An Image Is Worth More Than a Thousand Words - Is It Really?

Who hasn't heard or even used that famous saying, credited to a Chinese philosopher. It makes eminent sense, doesn't it? We all intuitively agree how difficult it is to adequately describe images. Images are so rich in information, they invoke different sensations that are indeed hard to capture in words. The challenge is perhaps not so much in the number of words but in finding the proper words to describe an image in a meaningful way.

Vision is our most impressive and complex sense. We see and analyse our environment continually, without conscious effort. Imagine you are about to cross a busy street. Reflected light from the street scene is incident on the retinaâ \in^{TM} s photoreceptors that transform the light intensities into nerve signals. Visual information is processed at various stages along the optical pathway, at the visual cortex and in higher cortical brain areas. The net result is a visual perception about the world. What we â \in^{TM} is not the image formed on the retina but its interpretation. What we store is a â \in^{TM} or visual perception, not the retinal pixel image. Perception causes us to plan and to properly respond to the environment.

Ever since computers became available, researchers in artificial intelligence and cognitive science have tried to mimic the process of seeing by machines, motivated by a desire to understand the human visual system. Researchers in computer vision and digital photogrammetry pursue more mundane goals, for example, navigating a robot through a cluttered environment, or finding objects like buildings, roads, and trees. Since humans are so incredibly adept at these visual tasks, one is easily lured into believing that a machine can do the same. After all, computers are so much faster and digital cameras render images of superior quality. Evidently the technology is here but not a detailed enough understanding for we have yet to see a machine that can reconstruct surfaces, recognise buildings, or detect changes, let alone produce maps automatically.

A major misconception is to confuse the image you obtain with a digital camera with the mental image we perceive from the same scene. Consider the mental image as a highly symbolic description of a scene and you realise that this is precisely what computer vision sets out to do. Now, if we want to compare apples with apples then we should compare the digital image with the retinal †pixel†image. And here is the fallacy: we do not have access to the retinal image. Imagine for a moment you were looking at the numbers of the huge matrix that represents a digital image and you would realise the enormous challenge which the computer faces to make sense out of all these numbers.

If a machine vision system is able to generate a meaningful description of a scene, however short and incomplete that might be, then I think that this is worth more than the image.

https://www.gim-international.com/content/article/an-image-is-worth-more-than-a-thousand-words-is-it-really