0$ on the Pareto frontier) between the man and the spear-bearer is large, due to the distortion caused when trying to embed the spear into the spear-less human shape. Similarly, the Gromov-Hausdorff distance between the man and the centaur is large due to the distortion caused by the bottom part of the horse body, which has approximately the same diameter as the spear. Hence, from the point of view of global intrinsic similarity, the man is approximately as dissimilar to the spear-bearer as it is dissimilar to the centaur. This is the only information we can infer from the scalar-valued distance.

However, if we consider the entire Pareto frontier, we see that the curve representing the set-valued distance between the man and the spear-bearer decays much faster compared with the one representing the distance between the man and the centaur. The reason is that in order to make a spear-bearer similar to a man, we have to remove only a small part (the spear), whereas in order to make a centaur similar to a man, we have to remove large parts (the horse body from the centaur and the legs from the man). Thus, from the set-valued distances, we can infer that the man is more similar to the spear-bearer than to the centaur, which corresponds with our intuition.

11.3 Scalar partial similarity

Although containing more information than scalar-valued distances, the inconvenience of Pareto distances is that they are not always mutually comparable. This problem stems from the absence of a total order relation between vectors: we can say that $d_{\mathrm{P}}(X,Y) < d_{\mathrm{P}}(X,Z)$ (meaning that X and Z are more partially dissimilar than X and Y) only if the curve $d_{\mathrm{P}}(X,Y)$ is entirely below $d_{\mathrm{P}}(X,Z)$ (in Example 11.1, the man–spear-bearer Pareto distance was below the man–centaur distance, therefore, we could say that a man is more similar to a spear-bearer than to a centaur). In order to define a total order between partial similarities, we have to convert our set-valued Pareto distance into a scalar-valued one. We refer to such a "scalarized" partial dissimilarity criterion as a *scalar partial distance* and denote it by $d_{\mathrm{SP}}(X,Y)$.

Straightforwardly, a scalar partial distance can be obtained by selecting a single point on the Pareto frontier. For example, with a fixed partiality λ_0, we can set a minimum threshold on the area of the parts we compare. This way, small values of λ_0 will make our criterion more reliable but at the same time, more restrictive: in order to say that two shapes are partially similar, they must have larger parts. The extreme case of $\lambda_0 = 0$ brings us back to the full similarity.

The scalar partial distance in this case is computed as the solution to a constrained optimization problem with a scalar valued objective,

$$d_{\mathrm{SP}}(X,Y) = \min_{(X',Y') \in \Omega} d_{\mathrm{F}}(X',Y') \quad \text{s.t.} \quad \lambda(X',Y') \le \lambda_0,$$