Refinement to the Chamfer matching for a "center-on" fit

Jingying Chen Paul Tan Terence Goh
Singapore Technologies Dynamics
249 Jalan Boon Lay Singapore 619523

Abstract

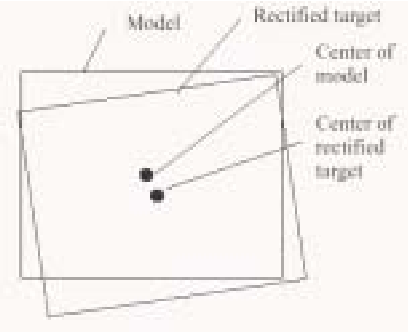
Matching is a central problem in image analysis and pattern recognition. In order to develop a matching algorithm that provides the optimal fit (i.e. "center-on" fit attempting to position the model points at equal spacing from the target image) between the model and distorted target images, a hierarchical modified Chamfer matching system with the simulated annealing is proposed in this paper. Since efficiency and accuracy are important issues in image matching, the proposed approach employs a hierarchical approach (i.e. multi-resolution pyramid) to speed up the matching process and then the simulated annealing is applied to the modified Chamfer matching scheme to achieve the optimal solution. Encouraging results have been achieved by the proposed technique. This suggests a strong potential alternative to automatic target recognition system.

Keywords: Chamfer matching, multi-resolution pyramid, simulated annealing, "center-on" fit, target recognition

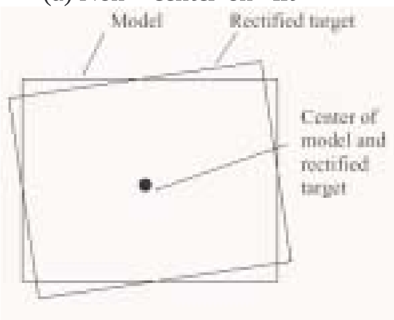
1 Introduction

Matching is a central problem in image analysis and pattern recognition, and edges are perhaps the most important low-level image features. Hence, the good edge matching algorithms are important. Chamfer matching is an edge matching technique, which was first proposed by Barrow *et al.* [1]. In their scheme, the edge points of one image were transformed by a set of parametric transformation equations to edge points of a similar image which was distorted somehow. Chamfer matching then attempts to find the best fit for the edge points from the two images by minimizing a generalized distance between them. However, this scheme requires good *a priori* knowledge on the parametric transformations, and is thus useful only in a limited number of applications [1]. Borgefors [2] proposed a hierarchical Chamfer matching method that has been embedded in a resolution pyramid. The matching is done not only in the original image resolution, but in a series of images, where each image is a representation of the original scene at a lower resolution. A number of initial positions are considered at some resolution level of the pyramid, where a Gauss- Seidel optimization procedure is used to find a local minimum for each initial position. Poor local minima are rejected. The remaining positions are considered at next higher resolution level of the pyramid, and the procedure is repeated until local minima are found at the highest resolution level of the pyramid. This technique (i.e. multi-resolution approach) speeds up the computations considerably, however, it still cannot guarantee the best fit in the event an attempt is made to match a distorted image to the corresponding model image. Both Barrow *et al.* and Borgefors applied edge detection to create a binary feature map for distance transform (DT). Distance transform (DT) converts the edge image into a DT image where each pixel denotes the distance to the nearest edge pixel. You *et al.* [3] used interesting points to replace edge points in distance transform for the chamfer matching. Liu and Srinath [4] employed curvature points as a basis for feature primitives in Chamfer matching. Based on their method, the estimates of rotation, scaling and translation parameters between the model and target images can be obtained, and thus matching can be successfully implemented for partial shapes.

Most of the target recognition applications assume that target image has been corrected for perspective distortions, and so matching can be successfully carried out. It is essential that the target image must first be rectified for rotation, scaling and perspective distortions. However, perspective distortions require warping transformations between the target and model images [5]. In the case where there are errors in the rectification parameters, the target image will not be rectified correctly. Consequently, it may not be easy to get a good fit between the rectified image and model image. In order to get the best fit between the poorly rectified target image (i.e. distorted target image) and the model image, a hierarchical modified Chamfer matching system with the simulated annealing is proposed in this paper. Since efficiency and accuracy are important issues in image matching, the proposed approach employs multi-resolution pyramid to speed up the matching process and then simulated annealing is applied to the modified Chamfer matching scheme to achieve the optimal solution, i.e. the “center-on” fit between the target image and its model (see Figure 1). It has strong potential application. The modified Chamfer matching consists of two terms: the Chamfer distance measure and smoothness constraint measure.



(a) Non- “center-on” fit



(b) “Center-on” fit

Figure 1: Illustration of “center-on” fit.

The outline of the paper is as follows. Some basic concepts are introduced in Section 2. Section 3 presents the novel technique, a hierarchical modified Chamfer matching system with the simulated annealing. Section 4 describes the results while Section 5 presents the conclusions and directions for future work.

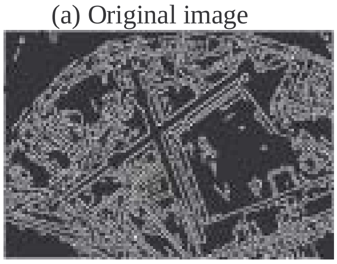
2 Basic concepts

Some basic concepts of distance transform, matching measure, hierarchical matching (multi-resolution pyramid) and simulated annealing are introduced below.

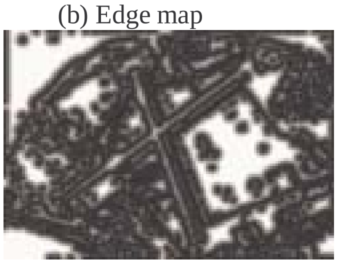
2.1 Distance transform

Distance transform (DT) is such an operation which measures the distance of non-edge pixels to the nearest edge pixel while the edge pixels get the value zero. The purpose of the distance transformation is to produce numeric image whose pixels are labeled with distance between each of them and their closet border pixel. It is important that the DT used in the matching algorithm is a reasonably good approximation of the Euclidean distance, otherwise the discriminating ability of the matching measure, computed from distance map (DT image), becomes poor [2]. One approximation of the Euclidean distance is the *chamfer metric* [1, 2], which is employed in this paper. Such distance transform makes use of two 3-4 DT masks (refer to Borgefors [2]) by two-pass algorithm (i.e. the forward pass moves from left to right and from top to bottom; and then the backward pass that moves from right to left and from bottom to top). One example of edge map and distance map is illustrated in Figure 2.

(a) Original image



(b) Edge map



(c) Distance map

Figure 2: Example of edge and distance map.

2.2 Matching measure

One of the common applications of distance transform is to evaluate image matching by overlaying the model over the distance transform map generated from the target edge map, and measure the extent of how well the model fits in terms of the pixel values in the distance transform map hit by the edge pixel in the model. In this way, the edge distance between the model edge map and the target distance map can be computed. The edge distance is the sum of the values of the pixels covered. Therefore, the likeliest match occurs when the sum of the edge distance is the minimum. This minimum takes place when the model edge map has been translated in such a way that its edges match those in the target distance map. Borgefors indicated that the root mean square average produce better edge distance for measuring the fit between the model and target images and yield fewer false alarm minima than any of the other averages. The root mean square average can be formularized as:

$$D_c = \frac{1}{3}\sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2} \quad (1)$$

where d_i is the Chamfer distance value and n is the number of edge points in model. As 3-4 DT mask is used for the distance transform of the target image, the compensation for the unit distance is made by dividing the root mean square by 3. The advantage of matching a model with DT map rather than with the edge map is that the resulting similarity measure will be smoother as a function of the model transformation parameters [6]. However, the matching scheme does not result in a "center-on" fit between the model and the poorly rectified target image. A "center-on" fit is very important in the automatic target recognition application. Hence, the modified Chamfer matching is proposed. It incorporates the smoothness constraint (refer to Section 3) attempting to position the model points at equal spacing from the target image.

2.3 Hierarchical matching

The hierarchical matching is based on the multi-resolution pyramid (i.e. the matching is done not only in the original image resolution, but in a series of images, where each image is a representation of the original scene at a lower resolution). Computational time can be saved using pyramid. The original image is the bottom, zero, level of the pyramid. From this level the next, first, level is generated by substituting one pixel at the next level for each block of four pixels. If the image is a gray level image, then the value of the upper level is an average, usually arithmetic or median, of the values of the lower level. This process is repeated until only one pixel is left. The image that will be converted into a pyramid in this paper are the edge map and distance map. Borgefors argued that the OR-pyramid (i.e. logical operation OR pyramid) is better than average pyramid for edge image. Thus, an OR- pyramid is used in this paper. Then, the matching is done in the pyramid. Examples of pyramids for edge map and distance map are shown in Figure 3 and 4.

2.4 Simulated annealing

Simulated annealing is a stochastic optimization technique which is based on the analogy between the annealing of solids and solving optimization problem. Simulated annealing has been applied to a wide variety of image processing applications such as segmentation [9] and shape detection [10]. The key point of the simulated annealing is the Metropolis algorithm [7], which simulates the evolution to thermal equilibrium of a solid for a fixed value of temperature *T*. At each temperature *T*, the solid is allowed to reach thermal equilibrium, characterized by a probability of being in a state with energy *E* given by the Boltzmann distribution:

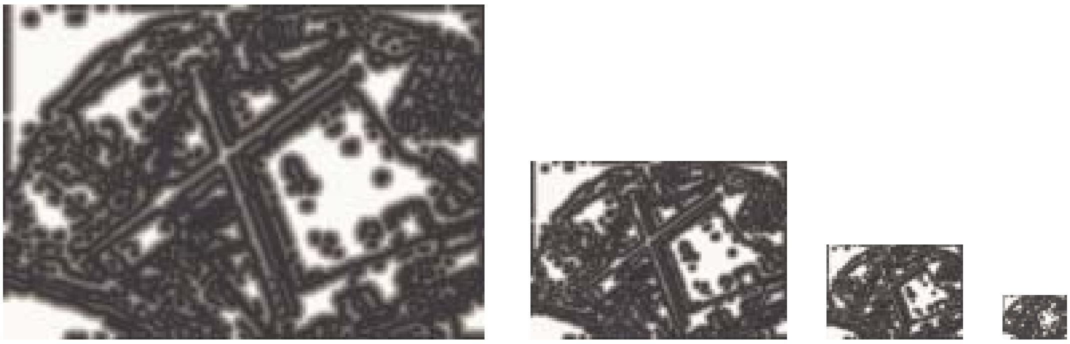
$$p(E) = \exp(-\frac{E}{KT}) \quad (2)$$

where *E* denotes the energy, *K* is the Boltzmann constant and *T* is the temperature. *p(E)* denotes the probability of a state having energy *E* and temperature *T*. The simulated annealing algorithm can be viewed as a sequence of Metropolis algorithm evaluated at a sequence of decreasing values of the control parameter (temperature *T*). It can thus be described as follows. Initially, given a high temperature value and a state S_i, another state S_j can be obtained by choosing at random an element from the neighborhood of S_i, which corresponds to the small perturbation in the Metropolis algorithm. Let $\Delta E_{ij} = E(S_i) - E(S_j)$, then the probability for state S_i to be the next state in the sequence is given by 1, if $\Delta E_{ij} \le 0$, and by $\exp(-\frac{\Delta E_{ij}}{KT})$, if $\Delta E_{ij} > 0$ (the Metropolis criterion). Thus, there is a non-zero probability of continuing with a state with higher cost than the current state. This process is continued until equilibrium is reached. Then the temperature is lowered in steps and the procedure mentioned above is repeated until a very low temperature is achieved. The algorithm is shown below.

Select a random state S_i.
Select a starting temperature T.
While T>0 do
  perturb $S_i \to S_j$, ΔE_{ij}
  if $\Delta E_{ij} \le 0$ *then accept else*
  if $\exp(-\frac{\Delta E_{ij}}{KT}) >$ *random number in [0,1] then accept*
  if accept then replace S_i *with* S_j *else*
  continue with current S_i
  if equilibrium is approached sufficiently closely then lower the temperature T.

3 Hierarchical modified Chamfer matching with the simulated annealing

In the proposed system, the model edge map is used to match against a rectified target image, in the form of a Chamfer distance map. The target image may not have been rectified properly owing to some errors in



(a) Target image size: 320×240 (b) 160×120 (c) 80×60 (d) 40×30

Figure 3: Pyramid of the distance map of target image.

(a) Model size: 80×64 (b) 40×32 (c) 20×16 (d) 10×8

Figure 4: Pyramid of the edge map of model.

the rectification input parameters. As a result, a "center-on " match may not be achieved by the conventional Chamfer matching. The lowest matching measure may not correspond to the best match between the model and rectified target images. A matching unevenness exists although the distance measure has been minimized; some pixels may have small distance magnitudes while other pixels may have large distance magnitudes, which are computed from the Chamfer distance map. Thus, there may be many variations close to the global minimum measure that may offer sub-optimal matches. Consequently, it is difficult to determine which is the correct match, even for the global minimum measure that has been obtained. Hence, the modified Chamfer matching incorporating the smoothness constraint with the simulated annealing is proposed.

In the simulated annealing optimization, an energy function *E* is introduced as,

$$E = \lambda \times D_c + (1-\lambda) \times D_s \qquad (3)$$

where λ is a pre-determined weight value, D_c and D_s are the respective Chamfer distance measure (see equation (1)) and smoothness constraint measure. D_s is given as follows:

$$D_s = \sum_{k=1}^{n-1} (d_k - d_{k-1}) \qquad (4)$$

where d_k and d_{k-1} are the Chamfer distance measures for two neighboring target points. n is the number of edge points. The conventional Chamfer matching process finds a match that corresponds to a minimization of the energy equation. In this case, the attempt is to minimize both the Chamfer distance measure and the smoothness constraint. The smoothness constraint attempts to position the model points at equal spacing from the target image. The algorithm for the modified Chamfer matching with the simulated annealing is described as follows:

a. An initial state S_i is randomly selected. S_i corresponds to the model edge map being overlaid at a random position over Chamfer distance map. The energy function E_i (E_i corresponds to state S_i, refer to equation (3)) is computed. The system initial temperature *T=100*.

b. The next state S_j is randomly perturbed from current state. The energy function E_j (E_j corresponds to state S_j) is computed.

   If $(E_j - E_i) < 0$

   $S_i = S_j$, $E_i = E_j$.

   The initial state S_i has now been re-assigned.

c. However,

   If $(E_j - E_i) > 0$

   Go to the Metropolis algorithm. If an arbitrary random number is greater than the Boltzmann function, accepted E_j as the new E_i state.

Otherwise, reject the new state E_j and retain E_i.

d. Repeat (b) and (c) for *L* iterations.
e. Lower the temperature by $T = T \times \alpha$.
f. Repeat (b), (c), (d) and (e) until the stop criterion is satisfied.

The parameter L was determined heuristically via a set of test images whose correct matches were known. A proportional cooling schedule is used here ($T = T \times \alpha$, where $\alpha \in (0,1)$). As for the stop criterion, it is clear however that when the equilibrium values of the energy function for successive stages are constant themselves, the iteration process can be stopped.

Since efficiency is an important factor in matching, hierarchical method (multi-resolution pyramid method) is incorporated to speed up the computations. The proposed method includes two steps:

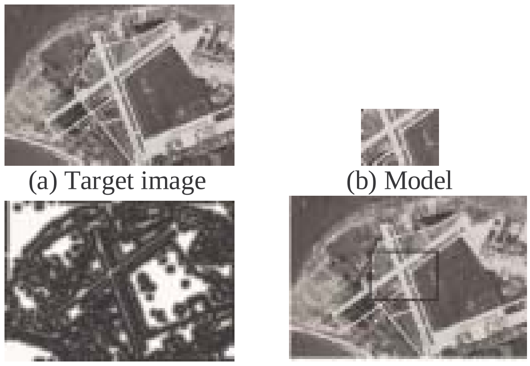
Step 1: Generate two multi-resolution image pyramids.

a. At the beginning, Canny edge detection [8] is performed on both model and target images.
b. Then, Chamfer distance transform is employed on the target edge map to generate the target distance map.
c. Two multi-resolution image pyramids are built for the model edge map and target distance map. The original image size is at level zero which is at the bottom of the pyramid, and at each higher level, the image size is scaled down by a factor of 2 on both height and width. The highest level is limited at 3 in this work.

Step 2: Perform the modified Chamfer matching with the simulated annealing on all image levels from the highest (e.g. level 3) to the lowest level (e.g. level zero).

3 Results

In order to assess the performance of the simulated annealing optimization in the hierarchical modified Chamfer matching, two sets of experiments have been conducted based on Pentium III using C. In the first set of experiments, the aerial images are used. The target image size is 320×240, the model image size is 80×64. One example is shown in Figure 5. The processing time is around 0.05 seconds. The first set of experiments is a test to determine if the matching system will result in a correct match. From the result, one can note that the model has been correctly matched in the target image (see Figure 5(d)).

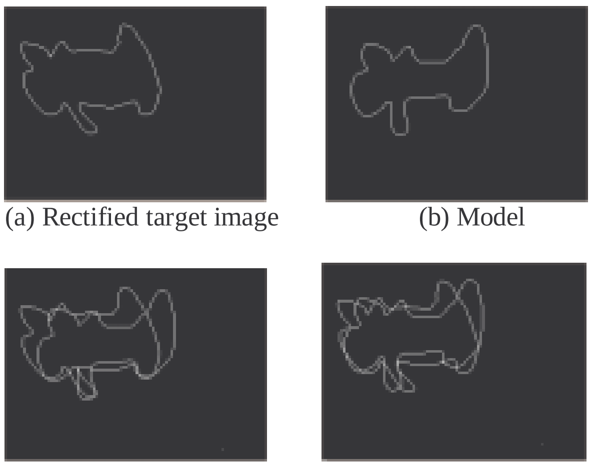


(a) Target image (b) Model

(c) Distance map of target image (d) Correct match

Figure 5: The matching result by the proposed method.

In the second set of experiments, comparisons have been made between the conventional Chamfer matching and the proposed method using images with irregular shapes. Figure 6 illustrates one example for lake matching (the image size is 320×240). The target image was first rectified (see Figure 6(a)). The rectification is not perfect in that the rectified image may not be properly registered with the model image (see Figure 6(b)). The best fits obtained between the target and model images by the conventional Chamfer matching method and the proposed method are illustrated in Figure 6(c) and (d). From the results, one can see that a better fit has been obtained by the proposed method which attempts to position the model points at equal spacing from the target image for a "center-on" fit. This suggests a strong potential alternative to automatic target recognition system.



(a) Rectified target image (b) Model

(c) The best fit by the conventional Chamfer matching (d) The best fit by the proposed method

Figure 6: The comparison made between the conventional Chamfer matching and the proposed method for lakes.

4 Conclusions

This paper presents a novel method tending to find a "center-on" fit between the distorted target and the model images. It has a strong potential for the automatic target recognition application. Since efficiency and accuracy are important issues in image matching, the proposed approach employs a hierarchical approach (i.e. multi-resolution pyramid) to speed up the matching process and then simulated annealing is applied to the modified Chamfer matching to achieve the optimal solution (i.e. the "center-on" fit between the target image and its model). The proposed method, the integration of multi-resolution pyramid and simulated annealing optimization in the modified Chamfer matching, has been applied on the real images with encouraging results. Nevertheless, this approach will need to be further tested against more real images under noisy conditions so as to further assess its robustness.

6 References:

[1] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles and H. C. Wolf, "Parametric correspondence and chamfer matching: Two new techniques for image matching", *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, Cambridge, MA, 1977, pp. 659-663.

[2] G. Borgefors, "Hierarchical chamfer matching: A parametric edge matching algorithm", *IEEE Trans. on PAMI*, vol. 10, No. 6, 1988, pp. 849-865.

[3] J. You, W. P. Zhu, E. Pissaloux and H. A. Cohen, "Hierarchical image matching: A chamfer matching algorithm using interesting points", *Proceedings of the Third Australian and New Zealand Conference on Intelligent Information Systems*, Perth, Australia, November, 1995, pp.70-75.

[4] H. C. Liu and M. D. Srinath, "Partial shape classification using contour matching in distance transformation", *IEEE Trans. on PAMI*, vol. 12, 1990, pp.1072 -1079.

[5] R. T. Collins and J. R. Beveridge, "Matching perspective views of coplanar structures using projective unwarping and similarity matching", *Proceedings of the IEEE Conference on Computer Vision*, NY, June, 1993, pp. 240-245.

[6] D. M. Gavrila and V. Philomin, "Real time object detection for 'smart' vehicles", *Proceedings of the Seventh IEEE Conference on Computer Vision*, Kerkyra, Greece, September, 1999, vol. 1, pp. 87-93.

[7] P. J. M. van Laarhoven and E. H. Aart, "Simulated annealing: theory and applications", Kluwer Academic Publishers, Boston, 1988.

[8] R. Jain, R.Kasturi and B. G. Schunck, Machine Vision, MIT Press and McGraw-Hill, 1995.

[9] P. Fieguth and S. Wesolkowski, "Highlight and Shading Invariant Color Image Segmentation Using Simulated Annealing", *Proceedings of the third International Workshop on EMMCVPR*, Sophia Antipolis France, September 3-5, 2001.

[10] M. Betke and N. C. Makris, "Fast object recognition in noisy images using simulated annealing", *Proceedings of the Fifth IEEE Conference on Computer Vision,* Cambridge, MA, June, 1995, pp. 523-530.

Filmmaking Production System with Rule-based Reasoning

Shen Jinhong, Seiya Miyazaki, Terumasa Aoki, and Hiroshi Yasuda
Faculty of Engineering,
The University of Tokyo, Tokyo, Japan
{j-shen, seiya, aoki, yasuda}@mpeg.rcast.u-tokyo.ac.jp

Abstract

This paper describes a software system designed to automate the production of digital movies with various visual effects like three-dimension animation, real image, and their composition. The production system can understand user's input screenplay through a parser then automatically interprets it into a relevant motion picture under the direction of a virtual director in place of a human one. The virtual director achieves user's intentions through knowledge-based approach by setting a scene, determining the corresponding shot types and shot sequence, and planning virtual camerawork dependent on the cinematic expertise stored in a domain knowledge base. We model the filmmaking knowledge and rule-based reasoning strategies in expert system language CLIPS. Video data is encoded in XML and tracked by the MPEG-7 standard.

Keywords: digital movie producer, rule-based system, 3D animation, knowledge representation, virtual director,

1 Introduction

There are three general ways to generate digital movie. (1) Originating from conventional film: Change the images of film into digital video (DV). (2) Utilizing digital video camera to shoot: DV camera works as a carrier to shoot. Usually the post-production of DV processes on computer. (3) Creating Computer Graphics (CG) movie: Whole movie is completely made by computer directly. Because these approaches are all time-consuming and/or expensive, it is still impossible for us to produce personal movie readily within short time when we come up with an idea for movie. Our research aims at developing an automation technique by which anyone can easily make and deliver his own movie.

Movies generated by the ways (2) and (3) mentioned above are considered "real" digital ones. Digital video production using DV camera encompasses acquisition, storage, selection/editing, and composition of video data. Except that actual shooting requires human' involvement, the process of video choosing and sequencing can be automated based on experienced editing knowledge. That is to say the process of digital video production is at most of automatic edition and composition. For the production of CG Movie, supposing the existence of a library that stores 3D models and actions mentioned in the script, it is possible to combine objects and actions according to the screenplay and to choose optimal placement for the camera automatically. Therefore automatic edition and computer animation are feasible.

The desktop moviemaking system *DMP* (Digital Movie Producer) we are implementing can interpret a verbal screenplay into a relevant motion picture automatically with various visual effects like real image, three-dimensional (3D) animation, or their composition [1-4]. The remainder of this paper is structured as follows. Section two introduces the system structure of DMP after analysing the related methods of automatic movie creation. The next section outlines the relevant filmmaking techniques utilized in the virtual 3D world from a film theoretic point of view and gives design of knowledge representation (KR) and reasoning strategies respectively. In the fourth section, the system implementation with an example piece of animation is showed to expound how to use cinematic 'rules of thumb' to make a scene. Finally, we will have a discussion about our work.

2 System Design

2.1 Related Work

Works on applying film theory for computer graphics generation have been put forward. Christianson et al. adopted the notion of *film idioms* from film theory and formalized them into a sequence of shots [5]. He et al. encoded the film idioms into hierarchically organized finite state machine applied in real-time system [6]. Amerson & Kime proposed a system *FILM* (Film Idiom Language and Model) for real-time camera control in interactive narratives [7]. New methodologies employ knowledge-based approach to address the tasks of graphics generation. In [8, 9], domain knowledge base was applied in automatically generating animation focusing on camera shot design while in [10] animation creation focused on human gesture. Cognitive modelling for intelligent agent was