Skin Patch Trajectories as Scene Dynamics Descriptors

B.Zhan, N.D.Monekosso and P.Remagnino
Kingston University
Surrey, UK
{B.Zhan,P.Remagnino}@kingston.ac.uk

T.Rukhsana, A.Mansur and Y.Kuno
Saitama University
Saitama, Japan
kuno@cv.ics.saitama-u.ac.jp

Abstract

There is an increasing interest in the concept of intelligent environments where a closed or delimited public space (shopping mall, station, museum, hospital etc) is endowed with some automatic ability to interpret human behavior. Intelligent environments interact with their users, aiding, serving and pre-emptying them. In a not too distant future, this paradigm - in Europe called ambient intelligence - will soon include robotic platforms. Both intelligent environment and robotic platforms will collaborate to better inform the inhabiting or visiting user. This paper presents some steps towards that direction, describing a study on some scene descriptors, which can be employed to provide an automatic interpretation of the clutter and dynamics of a complex scene.

1. Introduction

The main goal of this research is to estimate automatically the amount of clutter and the level of dynamics in a complex scene, frequented by an unspecified number of people.

This is important in applications where situation assessment is crucial to better inform people inhabiting a specific environment. For instance, in a shopping mall or in a museum, individuals and more or less large groups of people might pass or stop by to window shop or to observe an exhibit. In such cases information about merchandise or exhibit could be delivered in a more efficient manner, for instance with the aid of a robot.

An automatic estimation of clutter and dynamics is also important in crucial situations, where people must be informed of exits and escape routes.

What in Europe is now called ambient intelligence and in the United States goes under the name of smart or intelligent environments, is a paradigm which has been in the mind of artificial intelligence researchers for some time [9]. The idea is of a *living* environment, able to interact with the user to make their lives easier. The emerging phenomenon of robotic platforms, seen more as *companions* than mechanical machines, inspires the idea of a living environment, where both the surroundings and robots collaborate between them and with the user to improve productivity (factory or office), security (public space), safety (nursing home or hospital). This *symbiotic* existence actively assists the user, seen either as a casual passenger or pedestrian in the environment, or as a frequent visitor (station or shopping mall) or even the person inhabiting (home) the intelligent space.

This paper presents a method to estimate dynamics, offering a means to evaluate its amount and classify people behavior as interested or uninterested in the scene. One can them imagine the degree of interest in a scene being used to inform a robotic platform to deliver a specific message to the user.

The next sections describe the proposed method and illustrate some examples of how it could be employed to assess a situation.

2. Methodology

Conventional cameras, used in museums and shopping malls can capture full human figures and sometimes human faces. Video data of this kind can be employed to recognize and track people in a complex environment. Full figure chromatic and structure models can be built [7], and people physiognomy, gait and shape characteristics have indeed been used to suit this purpose. In this paper we use skin color to extract exposed patches of the human body figure and we show that those can be robustly tracked throughout a scene. The tracks are then employed to annotate the dynamics of individual patches and draw some qualitative and quantitative description of the global evolution of the scene.

The following sections describe in detail our method, employed to extract and track skin color patches, and estimate trajectory trends.

2.1. Robust skin color detection

This part of our method makes use of the already proved idea that skin color can be indeed modeled across races so long as a suitable color space is employed. In our experiments we convert the color frames from the RGB to the YUV space. The choice of the YUV color space is justified by a fast conversion and the factorization of chromatic and illumination features. The illumination Y component is easily factored out, and the chromatic UV plane is employed to estimate the skin color model (an example of color probability density function – PDF - of skin color is shown in Figure 1).

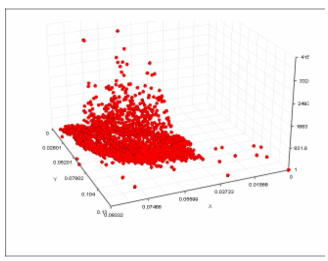

Figure 1: The PDF of skin color patch.

As already shown in other research, a skin color model can be estimated by acquiring video data of skin color patches and via the training of a color model using the expectation maximization algorithm [8]. In order to optimize the model, the skin color data can be studied and an optimal initialization defined in terms of number of clusters and initial positions and approximating functions.

Figure 1 illustrates the PDF of image data used to train the skin color model. All data is constrained in a small region of the $[u,v]$ plane and a mixture of Gaussians $\{\mu_i, \Sigma_i\}$ can serve as a good explicit approximation of the distribution.

Figure 2: PDF of data used to train the skin color model.

Figure 2 illustrates the probability masks of some of the data used to build the skin color model. The model is fairly robust to changes in illumination but it has the weakness of being specific to the camera used to acquire the training data. In all our tests, each new video camera we have used to acquire video footage had its own color model. As the training can be performed off line, the limitation is not prohibitive.

2.2. Color patch tracking

The MEANSHIFT method, based on an old idea of Fukunaga [1] and resurrected by Cheng [2] and Cominiciu [3], has been proven very robust for the tracking of objects and people in cluttered scenes. The MEANSHIFT algorithm tracks an object by estimating the drift of the underlying density function representing the evolving process. The limitation of the MEANSHIFT method stems from its inability to deal with time varying density functions. The CAMSHIFT algorithm proposed in [4] adapts to evolving PDFs by alternating cycles of the MEANSHIFT algorithm with a resizing of the search window. The window size is a function of the center of mass of the probability density map (zeroth moment).

Figure 3: LEFT: frame with bounding rectangles of recognized skin patches and RIGHT: related probability map.

Tracking color patches entails running the CAMSHIFT algorithm for each patch. However, this is not sufficient to maintain hypotheses in a rapidly evolving scene. That is why our method keeps track of a list of *alive* patches, by tracking them throughout the scene with the CAMSHIFT algorithm, removing those which have too low a probability associated for a number of frames and introducing new patches, whenever sufficiently large new patches appear in the scene with a sufficiently high probability.

Figures 4 and 5 illustrate four frames where skin color patches are identified and tracked throughout the scene.

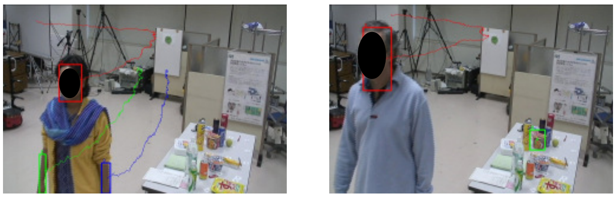

Figure 4: The two above frames show the low curvature trends of uninterested behavior: when people pass by an exhibit.

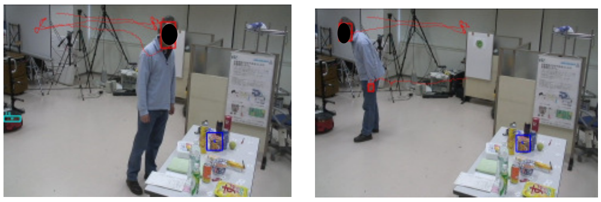

Figure 5: The above frames show when people are interested in the shown exhibits and they stop by the exhibit. Trends of such trajectories have higher curvature.

The trajectories of Figure 4 and 5 also illustrate two occasional problems: (i) FALSE NEGATIVES: sometimes not all exposed skin is recognized (because of the small size of the patch and because of the limitation of the color model) and (ii) FALSE POSITIVES: at times patches not of skin color are detected, these are commonly stationary objects and their stationary position can be used to eliminate misrecognition cases.

2.3. Dynamics estimators

Trajectories of skin patches identify people trajectories and can be seen as signatures of people behavior.

For instance, people interested in exhibits or merchandise have a more irregular signature, distinguished by curvature that becomes higher and changing more frequently, when the patches represent people looking at an object.

The amount of time spent in the scene also plays an important role: the shorter the time the smaller the interest shown in the exhibit/object. Frames in Figure 4 illustrate two examples of uninterested behavior, well correlated with a smoother (low curvature) trajectory, while Frame 5 clearly illustrates how the interest in an object is correlated with a change in curvature

Dynamics can therefore be estimated by studying the trajectories of the tracked skin color patches and making use of their trends. An in depth study of the trajectories led us to the following conclusions, all based on the assumption that the extracted skin patches belong indeed to people in the scene:

- Average number of patches and their speed over a period of time can be used to estimate the entropy of the scene: the higher the number of patches the more people are in the scene and the histogram of speed values over time and its change illustrates the amount of movement in the scene (the flatter the spread the higher the entropy),

- Fast patch movements indicate people in the scene are moving rapidly: the speed of each patch is estimated by the distance in pixels of a patch between frames,
- The curvature of a trajectory is a good indicator of how many twists and turns the trajectory trend has. Changes in curvature might occur more frequently in some moments then others: the frequency of change and the magnitude of curvature is an indicator of the person interest in some parts of the scene,
- A density signature of curvature peaks can therefore be estimated to describe people interest: the higher the density the higher the attention a person has for an object. Our study demonstrates that highly interested people will stop and move about in front of the object, uninterested or little interested people will move a lot in the scene and stop rarely and their curvature signature shows trends with small number of peaks of small number of high peaks. A suitable time window is defined to estimate the density: typically a number of seconds usually spent by a person to observe an object in the scene. This parameter depends on the application and can be learnt.

The figure below (Figure 6) illustrates the speed and curvature trends of a patch used to train the model of uninterested people. The speed becomes fairly high, however, the curvature remains lower than a low threshold; typically around 1.

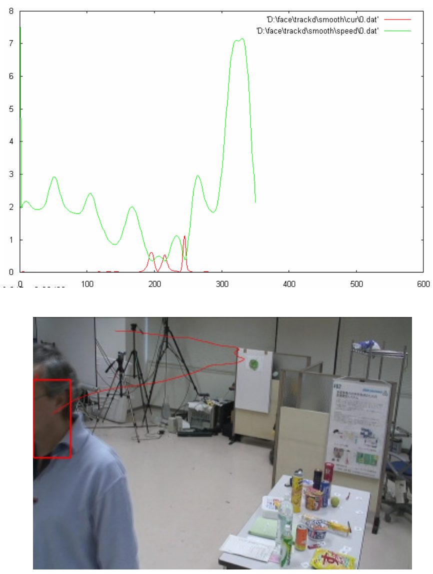

Figure 6: Speed (green) and curvature (red) of a patch of a person uninterested in the monitored scene.

The following graph (Figure 7) illustrates the signature of a patch related to a person who is interested in the scene. The speed is lower, indicating the person pays more attention to the scene. The frame in Figure 7 clearly shows the close occurrence of curvature peaks in two points of the scene, indicating that the person stopped, they looked around for a while, before moving to the next area of interest to stop again and observe, before leaving the scene. The yellow trajectory - associated with a hand – was picked up too late to illustrate the curvature phenomenon typical of an *interested* behavior.

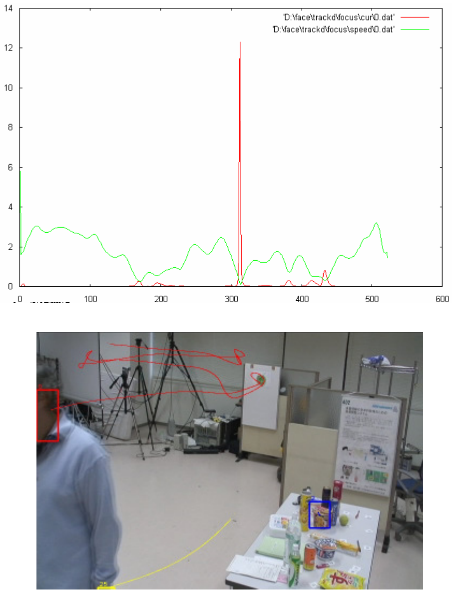

Figure 7: Speed (green) and curvature (red) of a path of a person interested in the object in the scene.

The following graph (Figure 8) illustrates what we call uninterested and animated behavior, characterized by patches of people uninterested in the scene objects, but where those people stay for longer in the scene and move about without really focusing on any object and they do not stand still in any particular position of the scene.

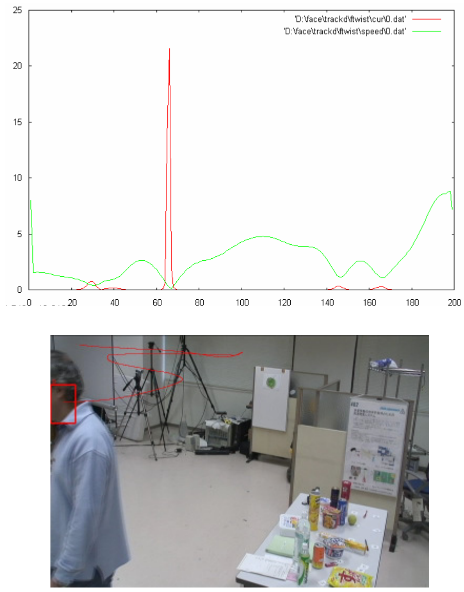

Figure 8: Speed (green) and curvature (red) of a patch of a person with animated behavior and uninterested in the scene.

As can be seen in the above graph, such behavior shows a large number of sparse high curvature peaks and it also correlates with a higher speed, indicating that the person did not stop for longer than the short period of time required to change direction in the scene a few times and then leave the scene.

3. Trajectory classification

Experiments were run in a University laboratory and all scenes filmed from a single camera.

A number of experiments were run with individuals performing the same action repeatedly more times. Mixtures of actions were then recorded with more people in the scene performing either the same or different actions.

The following graphs (Figure 9) illustrate how the density of curvature maxima can be employed to disambiguate between an interested, uninterested and animated behavior.

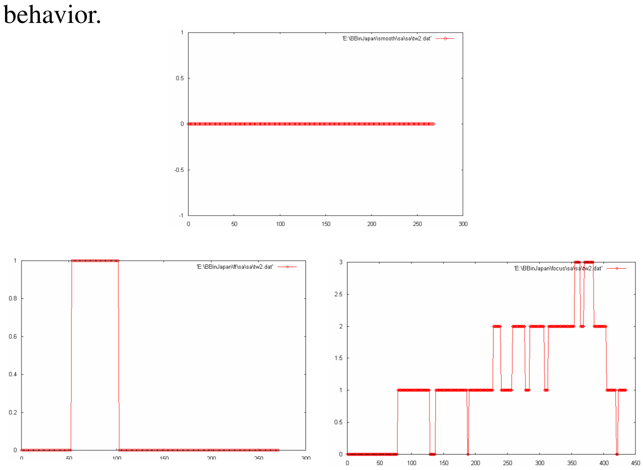

Figure 9: Examples of curvature density, estimated in a time windows of 100 frames. TOP: uninterested behavior; BOTTOM LEFT: animated behavior, BOTTOM RIGHT: interested/focused behavior

The graphs clearly indicate that whenever a person is interested in the scene objects, then they stop and spend time looking and while they do so they move about, building up a density of curvature maxima. Completely uninterested behavior shows no density at all for maxima above a threshold estimated by measuring the mean curvature of patches of uninterested people. Finally, animated behavior builds some density which, however, is not comparable with the density built for a focused behavior.

4. Conclusions

The paper has presented dynamics descriptors that make use of a skin color tracker and the trends of the tracked trajectories to infer a simple description of behavior in the scene. Preliminary experiments illustrate that curvature can be indeed employed to analyze trajectories and classify behavior. The amount of skin color patches in the scene and their life spans can shed some light on the clutter in the scene and their dynamics can be employed to assess a highly changing situation. The next step will be to further test our proposed method, provide a more automatic way to categorize scenes and the inclusion of robotic platforms, whose introduction in the scene is selected by the classification of dynamics. The introduction of robots and their interaction with the people present in the scene will then modify the dynamics and part of our future work will be to measure the dynamics "gradient" the human-robot interaction has caused. Briefly, one can envisage a robot being introduced to make an announcement, illustrate an exhibit or guide people in the scene. It is expected that a robot will reduce the entropy of the scene, and increase the interested of people.

5. Acknowledgements

Special thanks go to the Royal Academy of Engineering for funding Dr Remagnino's sabbatical in Japan, and to Professor Yoshinori Kuno for hosting both Remagnino's and Zhan's sabbatical at Saitama University. Dr Remagnino is also grateful to Dr Takashi Yoshimi, Humancentric Laboratory Corporate Research & Development Center Toshiba Corporation, and the Toshiba Science Museum for providing video data.

References

[1] K. Fukunaga and L.D. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," IEEE Transactions on Information Theory, vol. 21, pp. 32-40, 1975.

[2] Y. Cheng, Mean Shift, Mode Seeking, and Clustering, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 17, Issue 8, PP. 790-799, 1995.

[3] D. Comaniciu, V. Ramesh, P. Meer: Real-Time Tracking of Non-Rigid Objects using Mean Shift, IEEE Conf. Computer Vision and Pattern Recognition, Vol. 2, 142-149, 2000.

[4] G.R.Bradski, Computer Vision Face Tracking For Use in a Perceptual User Interface, Intel Technology Journal, Q2, Num. 15, 1998.

[5] S.Gong, S.J.McKenna, A.Psarrou, Dynamic Vision: From Images to Face Recognition, Imperial College Press, 2000.

[6] X.Wang, K.Tieu and E.Grimson, Learning Semantic Scene Models by trajectory analysis, European Conference in Computer Vision, 3, pp.110-123, 2006.

[7] Z.Zhang, K.Huang and T.Tan, Comparison of Similarity Measures for Trajectory Clustering in Outdoor Surveillance, International Conference on Pattern Recognition, Vol. 3, pp. 1135-1138, 2006.

[8] Arthur Dempster, Nan Laird, and Donald Rubin. "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society*, Series B, 39(1):1–38, 1977.

[9] P.Remagnino and G.L.Foresti, Ambient Intelligence: a New Multidisciplinary Approach, IEEE Transactions on Systems, Man and Cybernetics, Volume 35, Issue 1, pp. 1-6, 2005.