e-Shop with the atmosphere of physical shop

Ronald Chung, *SrMIEEE* Yong He
Department of Automation & Computer-Aided Engineering
The Chinese University of Hong Kong, Shatin, HONG KONG
e-mail: rchung@cuhk.edu.hk, yhe@acae.cuhk.edu.hk

***Abstract*--A new concept of e-shop is introduced. Existing e-shops are very much like the classical catalog marketing, offering potential customers a catalog of merchandise choices and expecting placement of purchase orders in return. The atmosphere of physical shopping, like having other fellow shoppers nearby, watching in live form what merchandises they crowd in to buy etc., are absent. We propose an e-shop that is constructed directly from, and in fact co-existing with, a physical shop. The e-shop has a number of cameras installed in the physical shop, and it allows e-shoppers to perceive all the actions taking place in the physical shop through internet transmission of the video data captured by the cameras. The key problem to solve in implementing the proposed e-shop is how to determine the intended viewpoint change or the intended merchandise-to-purchase of the e-shoppers. Using supermarket as an example, we propose a framework to tackle the problem. Experimental results with real images are also presented.**

***Index terms*—e-shop, view matching, homography**

I. INTRODUCTION

The popularity and advancement of internet in the last decade has helped generate and make possible a number of new concepts on traditional practices. A good example is e-mailing for delivering mails. Other examples include e-commerce for marketing and money transactions, tele-conferencing, and e-shopping. This paper is about a new concept of e-shopping that is further evolved from the current e-shopping.

The current e-shopping, also called virtual shopping by some, typically consists of a web site with a list of merchandises (and perhaps also pictures of them) for shoppers to browse and make purchase request of through the internet. However, looking at the concept more closely, one could find that the whole setup is not very much different from that of catalog marketing started decades ago; the major difference is perhaps the shorter turnaround time between the delivery of merchandise list to the customers and the receipt of purchase requests from them, due solely to the faster speed of internet communication. Many essences of physical shopping, including the atmosphere of having other fellow shoppers nearby, watching in live what merchandises they crowd in to buy, owning the luxury of pushing the trolley around in a supermarket without having a particular merchandise to buy initially, and so on, are absent. Such essences of physical shopping is however regarded by many as what make shopping enjoyable. The essences could also play a positive role in boosting the visit rate of the shopping site as well as the purchase rate of merchandises due to the "crowd" effect of shopping they might induce. Some of the current e-shops might have three-dimensional virtual space for customers to navigate within, but then all that is perceivable is still synthetic, and the above-mentioned essences in such virtual shops are still no match with those in physical shops.

In this paper we propose a new concept of e-shopping that provides customers the perception of physical shopping. We propose to have e-shop constructed directly from, and in fact co-existing with, a physical shop. Customers would be able to navigate in a physical shop, perceiving the presence of other physical shoppers and everything else that would be felt in physical shopping, except that they need not be physically there in the shop.

We have three goals to meet in the e-shop we propose:

1. It allows customers to navigate within the physical shop, perceiving all physical presence and actions that happen in the physical shop as if they were physically there.
2. It allows customers to select merchandises and make purchase request of them.
3. It allows concurrent shoppings of multiple customers, each customer having his/her own navigation route in the e-shop.

In this paper we describe how we materialize the concept using supermarket as an example. We also present some preliminary experimental results we have obtained in our implementation.

II. THE VIEWPOINT-CHANGE PROBLEM

The above three goals we set out to achieve point to an e-shop that is built from a physical shop equipped with cameras. With cameras, image data of both the merchandises (in a form they should appear in a physical shop) and the physical shoppers can be taken and transmitted to the e-shoppers, thus offering the e-shoppers the atmosphere of physical shopping.

There are two schemes for an e-shopper to move his viewpoint within the physical shop. One is to allow the e-shopper to remotely control the panning, tilting, and zooming degrees of freedom of each camera, so as to simulate the visual effect the customer would perceive should he move about physically in the shop. The other is to equip the physical shop with a dense net of cameras so as to have almost every possible viewpoint in the shop for the e-shopper to choose from. An e-shopper at any given

time would then have a particular camera as his current window ("eye") to the physical shop, and any "e"-movement of his would correspond to a change of the camera assigned to him.

However, if concurrent shoppings of multiple e-shoppers are to be allowed, the only choice is the latter scheme, since there could be multiple customers seeking the same viewpoint, and concurrent reading accesses of a camera's image data are far more plausible than concurrent motion control accesses of a camera.

To summarize, our materialization of the proposed e-shop consists of a physical shop equipped with cameras everywhere. Each e-shopper at any given time would have a particular camera as his current viewpoint, and he is allowed to switch his viewpoint across the net of cameras, each time from one camera to one of the neighboring cameras. With the e-shop co-existing with a physical shop, whatever observable in physical shopping will also be perceivable in e-shopping.

The key problem to solve in implementing such an e-shop is how we determine the intended change of viewpoint, say from a wide-angle view between two shelves to a close-up view of a particular merchandise on one of the shelves, of each e-shopper. The problem is simpler if the physical shop is structured so that it could be enforced at all time that each merchandise is designated a specific position in a specific orientation on a specific shelf. That way the cameras in the shop can each be positioned to be responsible for a particular function or merchandise. For instance, camera C_{ijk} is responsible for giving the kth distant view of the corridor between shelves i and j, and camera C_m is responsible for giving a close-up view and accepting purchase request of merchandise m. With all these preset positions of the merchandises as well as the cameras, for any view (say image I_{ijk} of camera C_{ijk}) within the shop, we could pre-determine which spot in that view is responsible for which merchandise and which camera is responsible for giving close-up view and receiving purchase request of that particular merchandise. If e-shopper e has view I_{ijk} as his current view, and he mouse-clicks a position around (x_m,y_m) in the view, we know he intends to grab the view of camera C_m (which shows merchandise m) in the next instant.

However, the above scheme would mean that most of the physical shops we have now on the market have to go through a major restructuring before they could have their e-shop brothers. The flexibility in adjusting the merchandise layout of a shop in accordance with the need is also lost; shop attendants could no longer move the merchandises to whatever position and refill them in whatever way and quantities they see fit.

Thus in our proposed e-shop we have this additional design requirement: the co-existence of the e-shop should induce minimum disturbance to the operation of the associated physical shop. In particular, shop attendants should be free to operate the shop and adjust the merchandise layout as if the e-shop brother does not exist.

The additional requirement implies that our e-shop does not know *a priori* "what" merchandises are observable in "which" image positions in the fields of view of "which" cameras. The key problem we have to solve is then how we determine the intended viewpoint change of an e-shopper at his mouse-clicking of position (x,y) in his current view I (of camera C). Since a viewpoint change must be gradual geometrically, we could reduce the choices to only the views of those cameras that are in the proximity of camera C. Information on the color intensity histogram of image I could also help reduce the choices. Nonetheless, in general, ambiguity over a number of cameras still remains.

The problem of viewpoint maneuvering can be solved by having a mapping from the mouse-clicks and mouse-drags to the neighborhood relationships of the cameras. For example, a mouse-drag to the left on the screen could be interpreted as a switching of viewpoint from the current camera to the camera on the left. The drawback of this approach is, all viewpoint changes have to be very gradual; the e-shopper cannot simply mouse-click on his screen a merchandise at the far end of a shelf in order to leap his viewpoint to there, but has to move one view-step at a time. More importantly, since cameras (and thus camera views) are not associated with merchandises, making purchase request based upon what is visible on the screen is not possible. Even if there is only one merchandise displayed on the screen, an e-shopper cannot just mouse-click the screen for placing the purchase order.

The problem of determining the camera view (or the merchandise template in the case of accepting purchase order) that *best* matches a mouse-clicked portion of a view (the current view) in terms of the visual data (of the merchandises) they display, for the purpose of either viewpoint change or accepting merchandise purchase request, is thus the key problem to solve in the proposed e-shop. We term this problem the ***viewpoint-change problem.*** The problem can be cast as a matching problem: given the current view I_C (of camera C) of an e-shopper and a small image window W selected (through a mouse-click) by him on I_C, finds the camera N, among all the cameras in the neighborhood of C, whose image I_N best matches W. The problem is illustrated in Figure 1.

In this paper we use supermarket as an example and propose an algorithm to this key problem.

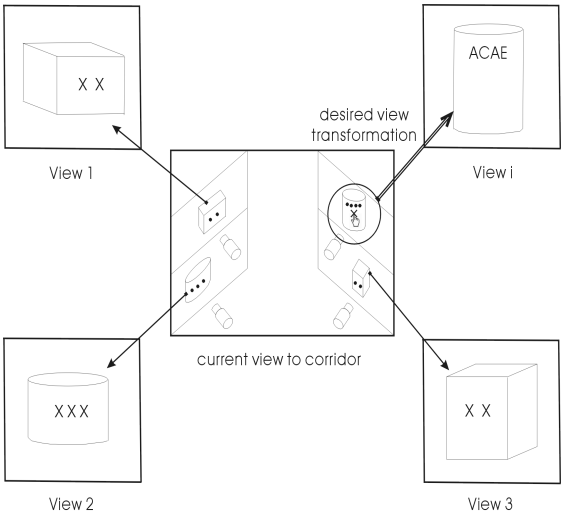

Figure 1: An important feature of the proposed e-shop: e-shopper is free to switch viewpoint in a physical shop from a wide-angle view (the current view) between shelves to a close-up view of a particular merchandise on the shelf (the intended view) through just a mouse click over the merchandise in the current view.

III. 3. E-SUPERMARKET THAT OFFERS PHYSICAL SHOP ATMOSPHERE

Information on the color intensity histogram of the current image I_C can reduce the search space in the matching problem. However, merchandises, especially those in supermarket, could have very similar color histograms. Precise determination of the intended new viewpoint still hinges upon cross-correlation of intensity data between W and each candidate image I_N .

However, there is this difficulty in measuring the cross-correlation: the two images to compare are not two-dimensional (2-D) transformations of each another, but image projections at different and unknown viewpoints of the same object in a three-dimensional (3-D) space. Furthermore, the object could consist of curved surfaces as likely as of planar surfaces. In other words the cross-correlation problem is a 3-D one. What makes the problem challenging is that intensity pattern in the image of an object changes with both the viewpoint and the object shape.

Our proposed solution can be described as the following. We take advantage of the fact most objects in supermarket are of either rectangular shape or cylindrical shape. Such shapes, which have zero or small curvature at most surface points, could be modeled to a good approximation as concatenation of a number of small planar patches. Thus if a number of initial correspondences can be established between template W and any candidate view I_N, such correspondences could not only allow the texture of I_N over the matched portion be represented as a number of contiguous triangular patches, but also allow such triangles be warped to the viewpoint of template W and be compared to W through cross-correlation in the same viewpoint. The final matched view, for viewpoint transfer or for identifying merchandise purchase request, will then be the candidate view which has the best cross-correlation with W.

An overview of our approach in solving the merchandise-view correspondence problem is outlined in Figure 2. Given template W, for every candidate view I (hereafter we drop the subscript N of I_N for simplicity), we first use the feature extractor and matcher named "Image Matcher", which is ported from INRIA [6], to establish point correspondences between W and I. Such correspondences, like those from any other matcher, could contain incorrect correspondences. The correspondences first go through the robust estimation method RANSAC [3] for estimating the epipolar geometry between views W and I , in terms of a Fundamental matrix [5]. In the robust estimation process incorrect correspondences are identified as outliners and discarded. With only the correct ones remained, the correspondences are used to divide the matched portion of I into a number of triangular patches. Here we use the classical method: the Delaunay Triangulation method [1]. Each of such triangles in I has three point correspondences with W. The three correspondences, plus the epipole information associated with the Fundamental matrix, are just enough to estimate an image-to-image mapping named homography [2] that allows all points within the triangle be warped from the image space of I to the image space of W. More precisely, every point (x,y) in the triangle (in view I_N) will have the image position (x',y') in the warped view (at the viewpoint of W) as follows:

$$[x',y',1]^{\mathrm{T}} \approx \mathbf{H}\,[x,y,1]^{\mathrm{T}}$$

where $\approx$ denotes equality up to a scale, and $\mathbf{H}$ is a 3×3 matrix representing the homography for all points in the small planar patch enclosed by the triangle. With the matched portion of I warped to the viewpoint of W, the two views can have normalized cross-correlation in the same viewpoint.

Notice that even if the same merchandise is shown in both W and I, it could be displayed with different scales. Matching two views of very different scales is much harder than matching two of similar scales. For this reason we propose to use a number of scales of W in the above process so as to ease the initial correspondence process. In our implementation, say when we are to transfer the viewpoint from a corridor view to a particular-merchandise view, we use five scaling-up ratios for the template W: 110%, 120%, 130%, 140%, and 150%.

Notice also that in the above process substantial errors in the warping due to imprecision of triangulation (i.e., the imprecision of the polyhedral modeling process) are still tolerable, as here what is needed is not a visually perfect warping, but a warping that gives a cross-correlation estimate which is accurate enough for the target view be identified from a number of candidate views.

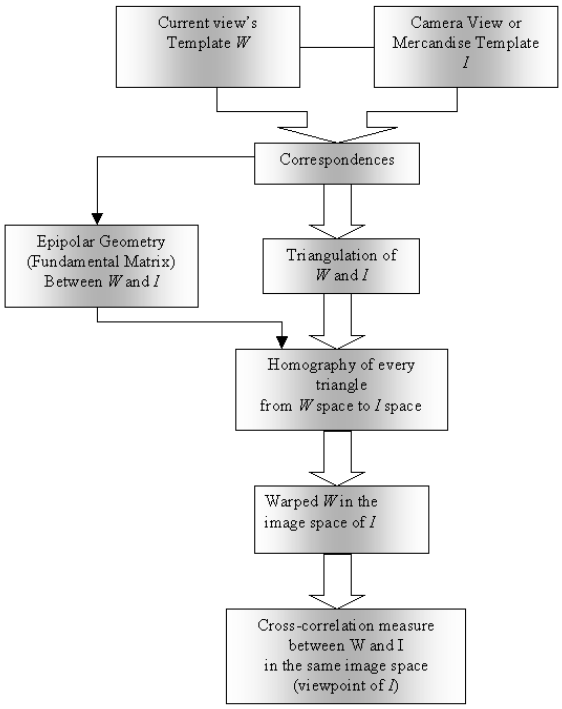

Figure 2: Overview of our solution to the merchandise-view correspondence problem.

IV. 4. Experimental Results

Here we show one set of experimental results to illustrate how our proposed solution performs. Figure 3 shows the current view of a particular e-shopper, which is a wide-angle shot of a number of cans on a shelf. Also shown in the figure are views that are geometric neighbors to the current view. They are Views 1, 2, 3, and 4, each displaying a close-up view of a particular can. Depending upon where in the current view the e-shopper mouse-clicks, either one of the four views should be the next view displayed to the e-shopper.

Should the e-shopper mouse-click a particular position in the current view, a template is cropped from the current view around the clicked position. The template is useful for identifying which of the candidates views (Views 1, 2, 3, and 4) are the view requested by the e-shopper. Figure 4 shows two templates (Templates *a* and *b*) that correspond to two different clicked positions over the current view.

As soon as the e-shopper clicks a particular position in the current view, a template is formed. A number of scalings (five scalings in our implementation) of the template are then matched with each of the candidate views. The scaling step is for easing the matching process: two images are easier to match if they have similar resolution. As mentioned above, we use INRIA's "image-matcher" algorithm [6] for the matching step. The extracted correspondences might not be perfect, yet the incorrect ones could be identified as outliners in the subsequent process: the robust estimation of the Fundamental matrix. The filtered correspondences then allow the template to be warped to the viewpoint of each of the candidate views. With the warpings, cross-correlation comparison between the template and each candidate view can be conducted in the same viewpoint.

Figure 5 shows the "inliner" correspondences between Template *b* and View 2, and the subsequent triangulation meshes constructed over the matched portions. Figure 6 shows the warping of Template *b* to the viewpoint of View 2 (the correct candidate view). It could be seen that although Template *b* and View 2 are drastically different in the intensity profile (of the target can), the view warped from the template shows an intensity profile very similar to that of View 2. For this reason, with the warping the cross-correlation result is much improved.

Table 1 shows the normalized cross-correlation results of different scalings of the template with the various candidate views, when the selected template is Template *a*. Table 2 shows the same when the selected template is Template *b*. In both cases the correct candidate view is identified. Cross-correlation values with the correct candidate view, under all scalings of the template, are consistently higher.

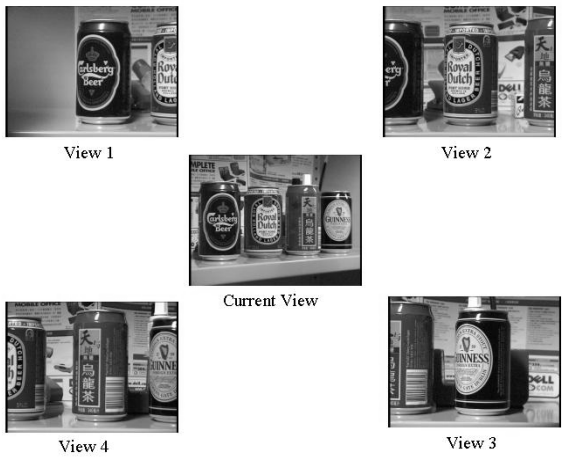

Figure 3: An example current view of e-shopper, and four candidate views (View 1, 2, 3, and 4) for viewpoint change. Depending upon where the e-shopper mouse-click on the current view, a sub-image is cropped from the current view around the clicked position and used as the template for triggering viewpoint change.

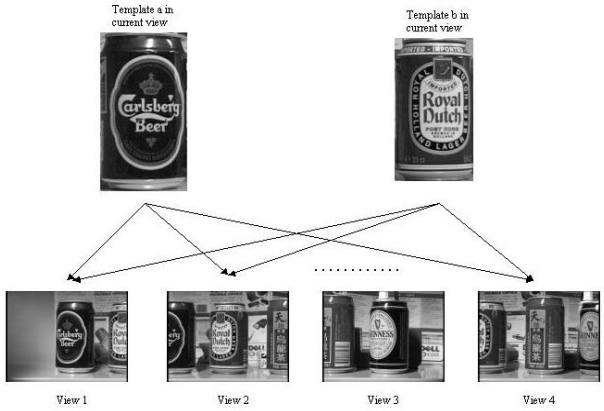

Figure 4: Input templates *a* and *b*, and the candidate views (Views 1, 2, 3, and 4).

Template *b* View 2

Figure 5: Consistent correspondences between Template *b* and View 2, and the resultant triangulation meshes over the matched portions.

(a) Template *b* (b) Template *b* warped (c) View 2

Figure 6: The warping of Template *b* (only the portion matched with View 2) from the viewpoint of the current view to the viewpoint of View 2. The warping is essential for cross-correlation comparison, in the same viewpoint, of the selected template (of the current view) with the candidate view (View 2 here).

Cand. View	**110% scaling of templ.**	**120% scaling of templ.**	**130% scaling of templ.**	**140% scaling of templ.**	**150% scaling of templ.**
View 1	0.56062	0.50791	0.63959	0.61164	<u>0.72031</u>
View 2	0.28183	0.1956	0.17774	0.41092	0.24265
View 3	0.04504	0.35909	0.11826	0.14999	0.168
View 4	0.33475	0.44884	0.34427	0.15872	0.23715

Table 1: Normalized cross-correlation estimates for all candidate views, when Template *a* is the selected template in the current view. The underlined estimate is the best cross-correlation, which corresponds to the correct candidate view (View 1).

Cand. View	**110% scaling of templ.**	**120% scaling of templ.**	**130% scaling of templ.**	**140% scaling of templ.**	**150% scaling of templ.**
View 1	0.26362	0.30441	0.08682	0.11114	0.24763
View 2	0.57578	0.69393	0.71286	0.71129	<u>0.77965</u>
View 3	0.25792	0.16092	0.20149	0.25123	0.2627
View 4	0.35898	0.18053	0.24444	0.30757	0.24762

Table 2: Normalized cross-correlation estimates for all candidate views, when Template *b* is the selected template in the current view. The underlined estimate is the best cross-correlation, which corresponds to the correct candidate view (View 2).

V. Conclusion and Future Work

We have introduced a new concept of e-shop that allows e-shoppers to perceive the atmosphere of physical shopping. The key problem to solve in implementing the proposed e-shop is how to determine the intended viewpoint movement or the intended merchandise to purchase of the e-shoppers. Using supermarket as an example, we have proposed a framework to tackle the problem. Experimental results with real images are promising.

Our proposed solution makes use of the multiple-homography model for representing merchandise. In a way we model each merchandise as a piecewise-smooth object. The multiple-homography model can approximate a large class of shapes: shapes that has have zero or low curvature at every surface point. However, for specific shapes like quadric surfaces (which include cylindrical surfaces and planar shapes as special cases), more accurate image-to-image mapping exists [4]. It is our plan to also investigate such mappings in the future work.

Acknowledgments

The work described in this paper was partially supported by a grant from the Research Grants Council of Hong Kong Special Administrative Region, China (CUHK Direct Grant; Project No. 2050230).

References

[1] M. Bern and D. Eppstein. Mesh Generation and optimal triangulation. In "Computing in Euclidean Geometry", by D.-Z. Du and F.K. Hwang, eds., World Scientific, 1992, pp. 23-90.

[2] O. Faugeras. *Stratification of three-dimensional vision: projective, affine, and metric representations*. In Journal of the Optical Society of America: A, March 1995, Vol. 12, No. 3, p. 465-484.

[3] M.A. Fischler and R.C. Bolles. *Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography*. Commun. Assoc. Comp. Mach., Vol. 24, 1981, pp. 381-95.

[4] A. Shashua and S. Toelg. *The Quadric Reference Surfaces: Theory and Applications*. In International Journal of Computer Vision, 1997, Vol. 23, No. 2, p. 185-198.

[5] P.H.S. Torr and D.W. Murray. *The development and comparison of robust methods for estimating the Fundamental matrix*. In International Journal of Computer Vision (IJCV), Vol. 24, No. 3, September 1997, pp. 271-300.

[6] Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong. *A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry*. In Artificial Intelligence, Vol.78, October 1995, pp. 87-119.