Advance in Head Pose Estimation from Low Resolution Images: A Review

Teera Siriteerakul

King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand

Abstract

In the task of analyzing image of individuals, head pose is one of the most vital pieces of information ones can acquired from the image. Thus, several researchers have attempted to find a way to automatically extract such information with sufficient accuracy. On the other hand, due to the increasing of surveillance system which generally produce head images in low resolution, many researchers has shifted their focus to techniques that would applicable to low resolution environment. As a result, there are several promising works in this area. In this article, difficulties of head pose estimation from low resolution images will be discussed along with an organized survey describing the development of the field. The survey will describe characteristic papers categorized according to their underlying techniques. In each category, advantages and limitations will be discussed followed by their potential to function in practical application. Finally, some potential research topics will be discussed.

Keywords: *Low-resolution head pose estimation, human-computer interfaces, low-resolution images, surveillance system*

1. Introduction

Head orientation provides crucial information for analyzing human behavior which can be useful for many applications (see Zeng et al. [41] and Cristani et al. [7] for the reviews of human activity analysis methods). There is several reported systems utilized head pose estimation in many areas. For examples:

Driving safety: Murphy et al. [15] proposed a driver-assistant system to signal an alert when there are pedestrians appear outside the driver's field of view. Similarly, Ray and Teizer [19] proposed a method to detect driver's blind spot. On the other hand, Schulz et al. [25] proposed a way to predict pedestrian path which can help drivers make some vital decisions.

Social behavior analysis: Stiefelhagen [31] and Zhang et al. [42] proposed a method which can be used to analyze behavior of people in a meeting. Later, Zen et al. [40] and Reid et al. [20] proposed a study of human interactions based on head pose as well. Similarly, Chen and Aghajan [6] provides framework for study human interaction in work environment based on relative head pose and proximity.

Advertisement: Smith et al. [29] proposed a system for tracking the visual focus of attention for a varying number of wandering people. The system can be used to count how many people look at a particular outdoor advertisement.

Surveillance: Benfold and Reid [4] proposed a system that can guide a surveillance system to focus on areas of attention of subjects in the video. On the other hand, Sankaranarayanan et al. [24] proposed a method to track visual attention of multiple subjects using multiple surveillance cameras.

While there are several reported systems utilized head pose estimation, the accuracy of the techniques for head pose estimation from low resolution images is subject to improvement.

Although the task of head pose estimation seems to be trivia for human, it can be very difficult for computer. Firstly, there are numbers of appearance variation of individuals such as skin color, hair color, or location of facial features such as location of the eyes. These variations effect both geometry and texture of head images. Secondly, in an unconstrained environment, there can be occlusions from individual themselves such as eye-wear, hairstyle, or hat. Additionally, the occlusions like trees, cars, or other individuals, can come from environment as well. Thirdly, scenes can have very different lightings subject to day time, night time, indoor, outdoor, or other environments. Lastly, image noises can be of problem similar to most of the computer vision system. However, while these difficulties are associated with head pose estimation, there are additional complexities when dealing with low resolution images.

In recent year, the usage of surveillance systems has been rapidly increasing. Since these systems usually produce head images in far view with imperfect image quality, the ability to estimate head pose from low resolution images robustly becomes more important. However, achieving high accuracy from low resolution environment can be difficult due to image quality and a limited number of pixels.

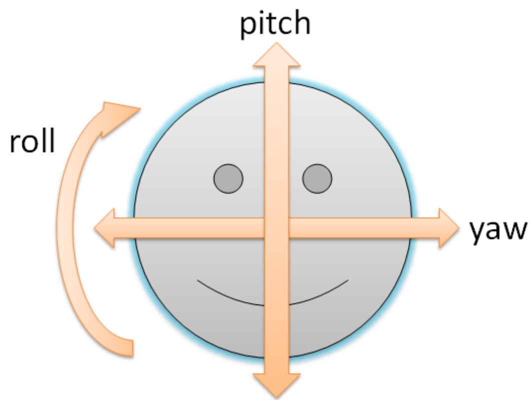

Fig. 1: Three degrees of freedom (DOF), which is adopted from Euler angles, used in head pose estimation literatures.

The insufficient quality and quantity of pixel can be problematic in several ways. Firstly, detail features will be lost or difficult, if at all possible, to track. Thus, most reported works rely on appearance based rather than model based. Secondly, pixelization can be of a problem. For example, when a low-intensity pixel that represents an eye moved from one pixel to the next in low resolution images, the eye location extracted from such images will be excessively different from each other. Lastly, image noise, especially non-Gaussian noise, can have profound effect on images with limitedly available information.

In the context of computer vision, head pose estimation is most commonly interpreted as the ability to infer the orientation of a person's head relative to the view of a camera [14]. At the finest level, head poses can be measured continuously up to three degrees of freedom (DOF). These DOF's are usually referred to as pitch, roll, and yaw as illustrated in figure 1. However, it usually not enough information for such refined estimation in low resolution images. Therefore, head pose are usually discretized into multiple pose classes (e.g. eight classes of yaw direction with 45 degrees of separation).

This article presents a survey of the head pose estimation methods for low resolution images. The survey begins by organizing methods by their common approaches. Then, advantages and limitation as well as their potential for practical application are discussed. Thereafter, this article discusses some research opportunities in each approach as well as in the area in general. Finally, the concluding remarks will be given with some overall discussion about the field.

2. Methods for Head Pose Estimation from Low Resolution Images

This section presents the categorization of the estimation methods based on the fundamental approach that underlies their implementations. Table 1 summarizes four general approaches for head pose estimation from low resolution images, example works for each approach, as well as the brief explanations of each approach. The details of each approach are described next. In each approach, functional requirements are discussed along with their advantages and limitations.

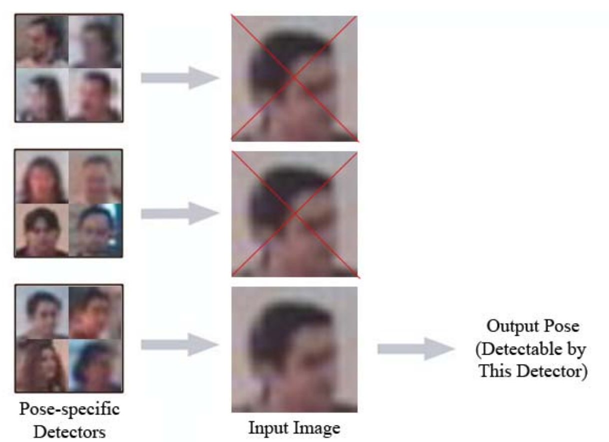

Fig. 2: Detector array applies multiple face detectors, each with different head pose or from different cameras, to detect face in the scene.

2.1 Detector Array Approach

Some of the early works dealing with head pose estimation from low resolution images are the methods that naturally evolved from face detection. In detector array approach, methods apply multiple face detections for different view or from different cameras to detect change in poses as illustrated in figure 3. Rowley et al. [23] applied a router network to resolve roll angle after a patch is assumed to be a face image. On the other hand, Huang et al. [11] used Support Vector Machine (SVM) to discriminate face poses in three discrete yaw angles. Later, Jones and Viola [12] improved the approach of Rowley et al. with AdaBoost, a technique from their famous work [35]. Then, Zhang et al. [42] proposed a system using multiple cameras and FastBoost, a variant of AdaBoost, to estimate head pose in seminar room.

Table 1: Approaches of methods for head pose estimation from low resolution images.

Approaches	Example Works	Brief Descriptions
Detector Array	Rowley 98[23] Huang 98[11] Jones 03[12] Zhang 07[42]	Using multiple face detectors, one for each poses, to estimate head pose from images.
Appearance Template	Robertson 06[21] Benfold 08[3] Orozco 09[16] Siriteerakul 10[27]	Compare input images to a set of templates specific to each discrete pose.
Functional Mapping	Rae 98[18] Stiefelhagen 02[30, 31] Tian 03[32] Seemann 04[26] Gourier 06[8] Voit 06[36, 37] Balasubramanian 07[1] Wang 08[38] BenAbdelkader 10[2] Tosato 10[33]	Create a functional mapping from the image space to the head pose angle space.
Tracking	Wu 00[39] Ho 04[10] Ross 04[22] Tu 06[34] Siriteerakul 11[28]	Estimate change in head pose from series of images taken from consecutive video frames.

Since detector array derives from face detection, the techniques using this approach can perform both face and pose detection which as be considered as a great advantage. However, since a face detector is trained for specific head orientation, multiple detectors or multiple cameras are required for head pose estimation. This results in increasing of computational cost or system requirement. Another major drawback of this approach is that each detector used in the system needs considered amount of training data, both positive and negative. Thus, the amount of training data needed can be in the neighborhood of thousands of images.

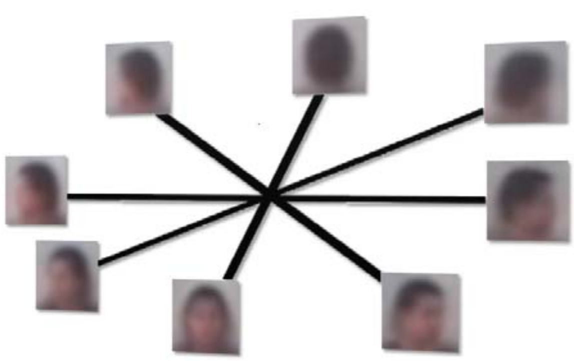

Figure 4: Eight pose classes along yaw direction used in low resolution setting.

2.2 Appearance Template Approach

Appearance template methods match head images into one of the discrete head poses using image-based comparison metrics as illustrated in figure 2. The method proposed by Robertson and Reid [21] matches 10x10 pixel head images to one of the eight poses (illustrated in figure 5) using descriptor based on hand-select skin histogram. Hand skin color selection requirements were later dissolved by Benfold and Reid [3] using randomized fern proposed by Ozuysal et al. [17]. On the other hand, Orozco et al. [6] asserted that explicit skin/hair separation by Benfold and Reid may not be feasible in low resolution images. Instead, Orozco et al. proposed another image descriptor based on pseudo Kullback-Leibler divergence between the input image and the mean images of each head pose classes computed from the training images. Later, Siriteerakul et al. [27] proposed image descriptors based on non-local difference of pixel values in head images to improve the classification accuracy

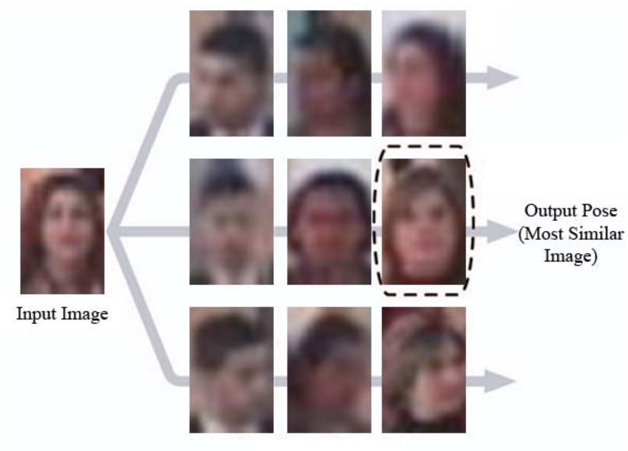

Figure 3: Appearance template methods compare input image with pose classes learned using parametric or non-parametric (illustrated in this figure) machine learning tools.

This approach has been reported to work well for low resolution images. One limitation of this approach is the fact that two images of the same person with different poses may appear more similarly than two images of the different persons with the same pose. Therefore, in order for the appearance template approach to work, images must be represented in such a way that the head pose images from the same orientation class are grouped together. Furthermore, these techniques require good head localization which itself can be considered a hard research problem.

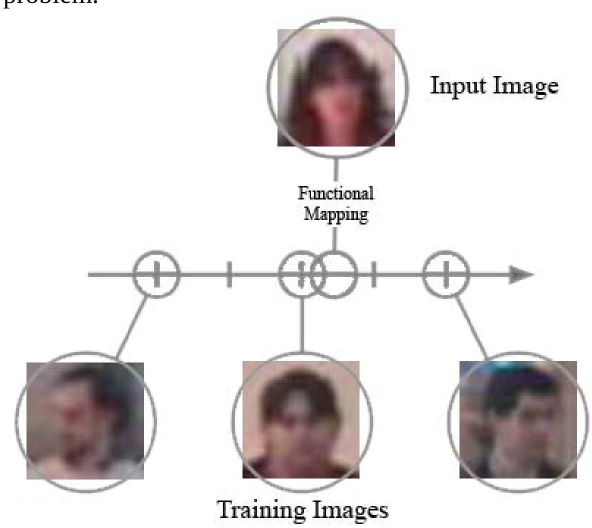

Figure 5: Functional mapping method learns the mapping from image space to head pose space using training data and use the mapping to map an input image to a head pose value.

2.3 Functional Mapping Approach

Rather than mapping head image to discrete poses, the functional mapping methods attempt to map information from image space to head pose space continuously. Guo et al. [9] provides a comparison study between these two. In the functional mapping approach, there are several methods adopted Artificial Neuron Network as their non-linear regression tool [8, 18, 26, 30, 31, 32, 36, 37]. On the other hand, there are many recent methods reported to use manifold embedding technique as their regression tool as well [1, 2, 33, 38].

The nonlinear regression approach can be very fast and accurate in practice. However, in order for the nonlinear regression approach to work, two head pose images with the same orientation should be mapped to the same point, or at least close to each others in the space of head orientation. Unfortunately, in the case of low resolution images, two images of the same person with different poses may be mapped closer to each other than images of different persons with the same pose. Therefore, this approach requires a robust image representation that is only, or at least more, sensitive to changes in head orientations. Another limitation of this approach is the fact that the refined mapping may not work in low resolution images due to limited information. Furthermore, similar to the appearance template approach, the nonlinear regression approach requires accurate head localization.

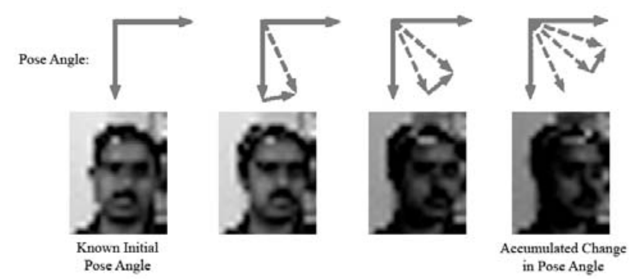

Fig. 6: Head pose are tracked for relative change in tracking methods.

2.4 Tracking Approach

Tracking approach utilizes temporal continuity and smooth motion of head poses in consecutive video as illustrated in figure 6. Similar to the detector array approach, the tracking approach evolved naturally from object tracking. Taking this approach, Wu and Toyama [39] uses an ellipsoid with a set of points on the surface where its north-pole is at the nose tip. Then, each point were trained to match local image features and used for tracking. Alternatively, Cascia et al. [5] proposed a method to track human head by placing the initial frame texture to a Cylindrical Head Model (CHM) and then perform translation and rotation to the CHM to match the current frame texture. Later, Ho et al. [10] and Ross et al. [22] independently adopted the CHM and replace the initial frame texture with adaptive PCA for texture matching.

Table 2: Mean angular errors and correct classification rates reported by each work.

Year	Publication	Mean error		Classification Rate	Dimension	Approach
		Yaw	Pit			
2000	Wu [39]	34.6	14.3	-	-	Tracking
2003	Tian [32]	-	-	79%	16x16	Mapping
2004	Seemann [26]	7.5	6.7	75.2%	24x32	Mapping
2006	Tu [34]	4	4	-	10x13	Tracking
2006	Robertson [21]	5.6	-	-	20x20	Template
2007	Zhang [42]	33.6	-	87%	-	Detector
2008	Benfold [3]	26.6	-	-	10x10	Template
2009	Orozco [16]	-	-	80%	20x20	Template
2010	Siriteerakul [27]	11.5	-	78.6%	20x20	Template
2010	Siriteerakul [28]	3.3	-	-	10x13	Tracking

Subsequently, Tu et al. [34] introduced a forgetting mechanism to the adaptive PCA. Later, Siriteerakul et al. [28] replaced the adaptive PCA with an LBP-based comparison in order to reduce the effect of noise and change in illumination.

While this technique can have fairly high accuracy, it usually requires an initial step such as frontal view or manual initialization which would limit practical use. Methods taking this approach also need some re-initialized step whenever tracking is lost. Furthermore, it is only applicable to continuous video.

3. Comparison of Head Pose Estimation Approaches

This section attempts to give relative performance of methods categorized in section 2. Firstly, the reported accuracy are compare to give some general idea of their relative performance. Then, a set of assumptions is listed for each reported works to provide some sense of real-world applicability.

3.1 Comparison of Publish Results

Due to lack of standard dataset of head pose images in low resolution, this article will rely on published results of each works. In these works, absolute angular error is a common information metrics for evaluating head pose estimation. Another common information metrics for the head pose estimation is the rate of correct template matching. These two information metrics are summarized in table 2. By correct matching rate, detector array method proposed by Zhang et al. [42] seems to yield better result than the methods taking functional mapping approach [26, 32], but their absolute errors suggested otherwise. The reason being Zhang et al. considers neighboring poses to be correct classification as well. It is rather hard to draw any conclusion about which technique performed better than the others since the data set and the assumptions used were diverse. One conclusion we can draw is both tracking [28, 34] and appearance template [16, 21, 27] techniques can yield decent accuracy results. However, further improvements are required for both approaches which will be discussed in section 4.

3.2 Comparison of Real-World Applicability

Similar to [14], in order to get some sense of how the reviewed methods would perform in the real world application, this article provides a list of assumptions used in the reported papers. These assumptions are

A Continuous video assumption Only small change of head pose allow in subsequence video frame.

B Initialization assumption The initial poses must be known before a system can begin the estimation.

C Head localized assumption Location of head in an image is known and head image can be cropped with sufficient stability.

D Multicamera assumption Subject is visible by two or more cameras at a small enough distance to discriminate depth information across the face.

Regardless of their reported accuracy, these assumptions would limit their practical used. List of papers and their assumption can be found in table 3. Generally speaking, methods taking tracking approach usually require assumption A and B while methods taking appearance template approach requires assumption C. Note that, although work by Wu and Toyama [39] requires no assumption, its accuracy is too low to be used in most practical applications.

Table 3: Comparison of real-world applicability.

Year	Publication	Approach	Assumption	DOF	Domain
2000	Wu [39]	Tracking	-	3	1024 poses
2003	Tian [32]	Mapping	D	1	Up to 12 poses
2004	Seemann [26]	Mapping	D	2	Continuous
2006	Tu [34]	Tracking	A, B	3	Continuous
2006	Robertson [21]	Template	A	1	8 poses
2007	Zhang [42]	Detector	D	1	5 poses
2008	Benfold [3]	Template	A	1	8 poses
2009	Orozco [16]	Template	C	1	8 poses
2010	Siriteerakul [27]	Template	C	1	8 poses
2010	Siriteerakul [28]	Tracking	A, B	1	Continuous

4. Research Opportunities

This section discusses research opportunities specific to each approach as well as in the head pose estimation from low resolution images in general. These research opportunities are essentially based on limitation of each specific approach.

4.1 Detector Array Approach

There is not much room for technical research in detector array approach. This is due to the fact that this approach makes use of face detector which has been well developed for number of years. However, one can try to make the detector array faster in general. On the other hand, one can try to make use of frontal head pose estimation result from this approach for some practical application as it was done by Smith et al. [29].

4.2 Appearance Template Approach

There are several rooms for improvement regarding appearance template approach. First, in order to match an input image to a correct template, there must be a robust image representation. The improvements from Robertson and Reid [21] to Benfold and Reid [3] to Orozco et al. [16] and to Siriteerakul et al. [27] are all included the improvement in image descriptors. The second point for improvement to this approach is to use or develop a better distance metric or a classification tool. On the other hand, if there is no image descriptor that is robust to the location of head in the input image, this approach needs an outside system that can crop head images automatically and precisely.

4.3 Functional Mapping Approach

Similar to appearance template methods, the functional mapping approach requires robust image feature extraction as well as a robust functional mapping. Although it should be noted that several works reported in this article require multiple cameras to achieve the presented accuracies.

4.4 Tracking Approach

With reported performances of tracking methods, one expected research opportunity would be to try to remove the requirement of initialization/re-initialization. One way to remove these requirements is to embed some detector array techniques into the system as done by Murphy et al. [15].

4.5 General Research Opportunities

Generally, all the reported works only consider head pose images as an only mean of gathering information. In case of low resolution images, this can be troublesome. To eliminate such limitation, one can consider information from other source, especially cues from environment (see recently published work by Krahnstoever et al. [13] for an example). This can be head pose of other persons in the scene, objects in line of sight, or even saliency map of the scene. Another great opportunity for research is to developing some practical application that make use available head pose estimation techniques. Although there are limitations in existed works, it may be applicable to some less demanding applications.

5. Discussion and Final Remark

From the reported results, the tracking methods can provide the highest head pose estimation accuracy while requiring some initialization. This can be solved by couple the techniques in this approach by the techniques from another approach such as the detector array approach. On the other hand, the nonlinear regression might be good for practical use if reasonable amount of training data can be obtained. The appearance template approach, which has been popular in the recent years, can be improved with better image descriptor and distance metrics as well as head localization. The last statement is applicable to the functional mapping approach as well. Lastly, the detector array methods can be very practical in limited usage where only frontal poses are required. While much improvement is still in need for head pose estimation from low resolution images, some of the system shows great potential. The system proposed by Tian et al. [32] and Seemann et al. [26] performed considerably well in small room environment while the system proposed by Robertson and Reid [21] and Benfold and Reid [3] did similarly well in unconstrained field.

References

[1] V. Balasubramanian, J. Ye, and S. Panchanathan, "Biased manifold embedding: a framework for person-independent head pose estimation", in Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–7.

[2] C. BenAbdelkader, "Robust head pose estimation using supervised manifold learning", in Proceeding of the 11th European Conference on Computer Vision, 2010, pp 518–531.

[3] B. Benfold and I.D. Reid, "Colour invariant head pose classification in low resolution video", in Proceeding of the British Machine Vision Conference, 2008.

[4] B. Benfold and I.D. Reid, "Guiding visual surveillance by tracking human attention", in Proceeding of the British Machine Vision Conference, 2009.

[5] M.L. Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models" in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, Vol. 22, pp. 322–336.

[6] C. Chen and H. Aghajan, "Multiview social behavior analysis in work environments", in Proceeding of the Fifth ACM/IEEE International Conference on Distributed Smart Cameras, 2011, pp. 1–5.

[7] M. Cristani, R. Raghavendra, A.D. Bue, and V. Murino, "Human behavior analysis in video surveillance: a social signal processing perspective", Neurocomputing, 2012.

[8] N. Gourier, J. Maisonnasse, D. Hall, and J.L. Crowley, "Head pose estimation on low resolution images", in Proceeding of the 1st International Evaluation Conference on Classification of Events, Activities and Relationships, 2006, pp. 270–280.

[9] G. Guo, Y. Fu, C.R. Dyer, and T.S. Huang, "Head pose estimation: Classification or regression?", in Proceeding of the 19th International Conference on Pattern Recognition, 2008, pp. 1–4.

[10] J. Ho, K. Lee, M. Yang, and D. Kriegman, "Visual tracking using learned linear subspaces", in Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, 2004, pp. 782–789.

[11] J. Huang, X. Shao, and H. Wechsler, "Face pose discrimination using support vector machines (svm)", in Proceedings of the Fourteenth International Conference on, 1998 Pattern Recognition, 1998, Vo. 1, pp. 154–156 .

[12] M. Jones and P. Viola, "Fast multi-view face detection", Mitsubishi Electric Research Laboratories TR2003-96, 2003.

[13] N. Krahnstoever, M. Chang, and W. Ge, "Gaze and body pose estimation from a distance", in Proceedings of the 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance, 2011, 11–16.

[14] E. Murphy-Chutorian and M.M. Trivedi, "Head pose estimation in computer vision: A survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, Vol 31, pp. 607–626.

[15] E. Murphy-chutorian, A. Doshi, and M.M. Trivedi, "Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation", in Proceeding of the 10th International IEEE Conference on Intelligent Transportation System, 2007, pp. 709–714.

[16] J. Orozco, S.G. Gong, and T. Xiang, "Head pose classification in crowded scenes", in Proceeding of the British Machine Vision Conference, 2009.

[17] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua, "Fast keypoint recognition using random ferns", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, Vol. 32 pp. 448–461.

[18] R. Rae and H.J. Ritter, "Recognition of human head orientation based on artificial neural networks", IEEE Transactions on Neural Networks, 1998, Vol. 9(2).

[19] S.J. Raya and J. Teizerb, "Coarse head pose estimation of construction equipment operators to formulate dynamic blind spots", Advanced Engineering Informatics, 2012, Vol. 1 pp. 117–130.

[20] I. Reid, B. Benfold, A. Patron, and E. Sommerlade, "Understanding interactions and guiding visual surveillance by tracking attention", ACCV International Workshop, 2011.

[21] N.M. Robertson and I.D. Reid, "Estimating gaze direction from low-resolution faces in video", in Proceeding of the 9th European Conference on Computer Vision, 2006, Vol. 3952/2006, pp. 402–415.

[22] D. Ross, J. Lim, and M. Yang, "Adaptive probabilistic visual tracking with incremental subspace update", in Proceeding of the European Conference on Computer Vision, 2004, pp. 470–482.

[23] H. A. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection", in Proceeding of the Conference on Computer Vision and Pattern Recognition, 1998, pp. 38–44.

[24] K. Sankaranarayanan, M. Chang, and N. Krahnstoever, "Tracking gaze direction from far-field surveillance cameras", IEEE Workshop on Application of Computer Vision, 2011, pp. 519–526.

[25] A. Schulz, N. Damer, M. Fischer, and R. Stiefelhagen, "Combined head localization and head pose estimation for video-based advanced driver assistance systems", 2011, Vol. 6835, pp. 51–60.

[26] E. Seemann, K. Nickel, and R. Stiefelhagen, "Head pose estimation using stereo vision for human-robot interaction", in Proceeding of the International Conference on Automatic Face and Gesture Recognition, 2004, pp. 626–631.

[27] T. Siriteerakul, D. Sugimura, and Y. Sato, "Head pose classification from low resolution images using pairwise non-local intensity and color differences", in Proceeding of the Fourth Pacific-Rim Symposium on Image and Video Technology, 2010, pp. 362–369.
[28] T. Siriteerakul, Y. Sato, and V. Boonjing, "Estimating change in head pose from low resolution video using LBP-based tracking", in Proceeding of the Intelligent Signal Processing and Communication System, 2011.
[29] K. Smith, S.O. Ba, J.M. Odobez, and D.G. Perez, "Tracking the visual focus of attention for a varying number of wandering people", 2008, Vol. 30(7), pp. 1212–1229.
[30] R. Stiefelhagen, "Estimating head pose with neural network-results on the pointing04 ICPR workshop evaluation data", in Proceedings of the ICPR Workshop on Visual Observation of Deictic Gestures, 2004.
[31] R. Stiefelhagen, J. Yang, and A. Waibel, "Modeling focus of attention for meeting indexing based on multiple cues", IEEE Transaction on Neural Networks, 2002, Vol. 13(4).
[32] Y. Tian, L. Brown, J. Connell, S. Pankanti, A. Hampapur, A. Senior, and R. Bolle, "Absolute head pose estimation from overhead wide-angle cameras", in Proceeding of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003, pp. 92–99.
[33] D. Tosato, M. Farenzena, M. Cristani, M. Spera, and V. Murino, "Multi-class classification on Riemannian manifolds for video surveillance", in Proceeding of the 11th European Conference on Computer Vision, 2006, Vol. 3952/2006, pp. 378–391.
[34] J. Tu, T. Huang, and H. Tao, "Accurate head pose tracking in low resolution video", in Proceeding of the 7th International Conference on Automatic Face and Gesture Recognition, 2006, pp. 573–578.
[35] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", 2001, Vol. 1, pp. 511–518.
[36] M. Voit, K. Nickel, and R. Stiefelhagen, "A Bayesian approach for multi-view head pose estimation”, in Proceedings of the IEEE International Conference Multisensor Fusion and Integration for Intelligent Systems, 2006.
[37] M. Voit, K. Nickel, and R. Stiefelhagen, "Head pose estimation in single and multi-view environments – results on the clear '07 benchmarks", in Multimodal Technologies for Perception of Humans, 2008, Vol. 4625 of Lecture Notes in Computer Science, pp. 307–316.
[38] X. Wang, X. Huang, J. Gao, and R. Yang, "Illumination and person-insensitive head pose estimation using distance metric learning", in Proceeding of the 10th European Conference on Computer Vision, 2008, Vol. 2, pp. 624–637.
[39] Y. Wu and K. Toyama, "Wide-range, person- and illumination-insensitive head orientation estimation", in Proceeding of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 183–188.
[40] G. Zen, B. Lepri, E. Ricci, and O. Lanz, "Space speaks: towards socially and personality aware visual surveillance", in Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis, 2010.
[41] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, Vol. 1, pp. 39–58.
[42] Z. Zhang, Y. Hu, M. Liu, and T. Huang, "Head pose estimation in seminar room using multi view face detectors", in Proceeding of the 1st International Evaluation Conference on Classification of Events, Activities and Relationships, 2007, pp. 299–304.

Teera Siriteerakul received M.S. Applied Mathematics in 2001 and M.S. Information Technology in 2003 from Rensselaer Polytechnic Institute, USA. Since 2006, he joined the Faculty of Science at King Mongkut's Institute of Technology Ladkrabang in the Department of Computer Science as a Lecturer. His research activities include Computer Vision, Image Processing, and Pattern Recognition techniques relevance to Computer Vision.