

# A framework for automatic progress assessment on construction sites using computer vision

Emanuele Trucco<sup>1</sup> and Ammar P. Kaka<sup>2</sup>

**ABSTRACT** | Site managers spend a significant amount of time measuring, recording and analysing progress on site. Information is continuously captured and exchanged between individuals, teams and firms for valuation purposes, productivity measurement, schedule or quality control. Images of construction sites provide an enormous amount of information about progress. Computer vision offers great potential for analysing this information automatically. This paper brings two contributions in this direction. First, we propose a computational framework to capture and measure construction progress automatically from video images taken on site. The novelty lies in the introduction of automatic assessments usable in IT systems while construction work is still in progress; normally, only information developed at the design stage is used. Second, we propose an algorithm for recognising objects and structures in unconstrained outdoors site imagery. This can be used to determine the location of a part of building or structure within a site, and for non-textual indexing and image selection in large image records from construction projects. We report and discuss the results of experiments with real images from two different building sites.

## 1 Introduction

### 1.1 Topic and Motivation

This paper reports on an algorithm for the automatic location of objects in unconstrained images of building sites, and sketches a computational framework to capture and measure construction progress automatically from video images taken on site. The measurement of work on site currently takes place using traditional building surveying techniques and/or visual inspections. The use of progress measurement is not limited to calculating interim payments but is essential for many business and project management

processes including schedule control, cost control, financial reporting, claims, etc. For these processes to be reliable and useful, measurements need to be accurate; the time (and therefore cost) needed to accurately measure work and in detail is often substantial. Conventional monthly measurements are not frequent enough and incorporate judgement, often guesswork, and shortcuts. This is likely to render them inaccurate, giving rise to under or over measurement and the likelihood of inaccurate cost/progress control data.

In the current era of computer integrated construction, digitally capturing and analysing progress will not only facilitate faster data generation and communication but

1. School of Engineering and Physical Sciences Heriot Watt University Riccarton, Edinburgh EH14 4AS, UK

2. School of the Built Environment Heriot Watt University Riccarton, Edinburgh EH14 4AS, UK

will also form an integral part of the overall information system connecting and supporting the sites of the future (Bjork, 1999). The recent developments in IT have stimulated substantial research and application in Construction and Construction Management. Information, in the form of drawings and database models generated at the design stage are currently being used in and integrated with many other disciplines (e.g. quantity surveying, costing, planning). Virtual reality (Powell 1995, Rezgui et al. 1996, Schmitt 1993, Aouad et al. 1998, McKinney and Fischer 1998) has also been found to be a useful tool for enhancing the presentation of buildings at the design stage. Current research on Facilities Management, an area that is growing rapidly, has also benefited from CAD and other modelling techniques (Al-Hajj and Aouad, G 1999).

The common factor in all these developments is that they make use of *information developed at the design stage*. Whilst these developments have and will contribute significantly to the efficiency of the construction industry, full Computer Integrated Construction will not be achieved unless the IT systems used during the construction process are able to digitally capture *data on site progress acquired during construction*. The automatic generation of such data from unconstrained site images is the object of our work.

This paper reports the first phase of a research project addressing the automation of the process of measuring work in progress on construction sites and integrating that with design and planning. A wider aim is the promotion of better access to and integration of site data within future communication systems needed for a truly global construction industry exploiting state-of-the-art information technology (Bjork, 1999). We describe the overall framework proposed for the assessment of work in progress from site images. We then describe the first module developed, a package capable of recognising parts of a building site and objects therein in real, unconstrained site imagery. The system is based on an iconic recognition system recently developed by one of the authors (Odone et al.

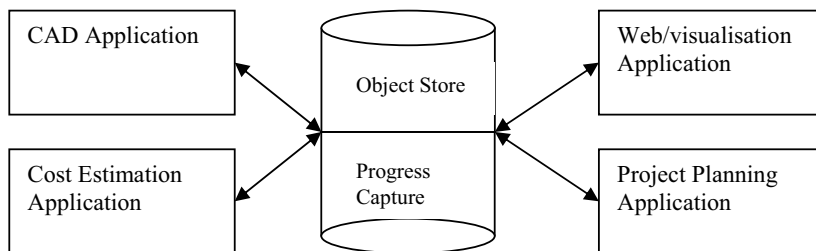
2001, Odone et al 2001a). We report experiments with real site imagery, and discuss the promise and current limits of the approach.

## 1.2 Computer Integrated Construction

The importance of integrated systems is reflected by the growing number of funded research initiatives in the US and Europe, including ICON (Aouad et al, 1994), COMMIT (Rezgui, 1996), CIMSTEEL (Watson, 1995), ATLAS (Nedreven, 1994), MOB (OTH, 1994), RATAS (Bjork, 1989), IRMA (Luiten, 1993), Kartam (1994), and other projects such as CONCUR and PROCURE. All these studies address the central issue of project based information exchange between building professionals. All these projects preach the use of a common building product model as a standard description of building objects, which can be subscribed to by all participants of the construction process. In addition, these projects highlight the importance of linking design, planning and cost control applications thus providing significant opportunities for the automation of these processes and functions. The proposed model is an integral part of the whole idea of Computer Integrated Construction (CIC).

The overall system proposed in this research effort will facilitate the automation and integration of work measurement with valuations, costing and plan updates. The system will consist of a digital camcorder, Progress Capture module (computer vision software) and an integrated object oriented database which will support four application modules: CAD Application, Cost Estimation Application, Project Planning Application and Web/Visualisation Application (see Figure 1).

The system will operate by using a well-defined standard procedure for 3D-image capture of the building being constructed. The digital data is then communicated from site to head office where a computer vision system will automatically assess the



**Figure 1.** Introducing Automated progress capture to Construction Integrated Systems

state of progress of the different structures and objects. This will be done by comparing site images with rendered images of the design objects/structures. The project planning application module will assist in the determination of what objects are expected to feature on site at any given point in time.

Once this is done, the results will then be communicated to project planning application in order to assign the percentage completion of each activity. This in turn will be used to do the following:

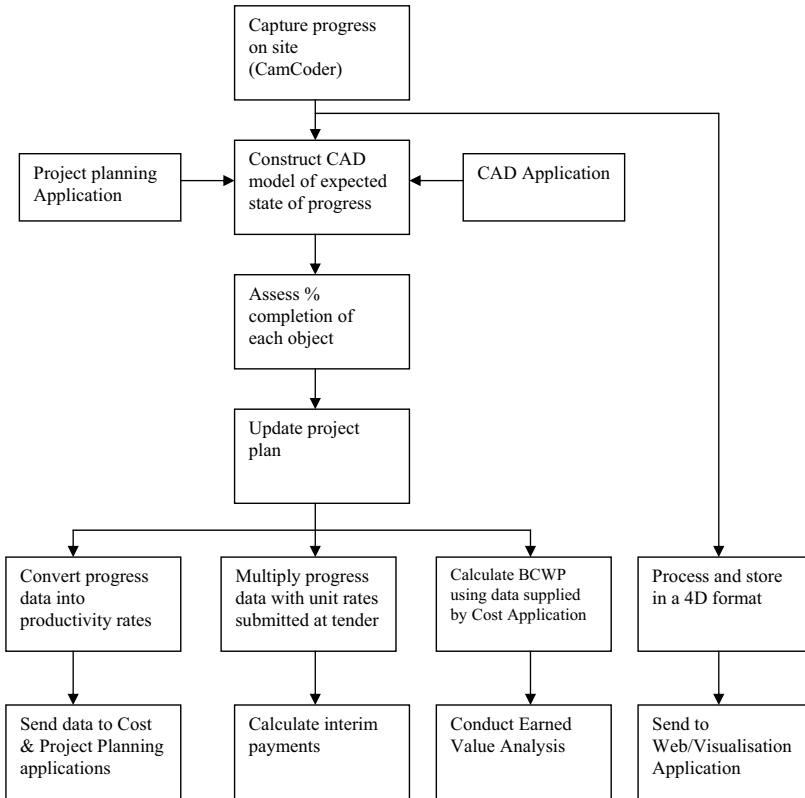
1. Earned Value Analysis: Percentage completion of each activity are multiplied by the corresponding total quantities and their estimated unit cost to calculate the Budgeted Cost of Work Performed (BCWP). This is then compared with the Actual Cost of Work Performed to assess cost variances.
2. Calculate Interim payments: Percentages completions of activities are multiplied by unit rates submitted at tender stage.
3. Produce productivity data: Percentages completion of activities are multiplied by the corresponding total quantities and divided by time taken to achieve level of progress will produce the rate of progress per week, day or even hour. This data is useful for future planning and cost estimation purposes.
4. Process and store images: Images of site in progress will be archived and accessed via a web browser. Clients and indeed any member of the supply chain would be able to inspect the site remotely which will lead to better site management and planning, improve safety and security, better quality control

(remote checks would increase frequency of checks hence avoid potential problems), claims resolution, etc.

The mechanism behind the proposed system is illustrated in Figure 2.

### 1.3 Computer Vision and its Potential for Construction Industry

*Computer vision* (Trucco and Verri 1998; Kasturi et al. 1995) is a vast field of research with great potential for the construction industry. Computer vision differs from photogrammetry, popular in the construction section (Knight and Kaka 1998, Streilein 1996, Suveg and Vosselman 2001), in that it concerns not only measurements, but also recognition, classification, indexing, motion analysis and much more. Here, we concentrate on 3-D *object recognition*, a specific area of computer vision. For the purposes of our discussion, we define 3-D object recognition as the problem of *detecting and possibly locating a 3-D given object, or model, in a video or set of images*. We consider two classes of solutions, differing by the nature of the model (Trucco and Verri 1998): algorithms using *CAD-like models* and algorithms using *iconic models*. **CAD-like models.** CAD-like models are *CAD-like descriptions of 3-D objects*, specifying the position and attributes of lines and surfaces in a world reference frame (Trucco and Verri 1998). Comparing such models to an image implies converting both model and image to a common, intermediate representation (Lowe 1993, Alter and Grimson 1993, Wunsch and Hirzinger



**Figure 2.** The mechanism associated with integrating computer vision with Project management functions

1997, Rosin 1999). Typically, this representation is expressed in terms of detectable image features (e.g., lines, curves, special surfaces). By modelling the camera imaging process geometrically, one can estimate the relative position of camera and scene (*3-D matching*), and predict the appearance of the features belonging to a 3-D model in a given position relative to the camera (*backprojection*). A 3-D model is located successfully in space if a camera position can be found which generates the distribution of image features appearing in the input image. Importantly for the discussion in Section 2, we notice that backprojection allows one to measure the degree of mismatch between model and image, and therefore potential missing parts. This is at the basis of our vision for automatic progress assessment.

**Iconic models.** Only a subset of objects can be

modelled by manageable 3-D CAD-like models for recognition purposes. For instance, articulated and flexible objects like human bodies, or objects changing in time like buildings under construction, lead to intractably complex models, or models which would require continuous updating. In such cases one can turn to *iconic models*, whereby a 3-D object is represented by a *set of significant views* (Murase and Nayar 1995, Leonardis and Bischof 2000). The model views are acquired from a number of viewpoints surrounding the target object, possibly in different illumination conditions. In comparison with CAD-like models, iconic models lack compactness (one may need many views per object), but can be compared directly with input images. This simplifies the matching algorithm, as there is no need for feature extraction.

The general framework we suggest draws from both

classes of algorithms; Section 2 explains how. The experimental work we report in Sections 3 and 4 focuses on iconic recognition.

## 2 A computational framework for site progress assessment

The computational framework we propose for automatic assessment of work in progress on building sites is illustrated in 3.

The overall objective is to analyse automatically video material acquired during site surveys to compute the following elements.

1. The *location* of a site view or building component (e.g., a wall, a column) within images of the site acquired later. This is vital to compare images of the same part of a site, taken at different times. Notice that this operation amounts to locating the view in a site, which in turn may be used to reference images automatically to a digital map of the site.
2. The *amount of progress* of a building, or of specific components thereof, compared with previous, recent surveys. This can be estimated by identifying instances of the same target (object, structure) in different images, and comparing the various instances after suitable geometric normalisations (as the size of imaged objects may

vary in different images).

3. The set of *images containing a target object*. This is, in essence, an image retrieval operation performed on an image database. The images of the target object, in this sense, constitute the query. The algorithm presented below addresses this problem.

Figure 3 illustrates the general operational idea of the framework. Image data from a site are acquired in batches over time. Such data can be a collection of still images or digital video footage, acquired on site by operators unskilled in computer vision; in other words, we do not impose constraints on scene, imaging conditions, or imaging procedures. Notice that the input imagery could equally be a pre-stored database, e.g., the photographic history of a construction site; this is relevant for point 3, as discussed below.

Each new image or video segment is automatically referenced, i.e., located in the site (point 1 above). The image database can therefore be automatically organised not only by time, but also by location. As this process is based on the detection of the same site objects, all images of a same objects (e.g., acquired at different times) can be labelled as such for object-based retrieval purposes.

Experience suggests that *quantitative* progress is best assessed against a common, external CAD model of a construction element. Several techniques for model-

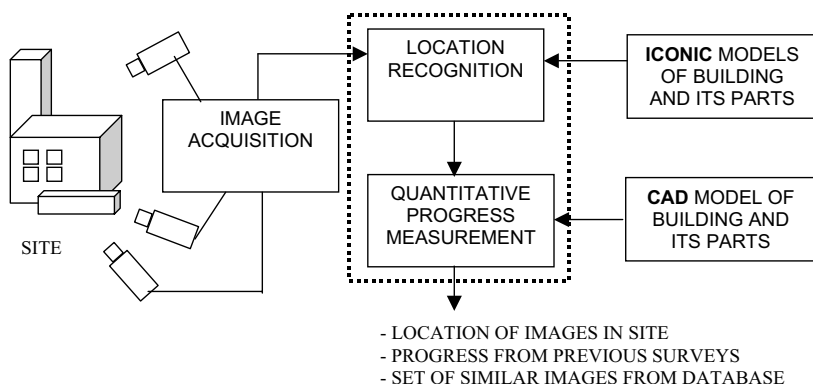


Figure 3. Computational framework for the automatic progress assessment

based recognition of 3-D objects in images exist (Trucco and Verri 1998). Models can be CAD or CAD-like (subsets of full CAD models), or iconic (Figure 3). The latter are suitable for identification and detection, the former for quantitative measurements, and feature therefore in several photogrammetric techniques. We now discuss the three objectives above, in the context of the modules shown in Figure 3.

## 2.1 Recognising Site Views and Components

Recognising a certain view of a site cannot be done by comparing the image with a CAD model of the building, as, in general, only a fraction of the structures expected in the completed building will be present, their real appearance cannot be predicted accurately from the CAD model, and a number of extra elements are likely to appear as well (e.g., people, tools, building materials). Notice that this is not to say that CAD models cannot be used *in general*, only that complex recognition problems with evolving objects call for an alternative approach. These difficulties suggest an *iconic approach* (Odone et al 2001 and 2001a, Murase and Nayar 1995, Rao and Ballard 1995), whereby the input view is compared with pre-stored, *real* images of the site, acquired of necessity previously, and representing significant parts of the site itself. These images are called the *model views*. “Significant” means, in this context, that a sufficient number of structural elements are visible to identify the site location univocally. Of course, the difficulty is that the model views are acquired some time before the input views, so that even images acquired from the same viewpoint will be different due to site progress, occasional disturbing elements like cars, people, etc, and changes in illumination conditions. A working solution to these problems is reported in our experimental work (Sections 3 and 4).

## 2.2 Measuring Progress of Structures and Components

Once the location of a structure is established, it is possible to index the CAD model of the site to extract

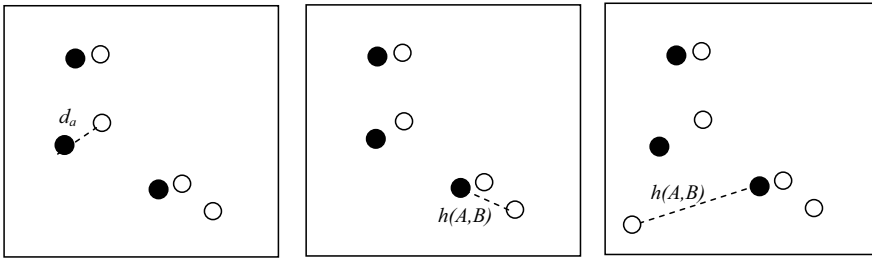
the expected, complete structure. It is then possible to apply 3-D matching methods (Lowe 1993, Alter and Grimson 1993, Wunsch and Hirzinger 1997) to determine the position of the partially built structure within the CAD model. This, in turn, enables us to estimate how much has been built, that is, what progress has been achieved from the last survey. We intend to address this process in future work, and point the reader to the references above.

## 2.3 Finding Images Containing a Target

Another important application of recognition is automatic image selection in an image database. This could be a large set of images from the history of a construction project, from a portfolio of company projects, or video footage from various construction sites operated by a multinational company. Iconic recognition can be used to detect images similar to a prototype image within the database. In practice, this relies on the very same techniques used for view and component location. We describe our solution in Section 4. Notice that this application is closely related to the fast-growing field of nontextual indexing and automatic video annotation and editing (Ahanger and Little 1996, Brunelli et al. 1999, Lienhart 2000).

## 3 Iconic recognition with unconstrained site images

This section describes our algorithm for automatic, iconic object recognition. The algorithm measures the similarity between two images using the notion of Hausdorff distance. Given two images, the method checks pixelwise if the grey values of one are contained in an appropriate interval around the corresponding grey values of the other. Under certain assumptions, this provides a tight bound on the directed Hausdorff distance of the two grey-level surfaces. The proposed technique can be seen as an extension to the grey level case of a matching method developed for the binary case by (Huttenlocher et al



**Figure 4.** Illustration of directed Hausdorff distance. Set B is formed by black circles, set A by white circles. First, compute distance  $d_a$  between each point in A and set B (left); the maximum of these distances is the directed Hausdorff distance (centre). Notice that  $h(B,A) \ll h(A,B)$ , and that a single outlier in A skews the distance (right)

1992). The method fits naturally an implementation based on comparison of data structures and requiring no numerical computations whatsoever. Moreover, it can match images successfully in the presence of occlusions and changes, an important feature for our application.

### 3.1 Hausdorff Distances and their Properties

Given two finite point sets  $A$  and  $B$  in the plane, the *directed Hausdorff distance*,  $h(A,B)$ , is measured in two steps. For a fixed point  $a$  of  $A$ , the first step computes the distance of  $a$  from each point  $b$  of  $B$ , and selects the distance  $d_a$  between  $a$  and the closest point of  $B$ . The second step takes the maximum of  $d_a$  for all  $a$  of  $A$ ,  $h(A,B)$ . Formally this can be written as

$$h(A,B) = \max_{a \in A} \{ \min_{b \in B} \{ \| a - b \| \} \}.$$

Note that order matters, as  $h(A,B) \neq h(B,A)$ . For example, if all points of  $B$  are “close” to some points of  $A$ , then  $h(B,A) \ll h(A,B)$  (Figure 4). We notice that the directed Hausdorff distance, which is not symmetric and thus not a distance mathematically, measures the degree of mismatch between  $A$  and  $B$ . Symmetry can be restored to obtain a proper mathematical distance by taking the maximum between  $h(A,B)$  and  $h(B,A)$ . This brings us to the definition of Hausdorff distance, that is,

$$H(A,B) = \max \{ h(A,B), h(B,A) \}.$$

As a distance,  $H(A,B)$  is zero if and only if  $A = B$ . Instead, the directed Hausdorff distance is zero if and only if  $A$  is a subset of  $B$ . A useful property of both measures for our purposes is their ability to *measure the distance between two sets with different number of points*. A less desirable property is their sensitivity to outliers: a few points far out skew the distance for many, closely arranged points. The next section discusses how this can be countered effectively.

An intuitive illustration of the Hausdorff measures is to think in terms of set inclusion. Let  $B_\rho$  be the set obtained by replacing each point of  $B$  with a disk of radius  $\rho$ , and taking the union of all of these disks. Effectively,  $B_\rho$  is obtained by dilating  $B$  by  $\rho$ . Then, the directed Hausdorff distance,  $h(A,B)$ , is not greater than  $\rho$  if and only if  $A \subseteq B_\rho$ . This follows easily from the fact that, in order for every point of  $A$  to be within distance  $\rho$  from some points of  $B$ ,  $A$  must be obtained in  $B_\rho$ . This geometric interpretation suggests how to counter the effect of outliers. Let  $A'$  the subset of the points of  $A$  contained in  $B_\rho$ . Assume that, for a given value of  $\rho$ ,  $A'$  is nearly equal to  $A$ . The directed Hausdorff distance between  $A$  and  $B$  can be distorted by the few points not in  $A'$ , but  $h(A', B)$  is still not greater than  $\rho$ , which means that the potential outliers are defined and identified in one step. This fact is exploited in our algorithm, explained next.

### 3.2 The Iconic Matching Algorithm

How can the above be used for finding a *model* image  $M$  (the reference image) in another image  $I$ ? The idea is to compare all possible translated versions  $M_t$  of  $M$  with  $I$ , and select the translation  $t'$  that maximises a similarity score based on the Hausdorff distance between the sets of the grey levels in  $M$  and  $I$ . The discussion on outliers above suggests that we can build a *tolerance* ( $\rho$ ) in the algorithm, allowing the system to accept a certain amount of difference between model and image (e.g., occlusion, rotation, viewpoint or illumination changes).

The algorithm consists of four steps. The input is formed by a model image,  $M$ , and an image  $I$  in which the model must be found. The output is a translation,  $\tau$ , which brings  $M$  onto the part of  $I$  most similar to  $M$  (i.e., locates  $M$  in  $I$ ).

*Step 1.* Expand the model image  $M$  into a 3D binary matrix,  $B_M$ , the dimensions of which are pixel coordinates  $i$  and  $j$  and grey values  $g$ . By definition,  $B_M(i,j,g)$  is 1 if  $M(i,j) = g$ , and 0 elsewhere. Build a 3D binary matrix  $B_I$  from the image  $I$  in the same way.

*Step 2.* Dilate the matrix  $B_I$  by growing its nonzero entries by fixed amounts in all three dimensions. Let  $B_G$  be the resulting 3D binary matrix. Typical dilations with medium-resolution, 8-bit monochrome images are 3 to 7 pixels, but values depend on the images being processed. This point is further discussed in the experiments section.

*Step 3.* Compute the size of the intersection between all possible translated versions of  $B_M$  and  $B_G$ , within a given search region. Call this number  $S(t)$ . Notice computing  $S(t)$  over a whole search region generate a function, which we regard as a *similarity surface*.

*Step 4.* Return the translation  $\tau$  for which  $S(\tau) \geq (S(t))$  for any  $t$ , that is, to the absolute maximum of  $S(t)$ , but only if the maximum is above a threshold (which can

be chosen in accordance with the level of similarity required).

We can now state how exactly the algorithm is based on the directed Hausdorff distance. If the dilation of the matrix  $B_I$  is isotropic in an appropriate metric, and  $S(\tau)$  takes on the maximum value possible, then the directed Hausdorff distance  $h(B_M, B_G)$  between  $B_M$  and  $B_G$  is not greater than the dilation. In other words, the maximum of  $S(t)$  signals that the difference (distance) of image and model within the tolerance (dilation) specified. This means that a successful match has been found.

With an appropriate choice of the data structures, the algorithm is reduced to a set of logical AND operations between entries of  $B_M$  and  $B_G$ , and no numerical computations are required. Implementations can therefore achieve high speed, a definite advantage for processing large quantities of images in a database.

Notice also that, as in any numerical optimisation, an answer is *always* produced, i.e., a maximum of  $S(t)$  is found even when the model is *not* present in the image. Two factors allow us to decide whether a given maximum suggests strongly the presence of the model: *the magnitude of the maximum and the variance of the surface  $S(t)$*  around the maximum. The former criterion allows us to compare answers from different input images  $I$ ; the latter is used within a single input image.

1. *Magnitude of the maximum.* Maximum values actually generated by the presence of the model are generally much higher than ones generated by any other objects. Intuitively, the closer the appearances of model and image, the higher the score  $S(t)$ . Clustering and allied statistical techniques (Lapin 1998) can be used to single out “significantly high” maxima in a set of maxima generated by comparing a model  $M$  with a set of input images; the results reported below were achieved using simply a fixed threshold on  $S(\tau)$ .



2. *Variance of similarity surface around the maximum.* If the model is actually present in the image, the surface  $S(t)$  has generally a pronounced peak around the correct location. Thus a narrow spread of surface values around the maximum suggests that the model has been detected. This property can obviously be measured by estimating the variance of  $S(t)$  around the maximum in the two axis directions.

## 4 Experimental results

### 4.1 Tests Description and Data

We summarise here results from two groups of tests conducted with real images acquired on different construction sites.

The first set of tests illustrates the system's capability to identify specific structures and components in site images. The images used were acquired by us in several visits to a building site in Bo'ness, Scotland, during the summer 2001, where 8 plots of small houses were being built. The images were taken during normal work, and no items or structure was changed in any way for the purposes of the tests. In all, 35, 48, 27 and 45 photographs were taken, respectively, during the four visits. The digital camera used was a Olympus Camedia C2500L and the images saved at 1712x1368 resolution.

The second set of tests illustrates the system's capability to identify a specific site view in a set of images of the same site. The images were taken of major single user office development in Edinburgh, Scotland. The purpose of the building was to provide a prestigious headquarters for a major financial institution as existing buildings owned by the organisation failed to adequately accommodate the growth in personnel and new technologies.

The enabling works for the construction commenced in November 1993 and was completed September 1996 at a cost of £60m. Spanning a gross floor area of 400,000ft<sup>2</sup> the development consists of two

independently functioning buildings connected by a shared reception. Both buildings are configured over seven floors and have an occupation capacity of 2500.

As part of the management contract procurement route, a requirement was made to capture images throughout the construction process. These primary purpose of these images were to show progress, but they also formed the legal documentation in the event of dispute. Using an independent photographer, images were taken at monthly intervals, over a thirty-six month period, using traditional photographic methods.

Four standard viewpoints were initially used to show the construction progress, however this was later reduced to two as the development grew. In addition to the viewpoint images, further images of the various site activities were also captured. Each month the at least 36 images, this number would vary dependant on the stage of construction providing a total of least 1200 images over the three year period.

From these images, the research team assessed those images deemed to be useful for the purpose of the research and converted them into a digital format. All tests were conducted on Pentium III PCs in the 550–800MHz range, under Linux and Windows. The processing speed, which depends on the size of the model image, varied from 2 frame per second (on the 500-MHz PC with model images of size 200x2200 pixel approximately) to 20 frames per second (on the 800-MHz PC with 60x60 model images approximately).

### 4.2 Detecting Small-Scale Components

The task here was to detect automatically the presence of a given construction element in unconstrained images of a site in progress. A single image of the construction element was available as model to the system. This experiment illustrates the ability of the

system to find images containing a specific element, as well as to locate that element within a larger image, with good resilience to disturbing factors like changes of viewpoint and position in the image, or the presence of spurious elements. We show results with two components, a partially built wall and a garage door.

*A garage door.* Figure 5 shows the model image (prototype) of a garage door, that is,  $M$  in our algorithm description (Section 3), and a selection of other images acquired on the same site at a different time. A white rectangle shows the location of the best match found by the system. Images appear by descending values of  $S(\tau)$ . We used dilations of 3 pixels for  $i$  and  $j$ , and 2 grey levels for  $g$  (see algorithm description, Step 2).

The right column visualises the similarity surface,  $S(t)$ , as a grey-level image (the brighter the higher the similarity score). The maximum position is highlighted. Surfaces were normalised for display purposes. A significant drop of similarity values was detected in this test below the value 2238, the lowest value at which the model was still actually present. It is clear that peaked surfaces are desirable for localization purposes. The actual variation of  $S(t)$  around the peak, and indeed the correct position of the peak, depend on the resemblance of model and image. If the model is excessively distorted geometrically (e.g., severely foreshortened shape) or photometrically (e.g., imaged in much brighter conditions) the algorithm may fail. Photometric distortion can be successfully combated by photometric normalisation (Trucco and Verri 1998). Shape distortion is addressed indirectly by keeping multiple images of the model, acquired from different views. This is typical of iconic models (Murase and Nayar 1995).

*A partially built wall.* Figure 6 shows the model image (prototype) of the wall and a selection of the other images acquired on site at a different time. As before, a white rectangle shows the location of the best match found by the system, and images appear by descending

values of  $S(\tau)$ . The images associated to the highest  $S(\tau)$  are the ones most probably containing the model. Dilation values were as before.

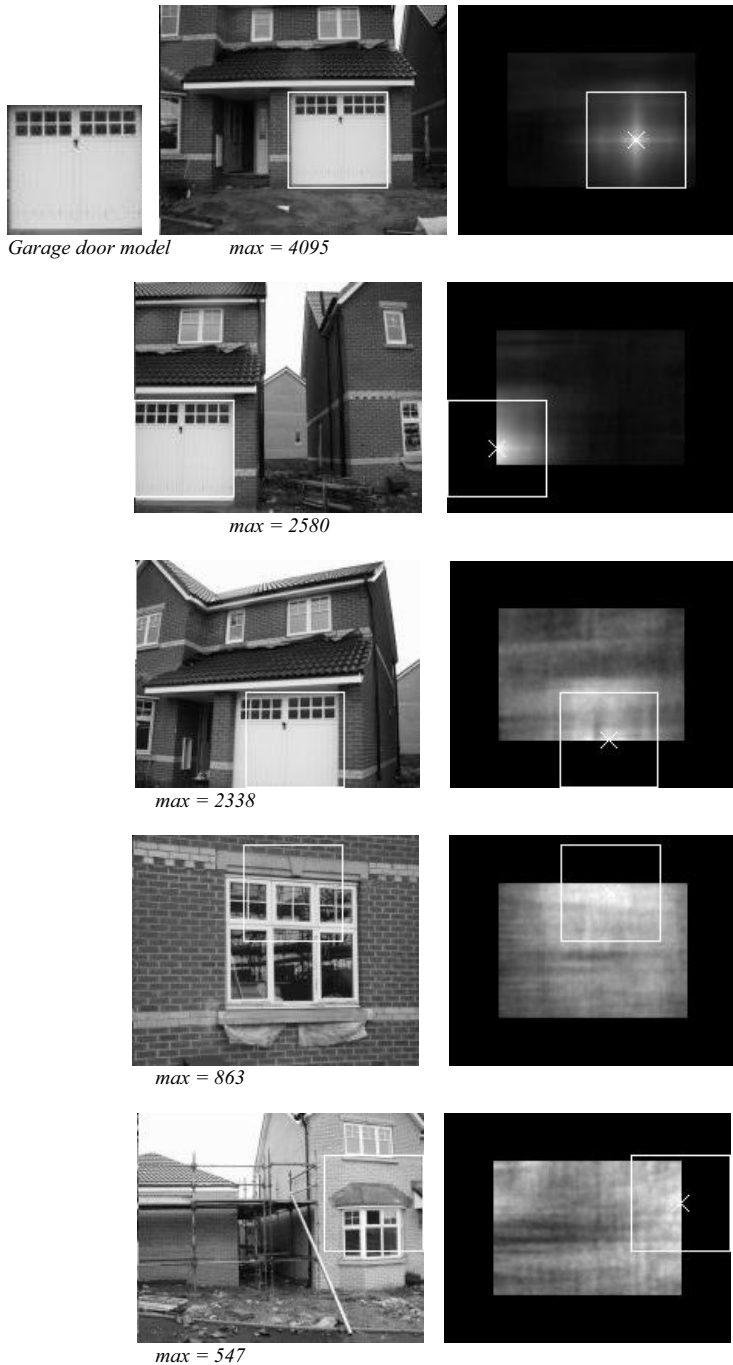
The right column visualises the similarity surface. Surfaces were normalised for display purposes. A significant drop of similarity maximum values was detected in this test below the value 1169, the smallest at which the model was actually present. Notice that the algorithm *always* returns a candidate match (for any image  $I$ ). The match represent an instance of the model is the score  $S(t)$  is sufficiently large. For practical purposes, “sufficiently” means 80% of the score achieved by matching the model with itself.

### 4.3 Recognizing Site Views

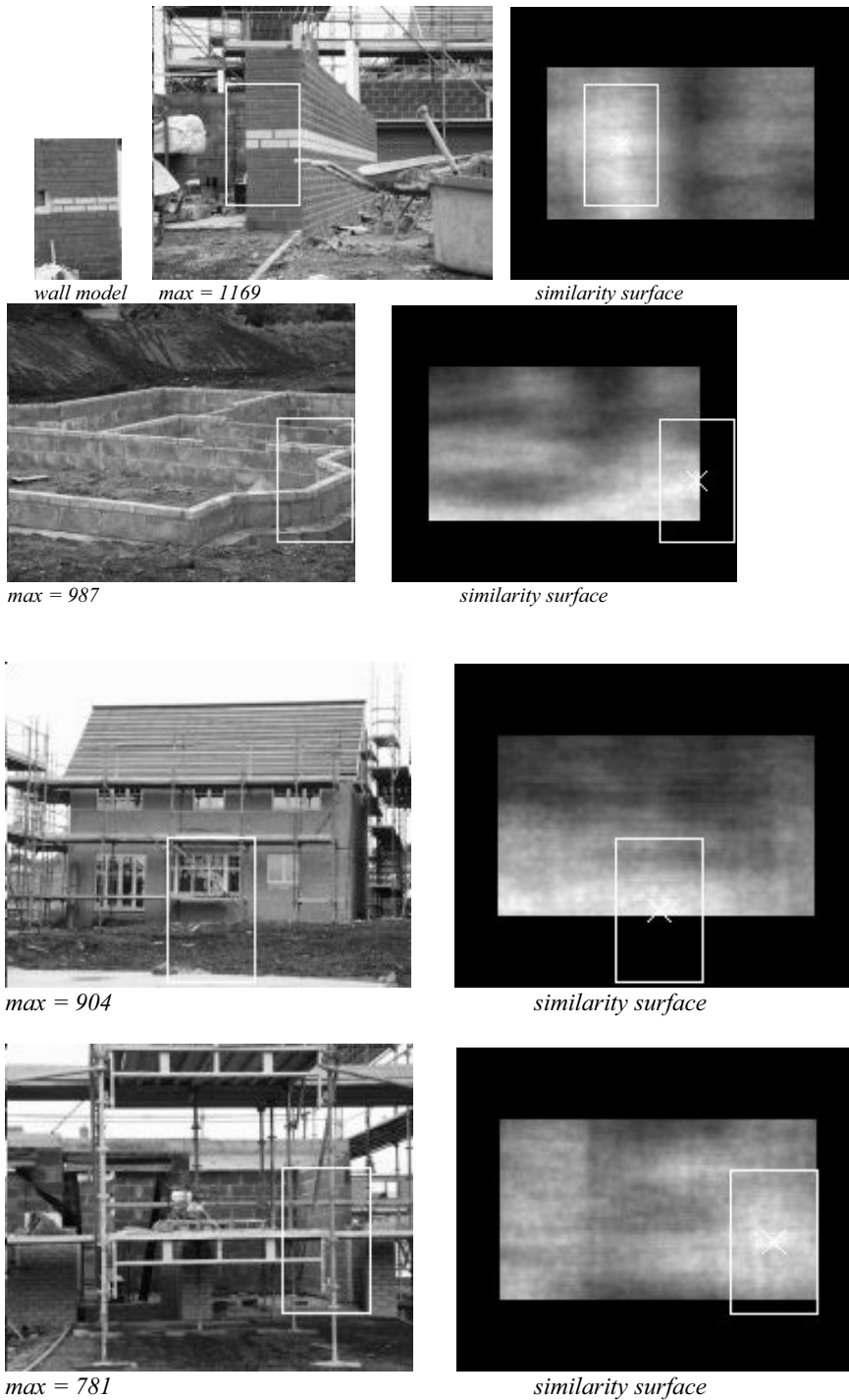
Given an image of a large building in progress acquired, the task here is to locate the same view in a set of images of the progressed site taken in later surveys. The test images are views of the Standard Life building site. Photographs were taken monthly by site personnel. No information were available on the photographic equipment. We scanned the photographic prints made available by Standard Life to obtain digital images of medium resolution (800x600 approximately).

We summarise results from two sets of tests: locating complete views of the building in progress, and locating a building part.

*Locating complete views.* Figure 7 shows location results with two models, before and after scaffolding were erected. Model and test images were acquired from approximately the same viewpoint. The algorithm can locate the model precisely in spite of changes of illumination conditions, object appearing (cars, people) and, to a smaller extent, size and viewpoint. The September 1995 – March 1996 match (bottom row) indicates that a view can be tracked in spite of considerable appearance changes like the removal of large parts of the scaffoldings. Dilation values were 4 for pixel co-ordinates and 3 for grey levels.



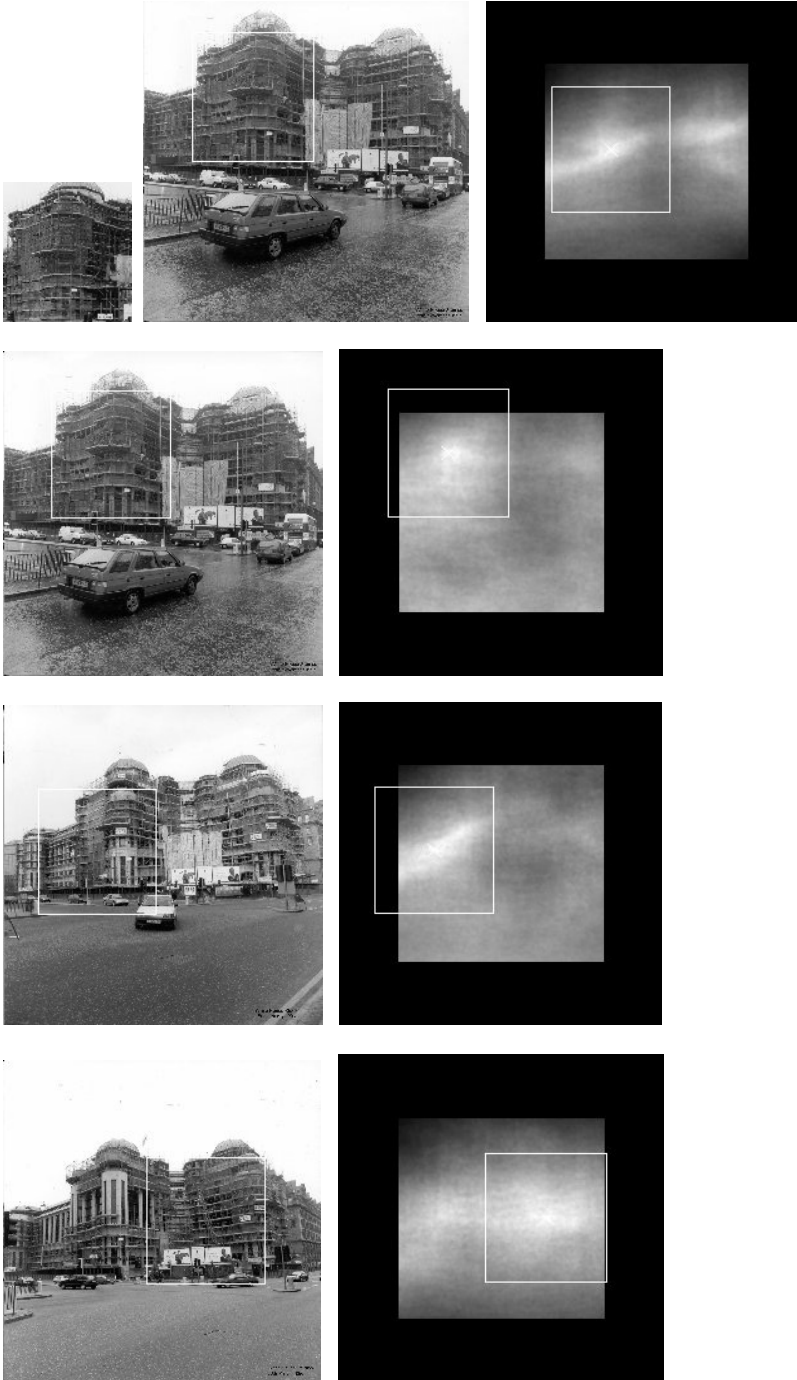
**Figure 5.** Garage door model (top left) and its computed location in a selection of input images (left column) with associated similarity surface (right column; shown as image, the brighter the larger the surface value), generated by the algorithm. The cross shows the position of the maximum. The black cornice is caused by the fact that the model must overlap the image completely, so its centre cannot reach the borders



**Figure 6.** Wall model (top left) and a selection of input images with located model and associated similarity surface generated by the algorithm (the brighter the higher the surface value)



**Figure 7.** Top left: model of the building in progress, March 1995. Top right: located model in same-viewpoint photograph taken in April 1995. Middle right: model of the building in progress, September 1995. Middle left: located model in same-viewpoint photograph taken in October 1995. Bottom: September model located in January and March 1996 photographs



**Figure 8.** Top row, left to right: model from September 1995, model located in October 1995 view, and associated similarity surface (maximum = 2905). Middle row: model located in November 1995 view and similarity surface (maximum = 2612). Third row: same with January 1996 view (maximum = 2473). Fourth row: same with March 1996 view (maximum = 2243)

*Locating a building part.* Here, the model was a part of the building in progress, obtained from the September 1995 full-building view used above. This model was then located in subsequent images of the complete building. Figure 8 shows the results obtained. Dilation values were the same as for the previous experiment.

The last rows shows an interesting example of false match, where the algorithm apparently fails to detect the correct part of the building. One must however remember that the algorithm simply looks for *most similar images*: the part found looks indeed the most similar to the model, as the appearance of the “right” part of the building changed radically after removing the scaffoldings.

We notice that the values of the maxima of the similarity surface decrease monotonically in time, that is, as the amount of visible changes increases, reflecting the continuous nature of visible changes.

#### 4.4 Discussion

The Hausdorff iconic matching algorithm has several attractive features. It detects (recognizes) a model image in spite of reasonable amounts of changes in viewpoint, occlusion, and presence of distracting elements. It works with unconstrained imagery, requiring no special training for personnel taking pictures. It is fast, as not based on expensive numerical computation. It has therefore high potential for automatic indexing in large volumes of images, for instance a record of a whole project after completion of work, or as part of a real-time retrieval system from a large portfolio of images from various projects. Within the framework proposed in Section 2, our matching algorithm serves the purpose of identifying specific site views or structures, so that they can be identified in a CAD model of the finished product, and used to quantify progress automatically. From a practical point of view, a narrow range of values of the dilation parameters seems to lead to good results with very different images, apparently removing the need for parameter tuning.

An interesting property of the Hausdorff similarity metric applied to building imagery is that maximum values reflect the amount of change; loosely speaking, smaller similarity maxima are tendentially associated with objects less similar to the model. For example, in the wall recognition test (Figure 6), the maximum value is 1169 occurs when target wall is found, 987 when another wall is found, and smaller values when altogether different objects occur. This property is very useful for image retrieval in databases. The exact relation between similarity maximum values and the probability that the model has actually been found remains to be investigated.

As all iconic matching algorithms, strong occlusions and significant scale or illumination changes may confuse the system. Our experiments indicate, however, that results are sufficiently robust to cope with variations occurring in real, outdoors site imagery. Good robustness is confirmed by preliminary studies with different images and targets (Odone et al 2001 and 2001a).

The quality and robustness of component and structure detection can be improved using *multiple-view models*, in which the single model image  $M$  is replaced by a set of images showing the target object in various orientation and sizes. Locating a target requires therefore comparing each input image with several model images, and fast iconic matching becomes essential. Our current algorithm lends itself perfectly as a fast component for multi-view recognition (Murase and Nayar 1995). Notice, however, that even a single model yields useful results, as reported above.

## 5 Conclusions

This paper is part of an ongoing research effort the aim of which is to automate and integrate progress measurement with design, costing and scheduling. The paper proposes an overall conceptual framework to achieve the above and presents the development

of two applications that are essential for such system/framework to be turned into a prototype. Current efforts aimed at integrating design, costing and planning will undoubtedly make significant contribution to the efficiency of the construction industry, but Construction Integrated Systems will not be achieved unless progress on site is captured efficiently and digitally.

The two main applications developed in this particular paper are the measurement of progress using consecutive images and the automatic determination of location of parts of buildings using images from different viewpoints. The novelty of the first application lies in the introduction of automatic assessments usable in IT systems while construction work is still in progress; normally, only information developed at the design stage is used. The second application proposes an algorithm for recognizing objects and structures in unconstrained outdoors site imagery. This can be used to determine the location of a part of building or structure within a site, and for non-textual indexing and image selection in large

image records from construction projects. Images can therefore be taken using portable digital cameras from arbitrary, non-fixed locations. This is very important for real life situations where fixed site views may be obstructed by construction activities.

Two actual case studies were used to assess the proposed algorithms and results demonstrate their robustness. It is worth pointing out that the tests performed in this paper were on few selected objects. Future work will continue the testing process particularly for interior objects. It is anticipated that the complete prototype will undoubtedly be intelligent and rely heavily on CAD generated models of buildings in progress.

### Acknowledgements

Standard Life who gave access to the Midlothian site  
 Wimpey who gave access to the housing site  
 Amos Haniff, Sufian Zainal and Colin Mathieson for identifying and collecting the images of the testing projects.

### REFERENCES |

- [1] Ahanger, G and Little, T, (1996) "A Survey of Technologies for Parsing and indexing Digital Videos", *JVCIR*, 7(1), 1996, pp. 28–43.
- [2] Albertz, J., Wiedemann, A., 1995. Acquisition of CAD data from existing buildings by photogrammetry. Proc. 6th Intern. Conf. on Computing in Civil and Building Engineering, Berlin, 1995.
- [3] Al-Hajj, A. and Aouad, G.,(1999),The Development of an Integrated Life-Cycle Costing Model Using Object Orientated and VR Technologies,8th International Conference on Durability of Building Materials and Components. May 30th-June 3rd, Vancouver, Canada.
- [4] Alter and Grimson W.E. (1993). Fast and Robust 3-D Recognition by Alignment, Proceedings of the 3<sup>rd</sup> IEEE Int Conf on Computer Vision.
- [5] Aouad G., Child T., Brandon P., and Sarshar M (1998). Linking design, time and cost planning of buildings through VR using an object oriented environment. Proceedings ARCOM 98, Reading UK
- [6] Aouad G., Betts M., Brandon P., Brown F., Child T., Cooper G., Ford S., Kirkham J., Oxman R., Sarshar M. and Young B. (1994) *Integrated databases for design and construction: Final Report*. University of Salford (Internal report), July 1994.
- [7] Bjork, B.C. (1989) "Basic structure of a proposed building product model"; *Computer Aided Design*, vol. 21 number 2, March 1989. pp 71–78.
- [8] Bjork, B.C. (1999) "Information Technology in Construction: domain definition and research issues", *Int. Journ. of Comp. Integrated Design and Construction*, vol 1 no 1, pp. 3–16.
- [9] Brunelli, R., Mich, O., Modena, C.M., (1999), "A Survey on the Automatic Indexing of Video Data", *JVCIR*, vol 10(2), 1999, pp. 78–112.



- [10] Hadwan M., Kaka A.P., Knight M. and Carter D. (2000). Application of photogrammetry in lighting calculations for obstructed interiors. *Journal of Lighting Research Technology* 32(1) pp. 13–17.
- [11] Isgro, F. and Trucco E. (1999). Robust estimation of motion, structure and focal length from two views of a translating scene. *Pattern Recognition Letters*, vol 20, no 8, pp 847–854.
- [12] Knight, MW and Kaka, A “The Application of Photogrammetry to Building progress measurement”, *proceedings of 1998 pp*
- [13] Kartam, N. (1994) “ISICAD: interactive system for integrating CAD and computer-based construction systems”, *Microcomputers in Civil Engineering*, 9 (1994), pp 41–51.
- [14] Jain, R, Kasturi, R, and Schunk, B. G (1995) *Machine Vision*. McGraw-Hill.
- [15] Lapin, L (1998), “Probability and Statistics for Modern Engineering, Second edition, Waveland Press.
- [16] Leonardis, A, and Bischof, H (2000) “Robust recognition using eigenimages”, *Computer Vision and image Understanding*, 78(1), pp. 99–118.
- [17] Lienhart R (2000), “Dynamic Video Summarization of Home Video” *SPIE 3972: Storage and Retrieval for Media Databases 2000*, 2000, pp. 378–389.
- [18] Lowe D. (1993) Fitting Parametrized 3-D models to Images, *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol 13, pp 441–550.
- [19] Luiten, B et al. (1993) “An Information Reference Model for Architecture, Engineering and Construction.”; in (Mathur, K.; Betts, M.; Tham, K. eds.) *Management of Information Technology for Construction*, World Scientific & Global Publication Services, Singapore 1993, pp. 391–406
- [20] Mason, M., Streilein, A., 1996. *Photogrammetric Reconstruction of 3D City Models*. Submitted to SA J. Surveying and Mapping.
- [21] McKinney, K., and Fischer, M. (1998) “Generating, evaluating and visualizing construction schedules with 4D-CAD tools.” *Automation in Construction*, 7(6), 433–447.
- [22] MOB (1994) *Rapport Final, Modeles Objet Batiment, Appel d’offres du Plan Construction et Architecture*, Programme Communication/Construction, (11).
- [23] Murase, H and Nayar, S. (1995) “Visual Learning and Recognition of 3-D Objects from Appearance”, *Intern Journ of Computer Vision*, 14, 1995, pp. 5–24.
- [24] Odone F., Fusiello, A., Trucco E (2000). *Robust Motion-based Segmentation for Content-based Video Coding*. RIAO 2000, Paris.
- [25] Odone F., Trucco F. and Verri A. (2001) *A Flexible Algorithm For Image Matching*, Proc. 11<sup>th</sup> IEEE/IAPR Int. Conf. on Image Processing and Applications ICIAPO1, Palermo (Italy).
- [26] Odone F., Trucco F. and Verri A. (2001a): *General purpose matching of grey level arbitrary images*. In C. Arcelli, L. Cordella, G. Sanniti di Baja editors, 4th International Workshop on Visual Form IWVF4, Lecture Notes in Computer Science, LNCS 2059, Springer-Verlag, pp. 573–582.
- [27] Powell, J (1995) “Virtual reality and rapid prototyping for engineering”, Proc. IT Awareness Workshop, Salford University, 1995.
- [28] Rao, R and Ballard, D (1995), “An Active Vision Architecture Based on Iconic Representations”, *Artificial Intelligence*, vol 12, 1995.
- [29] Rezgui Y, Brown A., Cooper G., Aouad G., Kirkham J. and Brandon P. (1996) “An integrated framework for evolving construction models.” *The International Journal of Construction IT*, vol 4, No 1, 1996, pp 47–60.
- [30] Rosin, PL (1999) “Robust pose estimation”, *IEEE Trans. on Systems, Man and Cybernetics*, vol 29 no 2, pp. 297 ff.
- [31] Sawyer, P., Bell, J., 1994. *Photogrammetric recording and 3D vis-ualization of Ninstints – a world heritage site*. IAPRS, Part 5, pp. 345–348.
- [32] Schmitt, G., 1993. *Virtual Reality in Architecture*. Chapter in “Virtual World and Multime-dia”, Nadia Magnenat Thalmann (Ed.), Wiley&Sons, New York, 1993.

- [33] Streilein A. (1996) Utilisation of CAD Models for the Object Oriented Measurement of Industrial and Architectural Objects. ISPRS Journal of Photogrammetry & Remote Sensing, Vol. 21 part B5, Vienna, pp 548–553.
- [34] Suveg, I and Vosselman, G (2001) “3-D econstruction of Building Models”, Proc IXI IAPRS Congress, 23(B2), 2001, pp. 538–545.
- [35] Tommasini, T Fusiello A, Trucco E and Roberto V. (1998). Making Good Features Track Better, Proc. IEEE Intern. Conf. on Computer Vision and Pattern Recognition CVPR98.
- [36] Trucco E and Verri, A. (1998). Introductory techniques to 3-D Computer Vision, Prentice-Hall.
- [37] Trucco, E., Fusiello A. and Roberto V. (1999). Robust Motion and Correspondences of Noisy {3-D} Point Sets with Missing Data. Pattern Recognition Letters, vol 20, no 9, pp 889–898.
- [38] Van Nederveen, S. (1994) *Atlas, View Type Model for Global Architectural Design*; occasional paper.
- [39] Watson A. and Crowley A. (1995) CIMSteel integration standard, in: Scherer R.J. (Eds.), *Product and Process Modelling in the Building Industry*, A.A. Balkema, Rotterdam, pp491–493
- [40] Waldhäusl, P., 1991. A test object for architectural photogrammetry: Otto Wagners underground station Karlsplatz in Vienna, Proceedings of the XIV. International Symposium of CIPA, October 2–5, 1991, Delphi, Greece, pp. 247–251.
- [41] Wunsch P and Hirzinger G,(1997) “Real-Time Visual Tracking of 3-D Objects with Dynamic Handling of Occlusions”, Proc. IEEE Int. Symposium on Robotics and Automation, 1997.